[SQUEAKING]

[RUSTLING]

[CLICKING]

BRYNMOR CHAPMAN: Hello? Can people hear me in the back? Are we good? So welcome to the last lecture of 604-- no, 6.1200. I'm totally sane. I can remember which class I'm teaching. Welcome. You made it through the semester. Well done.

So for today's material, we are going to wrap up our discussion of probability. And in particular, we will be discussing tail bounds. So what happens if you have a random variable? What is the probability that this random variable is very large or very small, very different from its expectation?

So let's start with a bit of a review. Variance. So I think we saw from the last-- not last lecture, two lectures ago, possibly, so the variance of a random variable R is defined to be the expectation of R minus the expectation of R squared. And the standard deviation is just the square root of that. If you've taken stats or anything like that, that's probably something that you've seen more often than variance.

Another, not a definition, but another useful formula, an equivalent formulation of this, is that the variance of R is also equal to the expectation of R squared minus the square of the expectation of R. So not the definition, but it's equivalent. And I think in recitation perhaps, you proved this. This formula is a particularly useful one for what we're going to be doing today.

We also saw that variance is linear, subject to certain constraints. Does anybody remember what you need for variance to be linear? Yeah?

AUDIENCE: Independence.

BRYNMOR CHAPMAN: Independence, OK. So if you have R1, R2, Rn are independent, then the variance of the sum is the sum of the variances-- oops. Is that the right constraint? Is it correct? Who believes the statement? Who disbelieves? Who's still asleep? Oh, OK, maybe we'll give it a few minutes for people to wake up.

Well, this is a correct statement. However, there is something slightly stronger that we can say. We can weaken this condition. We don't need mutual independence. What do we actually need? Pairwise independence, yeah. So if we have a bunch of pairwise independent random variables, we want to find the variance of their sum. We can instead just sum the variances.

So another very useful fact-- hopefully, you proved it already in recitation. So we don't need to go over the proof. But these-- well, these two formulas are going to be quite useful today.

Now, time for the first of our tail bounds. So before I give you the bound, let's think about an example. So looking at this room, suppose I pick a random student. And let the random variable R be the row that the randomly picked student is sitting in. So guesstimate that this is about-- the expectation of R is about 8.

So pick random student. R is the row. So for those of you who aren't in the room, those of you who are watching the video, what does that tell them about the probability that R is at least 16?

So what's the probability that a randomly selected student is sitting in the last five rows? I think there are 21 rows here. Yeah. Does anybody have any idea? So maybe I should label this as an example. Yeah?

AUDIENCE: 1 minus the CDF [INAUDIBLE].

BRYNMOR CHAPMAN: 1 minus the CDF. So the people watching the video, I assume they can't see the room. So they don't know what the CDF is. The only information they know is that the expectation is 8. Yeah?

AUDIENCE: Can we find the variance for that and see how far away you are from there?

BRYNMOR CHAPMAN: Oh, can we find the variance from that and then see how far away we are? No. [LAUGHS] That requires more information, right? We need to know a little bit more about what's going on. We need to somehow figure out the expectation of R minus 8 squared. So, we need to know something more about R than just its expectation. Yeah?

AUDIENCE: Do you know how many [INAUDIBLE]?

BRYNMOR CHAPMAN: Yes, there are 21, so 0 to 20.

AUDIENCE: Wait. Can we assume that it's uniformly distributed and [INAUDIBLE].

BRYNMOR CHAPMAN: Can we assume that it's uniformly distributed? No. And in fact, it is not. There seem to be more students towards the front than towards the back. Yeah?

AUDIENCE: [INAUDIBLE] that there are 20 rows. If you just think there are 16 row, we know that the way we find the expectec value is multiplying each possible outcome by its probability. Since a is 60 multiplied by 1/2, the probability is generally [INAUDIBLE] to be less than 1/2.

BRYNMOR CHAPMAN: Yeah.

AUDIENCE: If you factor in 16 or 20, it gets to be more fun math [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, so your intuition is great, though. So, well, we're using the observation that 16 is twice 8. We're also using the observation that the rows are numbered 0 through 20. Like they're non-negative. So what happens if more than half of the students are sitting in the back five rows? What does that tell us about the expectation?

Well, if more than half of the students give us R is at least 16, that tells us that the expectation should be at least 8-- or sorry, more than 8, if more than half of the students are in the back. So the converse of that is that there are at most, half of the students in the back five rows.

