PROFESSOR: If we're going to make use of Chebyshev's Bound and other results that depend on the variance, we'll need some methods for calculating variance in various circumstances. So let's develop that here.

A basic place to begin is to ask about the indicator variables and their variance. Remember, i is an indicator variable that means that it's zero one value. It's also called the Bernoulli variable. And if the probability that it equals 1 is p, that's also its expectations. So we have an indicator variable with expectation of the indicator is p, and we're asking what's its variance, which by definition is the expectation of i minus p squared.

Well, this is one of the sort of almost mechanical proofs that follows simply by algebra and linearity of expectation. But let's walk through it step by step, just to reassure you that that's all that's involved. I would recommend against really trying to memorize this, because it's-- I can remember it anyway, I just reprove it every time I need it. And so, let's see how the proof would go.

So step one would be to expand this i minus p squared algebraically. So we're talking about the expectation of i squared minus 2pi plus p squared. Now we can just apply linearity of expectation, and I get the expectation of i squared minus 2p times the expectation of i plus p squared. Of course, the expectation of a constant is the constant. So when I take expectation of p squared, I get p squared.

But now look at this. i squared is zero one value. So in fact, i squared is equal to i and the expectation of i has now appeared here, that's p. So this term simplifies to expectation of i, and this term becomes 2p times p plus p squared.

Of course that expectation of i is a p. So I've got p minus 2p squared plus p squared. The p squareds cancel, and I get p minus p squared. If you factor out p, that's p times 1 minus p, or pq, which is the standard way that you write the variance of an indicator variable. It's p times 1 minus p. OK, that was easy, and again, completely mechanical.

There's a couple of other rules for calculating variance of new variables from old ones that are basic. Like [? additivity ?] of expectation, but it doesn't quite work so simply for variance. So the first rule is that if you ask about the variance of a constant times r plus b, that turns out to be the same as a squared times the variance of b-- of r. The b doesn't-- the additive of be

doesn't matter, and the-- because the variance is really the expectation of something squared, when you get rid of that constant a, you're factoring out an squared. And this is the rule you get here. OK.

Another basic rule that's often convenient, instead of working with variance in the form of the expectation of r minus mu squared, is to say that it's the expectation of r squared minus the square of the expectation of r. Now, this expression-- the square of the expectation of r comes up so often that there's a shorthand for it. Where instead of writing [? parends ?] you write e squared of r, just means the same as expectation of r squared.

And so much for the second rule, which we'll use all the time, because it's a convenient rule to have. I'm going to prove the second one, just again, just to show you have nothing to worry about. You don't even have to remember how the proof goes, of course, you can reconstruct it every time. So it's again simple proofs just by linearity of expectation and doing the algebra.

So the variance of r is by definition the expectation of r minus mu squared. Let's expand our minus mu squared. It's the expectation of r squared minus 2mu r plus mu squared. Now we apply linearity to that. I get the expectation of r squared minus 2mu expectation of r, plus the expectation of mu squared, if I'm really being completely mechanical about linearity of expectation.

Now expectation of a constant mu squared is simply mu squared. And here, I've get the expectation of r, that's mu again. So I wind up with the expectation of r squared minus 2mu mu plus r squared. This is 2mu squared-- minus 2mu squared plus mu squared. It winds up with minus mu squared. And of course, mu squared is the expectation squared of r, I've proved the formula. Again, as claimed, there's nothing interesting here, just algebra and linearity of expectation. And the first result about factoring out an a and squaring it follows from a similar proof, which I'm not going to include here.

So let's look at the space station Mir again, which we used as an example of calculating mean time to failure. So the hypothesis that we're making is that with probability p, the Mir space station will run into some huge space garbage that will clobber it. And the probability of that happening in any given hour is probability p.

So we know that that means the expected number of hours for the Mir to fail is 1 over p, that's the mean time to failure. And what we're asking is what's the variance of f, if f is the number of

hours to failure? What's the variance of f?

Well, one way we can do is just plug-in the definition of expectation and this will work. The probability that it takes k hours to fail is-- we know the geometric distribution. The probability of not failing for k minus 1 hours, and failing after that, q to the k minus 1 times p.

So the variance of f, using our previous formula about the expectation of f squared minus the expectation squared of f, this becomes a minus 1 over p squared. We can forget about that, we want to focus on calculating the expectation of f squared. So f is 1, 2, 3, and so on. That means f squared is 1, 4, 9, k squared. The point being that the only values that f squared can take our squares, so we don't have to worry about counting them in to sum that defines the expectation. So let's go look at that.

So the expectation of f squared is the sum over the possible values that f squared can take, namely the sum from k equals 1 to infinity of k squared, times the probability that f squared is equal to k squared. Well of course, the probability that f squared is equal to k squared is the same as the probability that f equals k.