So let's try and formalize this intuition. So this is what's known as Markov's inequality. So let R be a non-negative random variable. 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.

So for the example that we just saw, the probability that R is at least 16, it's at most the expectation, which is 8 divided by 16. So that gives us 1/2. Does the statement make sense to people?

So basically, it's saying that you can't have too many super large values, because that would push the expectation up too high So how could we prove this? And it is, how do we formalize the intuition that we were just talking about?

Let's try using the law of total expectation. So how can we write the expectation of R? What might we condition on? We don't have too many events that we can work with. Which of them might be useful? Nobody?

So let's try conditioning on x greater than-- oh wait, x-- yeah, or greater than or equal to x. So what does the law of total expectation tell us? Well, this is going to be the expectation of R conditioned on R being greater than x multiplied by the probability that R is greater than x plus the expectation R conditioned on R less than x times that probability. Do people remember that? Is that coming back? Hopefully. Yes. No Maybe. A few nods.

So what do we know about each of these values? Let's start with this one. Can anybody bound this? Yeah?

AUDIENCE: It's at least x.

BRYNMOR CHAPMAN: Yeah. It's at least x. If you conditioned on R being at least x, the expectation of R is also going to be at least x. Every outcome is at least x. So this is going to be at least x times-- well, this is the probability that we're looking for. So let's just write it out as it is.

Now, what about the bottom line? What do we know about this expectation? Yeah?

AUDIENCE: Is it 1 minus the expectation of the other one?

BRYNMOR CHAPMAN: Is it 1 minus the expectation of the other one? No. So the probability is 1 minus the other probability. The expectation does not work like that, though. Yeah?

AUDIENCE: At most, x.

BRYNMOR CHAPMAN: It is at most x. That is true. We're actually looking for a lower bound, though. Yeah?

AUDIENCE: At least 0.

BRYNMOR CHAPMAN: It's at least 0, yeah. We know that R is always non-negative. So conditioning on anything, R is still always going to be non-negative. So its expectation has to be non-negative. So we can drop this. This is just going to be 0.

Now we've got the expectation of R is at least x times the probability that R is at least x. Now we can just move this x to the other side, assuming x is non-negative. If x is negative, then it's a very silly question-- or sorry, assuming x is positive. Are people happy with this proof? Does anybody have any questions?

So another form of Markov which you may often encounter is the following. So the probability that R is at least C times its expectation is at most 1 over C for any positive C-- so, essentially, equivalent statement, just you might sometimes see it in this form instead.

So let's take a look at a couple more examples. So suppose R is the test score of a randomly chosen student instead of the roll that they're sitting in. And suppose that these are all in the range, say, 30% to 100%.

Can we use Markov to bound the probability that R is at least, say, 90%? How might we do that? Sorry, I completely forgot to give you the expectation. The answer is no. OK. Now can you do it? Hopefully, it's a little bit easier now. Mm-hmm?

AUDIENCE: What if [INAUDIBLE] value of C such that C times 75 is 90.

BRYNMOR CHAPMAN: Yeah. So find the value of C such that C times 75 is 90. So I guess 75-- or 90 divided by 75, that's, what's, 1.2. Is that right? 75 is 5 by 15, 96 by 15. OK, nice, arithmetic, score.

So then using this formulation here, we have the probability that R is at least 1.2 times expectation of R is at most 1 divided by 1.2. So this is going to be something like 83%. People reasonably happy with that? Yeah?

AUDIENCE: Can you please expand a little bit [INAUDIBLE]?

BRYNMOR CHAPMAN: Oh, the question is, can we explain the difference between these two formulations?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Sorry?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: So yeah, essentially, there is no difference. They're just different ways of thinking about the bound. You can either think about bounding it in terms of some absolute threshold. Or you can think about it as in terms of some relative threshold, multiply the expectation by some scaling factor.

So, depending on which of them feels more natural to you, we could also use the first formulation to say, probability that R is greater than 90 greater than or equal to is going to be at most, 75 over 90, which same thing, 83%. So either of them works. They both give the same answer. They're equivalent formulations. It's just whichever one you find easier to think about. Question?

AUDIENCE: Based on the fact that we know the range is 30 to 100, not 0 to 100, is there a way to make the final probability more accurate?

BRYNMOR CHAPMAN: Yes, exactly. So the question was, we don't just have a non-negative random variable R. Right. We've actually got an even stronger bound. R is at least 30, not just at least 0. So can we leverage that to get an even stronger bound? So how could we do that?

Well, let's try defining a different random variable. So let R prime be R negative 30. So this is another random variable. We're just going to take our original R. We're going to subtract 30. We're not going to do this to people's exam scores in reality. Don't worry. But that this makes the minimum exam score 0.

So the hope is that now we're satisfying the condition to Markov's inequality more tightly. So hopefully, Markov will give us a tighter bound. So what bound does it give us now? What is the probability that R is at least 90? Well, what is that in terms of R prime. What is that event? Yeah?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah. So this is the same event, and so the same probability that r prime is at least 60. Now, what happens when we apply Markov to this? Well, this is going to be at most, the expectation of R prime divided by 60. What is the expectation of R prime? Yeah?

AUDIENCE: 45.

BRYNMOR CHAPMAN: 45, yeah. So this is going to be the expectation of R minus 30 divided by 60, which is the expectation. So linearity of expectation, we can pull this-- or we can split this into two expectations. So this is the expectation of R minus the expectation of 30. There's not really any randomness involved there. It's just 30 over 60. So this is 45 over 60. Does that make sense to everybody?

So now, instead of this bound of it's at most 83%, now we've got the better bound. It's at most 75%. So by shifting our random variable, we've managed to get a tighter bound.

What about the probability that R is at most 65? Yeah?

AUDIENCE: For the previous one, where R is from 30 to 100, is that only the [INAUDIBLE] assumed that the lower bound for x is 0, and that somehow, the lower bound for R is 30. So I [INAUDIBLE] do a more complicated thing.

BRYNMOR CHAPMAN: So that's kind of what we're doing. The question was, in the proof of Markov, we're assuming that the lower bound is 0. Here we've got a lower bound of 30 instead. So do we need to do something more complicated? The answer is no, we don't have to. But if we do, it gives us a stronger result. So that's what we've done here.

But we don't have to do it. Because 0 is still a valid lower bound. We could say that the co-domain of R is 0 to 100 instead of 30 to 100, even if the range stays unchanged. So it's still a non-negative random variable. So we're still allowed to apply Markov to it. It's just that doing it on R instead of R prime gives us a weaker result. Does that make sense to everybody? Another question?

AUDIENCE: [INAUDIBLE] you mean use a tighter bound when you shift on. Does a tighter bound mean that the probability is lower or higher?

BRYNMOR CHAPMAN: So the question is, what do I mean by a tighter bound when we shift R? So tighter in the sense that it is stronger. So R at least 30 implies that R is at least 0. So R at least 30 is a stronger condition than R at least 0. So that's what I mean by tighter.

AUDIENCE: So what's the difference between the 75% that we got here for R is greater than or equal to 90 versus the 83%?

BRYNMOR CHAPMAN: So the difference is how we're applying Markov. In the first case, we're applying Markov to R directly. In the second case, we are creating a new random variable. And we're applying Markov to that. And because we've got a stronger condition, Markov's-- what do you call it-- precondition is tight now, we can get a stronger result. Yeah?

AUDIENCE: I also have a question. So does that mean it's depending on the situation to [INAUDIBLE] it could go higher than 83%?

BRYNMOR CHAPMAN: Could it go higher than 83%?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: So we will get to that in a moment. The question is, if we're in a different situation, could shifting it give us something that's higher than 83%? And the answer is yes-ish. In the case where we've applied Markov incorrectly the first time, then yes. But if we've applied Markov correctly, then the first is a valid bound.

So we could, in principle, improve it. But we're not going to do any worse by imposing tighter conditions. Does that at least kind of intuitively make sense? I'll give you a more complete answer in a moment. Any more questions about this before we move on to the next example? Yeah?

AUDIENCE: So it's a good thing that the 75% is lower than the 83%?

BRYNMOR CHAPMAN: Yes, it is a good thing. So the question is, is it a good thing that the 75% is smaller than the 83%? Yes. We are looking for an upper bound on the probability. So a smaller upper bound is a stronger condition. So that's better. It's a stronger bound.

AUDIENCE: We want a smaller upper bound or a stronger [INAUDIBLE]?

BRYNMOR CHAPMAN: So they're the same. Smaller upper bounds are stronger upper bounds. Yeah. So less than or equal to 83% implies less than or equal to-- sorry, less than or equal to 75% implies less than or equal to 83%. So 75 is stronger.