And we know what the probability of f equals k is, it's a geometric distribution, so the probability that f equals k is q to the k minus 1 times p. If I factor out a p over q, this simplifies to the sum from k equals 0 to infinity of k squared, q to the k. And this is a kind of sum that we've seen before that has a closed form and we could perfectly well calculate then the expectation of f squared by appealing to our generating function information to get a closed form for this, and then remember to subtract 1 minus p squared because the variance is this term minus the square of the expectation of f.

But let's go another way and use the same technique of total expectation that we used before. That is, the expectation of f squared, of the failure of time squared, is equal, by the law of total probability, to the expectation of f squared, given that f is one. That is, we fail f on the first step, times the probability that we fail on the first step. Plus, the expectation of f squared, given that we don't fail on the first step, that f is greater than 1, times the probability that f is greater than 1.

Now, what's going to make this manageable is that this expression, the expectation of f squared when f is greater than 1, will turn out to be something that we can easily convert into a nonconditional probability, and find a value for. So the limit that we're using here is the following. What I'm thinking about in mean time to failure-- if I think of any function whatsoever,

g of the mean time to failure. And I'm interested in the expectation of g of f, And I'm interested in the expectation of g of f, given that f is greater than n. That is, it's already taken n steps to get where I am.

Then the thing about the mean time to failure is that at any moment that you haven't failed, you're starting off in essentially the same situation you were at the beginning in waiting for the next failure to occur. And the probability of failing in one more step is the same probability-- is the same p. And the probability of you're failing in two more steps is qp, and three more steps is qqp.

The only difference is that the value of f has been shifted by n. In the ordinary case, we start off with f equals 0 and look at the probability that we fail in one more step, two more steps. Now we're starting off with f having the value f plus n, and asking about the probability that it fails in the next step or the next step, or the next step.

So the punchline is that the expectation of g of f, given that f is greater than n, is simply the expectation of g of f plus n. And I'm going to let you meditate that and not say anymore about it. But the punchline is the corollary that the expectation of f squared, given that f is greater than 1, is simply the expectation of f plus 1 squared.

And that lets us go back and simplify this expression. That we've had from total expectation, we now have-- here's the expectation of f squared, given that f is greater than 1. And let's look at these other terms. This is the expectation of f squared, given that f equals 1. Well, the expectation of f squared given that f equals 1 is 1 squared, because we know what f is and that's the end of the story. Times the probability that f equals 1, that's p, the probability of failure on a given step.

This is the probability that f is greater than 1, which is q, that we didn't fail on the first step. And we just figured out that this term is the expectation of the square of f plus 1. So there's the 1 and the p. And that becomes a q, and this is the expectation of f plus 1 squared.

Now again, I apply limit linearity. I'm going to expand f plus 1 squared into f squared plus 2f plus 1, and then apply linearity of expectation. And I'm going to wind up with the expectation of f squared plus twice the expectation of f, which remember, is twice over-- 2 over p, plus 1, times the q.

And now what I've got is a simple arithmetic equation between the expectation of f squared

and some other arithmetic and the expectation of f squared. It's easy to solve for the expectation of f squared. And I'll spare you that elementary simplification.

But the punchline is, when-- we also remember to subtract one over p squared, because that was the expectation of the square of f of the expectation of f. We came up with this punchline formula. The variance of mean time to failure is 1 over the probability of failure on a given step, times 1 minus 1 over-- times the probability-- 1 over the probability of the failure in the first step, minus 1.

That's just for practice and fun, let's look at the space station Mir again. Suppose that I tell you that there is a 1 in [? 10,000ths ?] chance that in any given hour, the Mir is going to crash into some debris that's out there in orbit. So the expectation of f is 10 to the fourth, about 10,000 hours.

And the sigma is going to be the variance of f, which is about 1 over ten thousandths, that is 10,000 times 10,000 minus 1, which is pretty close to 10,000 squared for the variance. And when I take the square root, I get back to 10,000. So sigma is just a tad less than 10,000, is 10 to the fourth.

So with those numbers, I can apply the Chebyshev's Theorem and conclude that the probability that the Mir lasts more than 4 times 10 to the fourth hours is less than 1 chance in four. If we translate that into years-- if it was really the case that there was a 1 in 10,000 chance of the Mir being destroyed in any given hour, then the probability that it lasts more than 4.6 years before destructing is less than 1/4.

So another rule for calculating variance, and maybe the most important general one, is that variance is additive. That is, the variance of a sum is the sum of the variances. But unlike expectation, where there's no other side condition, and it does not in any way depend on independence, it turns out that variance is additive only if the variables being added our pairwise independent.

Now you might wonder where the pairwise came from, and it's because variance is the square of an expectation. So when you wind up multiplying out and doing the algebra, you're just getting quadratic terms for variances of-- for expectations of ri times rj. And so you need to factor those into expectation of r times expect-- ri times expectation of rj, which you only need pairwise independence for.

So that's a fast talking through the algebra that I'm going to leave to you. It's in the text, and it's again one of these easy proofs.