Have people had enough time to think about the probability that R is at most 65? How might we use Markov to figure this one out? Yeah?

AUDIENCE: Can you do 1 minus probability that R is greater than 65. So [INAUDIBLE].

BRYNMOR CHAPMAN: So R greater than or equal to 66, so if we assume integer scores. OK. And what does Markov tell us about this? Well, the probability that R is at least 66 is going to be at most 75 over 66, which is bigger than 1.

So applying Markov does not tell us anything useful. It tells us that this is a probability that's bounded by 1, which we already knew. So it is a good idea. It's not quite what we want here.

So what if instead of fiddling with the event, we fiddle with R again, kind of similar to what we did before? Yeah?

AUDIENCE: [INAUDIBLE] that R [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, exactly. So we're going to flip R. So we're going to define a new random variable. Let's say S equals 100 minus R. So now the probability that R is at most 65 is the probability that S is at least, what, 100 minus 65? 35. OK.

Can we apply Markov to that? I'm seeing some nods. Does anybody want to do that and spare me the trouble of doing arithmetic? Nobody? Sadness.

Well, what does Markov say? Well, it's at most the expectation of x-- oh, sorry S-- divided by 35. What is the expectation of S? Well, using linearity in the same way that we did a moment ago, this is going to be equal to, what, 100 minus the expectation of R, which is 75, over 35. So that's 25 over 35. This is 5/7. So does that make sense to everybody? There is at most, a 5/7 chance that you did worse than 65.

Now, perhaps not terribly realistic in terms of exam scores, but what if, say, some of the students did so badly that they got scores that are less than 0? Maybe you've got R is in the range minus 30 to 100. Same deal, expectation of R Is 75. A question, or-- sorry.

So what can we say about the probability that R is at least, say, 90? Mm-hmm?

AUDIENCE: So Markov [INAUDIBLE]?

BRYNMOR CHAPMAN: Yeah, exactly. So the observation was if we try to apply Markov to this-- maybe I should call it something different. Let's call it T. So if we try to apply Markov to T, it's going to be the same thing as when we applied Markov to R before. It will give us a bound of 83%.

The problem is that that's wrong. T does not satisfy the precondition for Markov. So what we want to do instead is define T prime equals T plus 30. And now if we apply Markov to T prime, what do we get? So the probability that T is greater than 90 is the probability that T prime is at least 120.

And by Markov, this is going to be expectation of T prime over 120. And same deal as before, linearity here gives us 75 plus 30 over 120. 105 over 120, that's 7/8. Is that right? Yeah.

So in this case, we do get, in some sense, a weaker bound than we would have gotten by applying Markov to T directly. But applying Markov to T directly doesn't actually work. This looks weaker, but it's a correct bound. So hopefully, that answers your question more fully.

Does anybody have any questions? So even though we couldn't apply Markov directly to T, we can create a new random variable T prime, to which Markov does apply. Yeah?

AUDIENCE: Got a question before about 30 being a lower bound to make our statement any stronger for S [INAUDIBLE].

BRYNMOR CHAPMAN: So using the lower bound of 30 making the statement stronger for T?

AUDIENCE: For S.

BRYNMOR CHAPMAN: Oh, for S. Which one was S. Ah, S here? So that does not really help us in this case. Because that corresponds to an upper bound rather than a lower bound. And upper bounds don't really appear in the proof of Markov. We're not really using them in any way. So imposing a stronger upper bound doesn't really help. Does that answer your question? OK, cool. Yeah?

AUDIENCE: Is that always the case that the lowest lower bound appears [INAUDIBLE] lowest value here includes 0?

BRYNMOR CHAPMAN: Is it always the case that the best lower bound appears-- or sorry, the best upper bound, I guess, appears when the range actually includes 0? Essentially, yes, as far as correct bounds go, anyway.

So I guess there's a question of morally, why do we need this non-negativity? And perhaps a very simple illustrative example is, suppose we have the random variable R, which is plus or minus 1, depending on a fair coin flip. So it's uniform on positive 1 and negative 1.

So if we try to apply Markov to that directly, what does that tell us about the probability that R is greater than 1/2? Well, if we try to apply Markov directly, the expectation of R is 0. So no matter what we put here, Markov would try to tell us that the probability is at most, 0.

That is clearly rubbish. Because R can be greater than 1/2. R takes value 1 with probability 1/2. So basically, in our proof-- where was it? Yeah. So if we get rid of the non-negativity constraint, then this no longer works. We can't discard this term, because that 0, if we don't know that the random variable is non-negative, we could, in principle, have a negative value there, which we can't get rid of.

So another question, is Markov tight? So, in other words, can we make a stronger statement than what Markov tells us out of the 10, just with the information that we have the mean and, we have a non-negative random variable? Is there anything stronger that we can say?

Who thinks yes? A couple of tentative hands. Who thinks no? A couple of less tentative hands. Who has no idea? Well, let's think about some simple examples then.

So let's go back to the cell phone check problem from, what was it, last week? Don't remember which of the lectures last week. Do you say Thursday? Both? Both, OK.

So if we consider the Lazy Susan version, and we let R be the number of people who get their phone-- so this is the Lazy Susan How could we use Markov to bound the probability that everybody gets their phone back? So if everybody gets their phone back, that's R equals n. What does Markov tell us about this probability? Yeah?

AUDIENCE: At most, 1 over n.

BRYNMOR CHAPMAN: Yeah, it's at most, 1 over n. Why is that? Well, the expectation of R is 1, as we saw last week. The bound we're looking at is n. So there's at most, a 1 in n probability that everybody gets their phone back. What is the actual probability that everybody gets their phone back? Yeah?

AUDIENCE: 1 over n.

BRYNMOR CHAPMAN: Yeah. The actual probability is 1 over n. So this is tight. We can't say anything stronger than Markov without additional information. Does that make sense? So if we know something else about our distribution, then we might be able to say something more. But if all we know is the expectation and the non-negativity, then Markov is the best that we can do.

On the other hand, suppose we're permuting all of the phones instead of putting them on the Lazy Susan. So in this case, what does Markov tell us about the probability that everybody gets their phone back?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Well, it's still 1 over n. We've still got a non-negative random variable with the same mean. We're still looking at the same threshold. So Markov gives us the same bound. What is the actual probability that everybody gets their phone back?

Well, that's exactly when you have the identity permutation. You've got n factorial possible permutations. One of them gives everybody's phone back. So what we're looking for is actually 1 over n factorial. So Markov is very much not tight in this example.

So the next question is, what can we say if we do have more information about a random variable? So, suppose we know the variance. So the variance of those two distributions is very different. In the case where we put the phones on the Lazy Susan, we've got a very high variance. In the case where we have a random permutation, it's much lower.

So maybe the variance can tell us more about these probabilities. So this is a different tail bound, due to Chebyshev. So probability that R minus the expectation of R, an absolute value is at least x, is less than or equal to the variance of x divided by x squared. So note that we no longer require non-negativity here. We just require information about the variance And, well, we have to compute it relative to the mean.

But R can be any random variable. We're no longer constrained to looking at non-negative ones. It can be anything. We just have to compute the variance. How might we prove Chebyshev's inequality? Any ideas? Is that a hand or a stretch? Yeah?

AUDIENCE: [INAUDIBLE] Markov?

BRYNMOR CHAPMAN: Yeah. So we're going to prove Chebyshev using Markov. So in order to use Markov, what do we need? Well, we need a non-negative random variable.

So similar to what we were doing earlier with shifting our random variable, we're going to define a new random variable that we can then apply Markov to. And that should give us this result here. Does that make sense to everybody? Yeah?

AUDIENCE: For a [INAUDIBLE] is variance of x supposed to be R?

BRYNMOR CHAPMAN: Oh, variance of R-- sorry. Thank you. Yes, variance of x does not make any sense. So what might be a useful random variable that we can apply Markov to? Well, maybe here's the way we can cheat.

In Markov's inequality, what's the bound that we get? The bound should be the expectation of a random variable. So can we define a random variable whose expectation is the variance of R? Yeah?

AUDIENCE: [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, exactly. So R prime equals R minus its expectation squared. So this is non-negative. It's a square of some value. Its expectation by definition is the variance of R.

And also, if we squint, we can fiddle around with this to actually threshold R prime instead of this absolute value thing that we've got here. How do we do that? Well, we just square the stuff in the absolute value. It has the same square as if we didn't have the absolute value.

So the probability that R minus its expectation greater than or equal to x equals the probability that R prime is greater than or equal to x squared-- because those are exactly the same event. They're the same event, so they have the same probability. And now this looks like something we can apply Markov, less than or equal to the expectation of R prime divided by x squared. Does that make sense to everybody?

So let's take a look at an example. So let's look at those test scores again. Oops, sorry, my bad. So R, same deal, we've got test scores. Expectation is-- what did we have? 70, 75? 75. And say that the variance. Is 25.

So now what can we say about the probability that R is at most, 65? Any ideas? How could we apply Chebyshev to this? That doesn't quite look like the theorem statement, right?

But we can observe that this is at most, the probability that R is at most, 65, or R is greater than or equal to 85. If we take the union of two events, that union has at least the probability of one of those two constituent events. So this is only increasing the probability, if we add in this extra half of the event.

And what do we know about this probability? How could we rewrite that in a nicer way? Yeah?

AUDIENCE: Probability of actually having [INAUDIBLE] expectation is greater than or equal to 10.

BRYNMOR CHAPMAN: Exactly. So this is just saying R deviates from its mean by at least 10. So now we've got it in the form that Chebyshev likes. So we can apply Chebyshev. This is at most, variance of R divided by-- R threshold is 10? Yeah, 10 squared. 25 over 100 equals 1/4. Now, we've got a much stronger bound this time than we did earlier. Yeah, question?

AUDIENCE: Can you explain one more time how you got the probability of R is that [INAUDIBLE] 65 or greater than 85 to the next probability?

BRYNMOR CHAPMAN: Ah, OK, so explain this equality?

AUDIENCE: Yes.

BRYNMOR CHAPMAN: So R less than or equal to 65 or R greater than or equal to 85, we could rewrite this as R minus-- well, R minus 75 is less than or equal to negative 10. And this one is R minus 75 greater than or equal to positive 10. So if you absolute value both of these, they're both greater than or equal to 10. Does that make sense? OK. And other questions about this example?

So because we have more information than we had here, we can get a stronger bound. Knowing the variance helps us out. So equivalently, we could rephrase this in terms of standard deviations. Where would be a good place to write this. How about here?

So the probability of being at least two standard deviations from the mean is at most, 1/4. Does anybody have any objections to that last statement? So if you've taken a stats class, you probably heard something very different from that. No? Nobody's taking stats? That's good. High school stats is terrible. Oh, that's on camera. Never mind. Stats are great. it's good that you've saved it for university instead of doing it in high school.

So if you're looking at a normal distribution, you actually have a much better bound than 1/4. It's something like a few percent. So why is that. Any ideas?

Well, for a normal distribution, we know a lot more about the random variable than just its variance. So if we know even more about a random variable than its variance, it turns out we can make even stronger statements. Just like adding in the variance was an upgrade from Markov, let's see how we can upgrade this again.

AUDIENCE: [INAUDIBLE]?

BRYNMOR CHAPMAN: Yeah, what's up?

AUDIENCE: What's a normal distribution?

BRYNMOR CHAPMAN: Oh, what's a normal distribution? Think of a normal distribution as numb er of heads, if you flip a whole bunch of coins. That's not quite it, but roughly speaking-- good catch. So as I just said, if you flip a whole bunch of coins, let your random variable R be the number of heads. You know a lot more about this random variable than just its variance.

You can decompose your random variable into a sum of a whole bunch of indicators. And all of those indicators are mutually independent. In order to use Chebyshev, you don't need mutual independence. If you had something of that form, you just need pairwise independence. You need to compute the variance.

But if, on top of that, you have mutual independence, it turns out you can do a lot better. So suppose T1, T2, T3, up to Tn are mutually independent random variables. And, well, let's say, in addition, they are bounded. And the random variable we actually care about is T, which is their sum.

Oh, and say we've also got some threshold C, which is at least 1. The probability that T is at least C times its expectation-- I'll close that as well-- less than or equal to e to the negative C log of C minus C plus 1 times the expectation of T. So a little bit of a mouthful, but what exactly is going on here?

So basically, Markov was saying that the probability that you deviate from the mean by a large amount, it's at most linear in how far you're deviating. Chebyshev said it's at most, quadratic. This is saying it's at most, exponential, or inverse exponential. So you have a much stronger condition that gives you a much stronger bound.

And how could we prove it? Not like this.

Any ideas? Well, maybe a proof sketch-- we won't go through the entire proof.

So what tools do we have already? What might we be able to do?

Well, what happens if we apply Markov? How can we apply Markov? In order to use Markov, we need a non-negative random variable. So in order to prove Chebyshev, we took a square that made something non-negative. What might be a reasonable thing to try in this case? Yeah?

AUDIENCE: Add 1.

BRYNMOR CHAPMAN: Add 1. OK, so look at T plus 1 instead of T? So that's actually not going to help us. T is already the sum of a bunch of non-negative random variables. So it's already non-negative.

We're not looking to turn a non-negative-- or sorry, turn an arbitrary random variable into a non-negative one. We are trying to find a more useful non-negative random variable. Yeah? Yeah?

STUDENT: Log of both sides?

BRYNMOR CHAPMAN: Take the log of both sides-- close.

What happens if we exponentiate? So take T prime equals C to the T.

What do we know about T prime?

Well, we're taking something that's at least 1, right? We're raising it to some positive power-- well, some non-negative power. So T prime is a non-negative random variable. Do we know anything more about it? Yeah?

AUDIENCE: It's at least 1.

BRYNMOR CHAPMAN: It's at least 1, yeah. So maybe we try applying Markov to this. So now we've got a random variable that takes values in the range from 0 to C to the n, C to the n minus 1.

Now, how could we rewrite this condition, T greater than or equal to C times expectation of T in terms of T prime?

Well, because C is at least 1, exponentiation is monotone. So this is the same as C to the T greater than or equal to C to the C times expectation of T. And we don't have too much time. So I think I will refer you to the book, if you are interested.

But the main idea is this step here. You take the random variable you care about, and you exponentiate it. And then if you apply Markov to that, it actually gives you a much stronger bound. If you're using the mutual independence of all of the component random variables, this will give you a much stronger bound than even Chernoff.

So I promised you last week that we would revisit the coin tosses. So let's do that briefly. So suppose R is the number of heads if you flip n coins. So if you recall, we were looking at the probability that R is equal to n over 4 before. Or maybe equivalently, 3n over 4 might be a better way to analogize.

And I promised you that we would revisit this probability, R greater than or equal to 3n over 4. Now, what happens if we apply each of our three tail bounds to R? Well, first, what happens with Markov?

So let's call this p. p is at most, what?

Well, the probability that R is at least some threshold, it's going to be the expectation of R divided by this threshold. What's the expectation of R? n over 2. So this is going to be 2/3.

What about Chebyshev? So, similar to what we did with the first Chebyshev example, p is going to be at most, the probability that R differs from its expectation in either direction by another 4. So what does that give us? Probability R minus, let's say, n over 2 greater than or equal to n over 4 is at most var of r divided by n over 4 squared.

So what is the variance of R? Does anybody remember? We've got a sum of a bunch of indicators. They're all mutually independent. So the variance is linear. What's the variance of a single indicator? Nobody? Yeah?

AUDIENCE: It's n over 4.

BRYNMOR CHAPMAN: Yeah, n over 4-- well, for R, not for a single one. So n over 4 divided by n squared over 16 equals 4 over n. So, much better, much better than the 2/3 we just got.

What happens if we use Chernoff?

Any ideas? Yeah?

AUDIENCE: Plug in 3/2 for C.

BRYNMOR CHAPMAN: Plug in 3/2 for C. So C equals 3/2. OK. Why don't I put Chernoff back up here. and what does that give us? Probability that R is at least 3/2 times expectation of R is less than or equal to e to the minus C log C minus C plus 1.

So this is a constant times the expectation of R. So the expectation of R is n over 2. This is a constant. So we've got something that's exponential, or inverse exponential, in n-- oops-- e to the minus order of n.

So, Markov gave us a really bad bound; Chebyshev, linear-- pretty reasonable. Chernoff gives us an exponential. So that's all we have time for today. Sorry, I went a little bit over time.

It's been a pleasure teaching you all this semester. And let's see, we have the rest of our instructors and at least one TA. Hello, Lily. If anything went wrong, it was probably my fault. Anything that went right, you have your TAs to thank for it, including Lily. Lily's amazing.

Perhaps I will see some of you next term in 6.1210. Good luck on the final. Make sure you take the right one.

ERIK DEMAINE: Thanks for being students in this class. It's been lots of fun.

BRYNMOR CHAPMAN: Yeah.

[APPLAUSE]