[SQUEAKING]

[RUSTLING]

[CLICKING]

BRYNMOR CHAPMAN: Hello out there. Can people hear me? Oh, nice one. People at least can intuitive what it means when I wave at them.

So today, we are going to continue our discussion of probability. And in particular, we're going to talk about a very naughty word, what we call "random variables." Unfortunately, a random variable is neither random nor a variable.

It's just called this for historical reasons. It's very silly. I'm just going to be referring to them as RVs through the lecture. This is what it stands for. But it's a very silly name. It's a bit of a misnomer.

So what is an RV? So basically, what an RV is, it's neither random nor a variable. It is a function from our sample space to, say, the real numbers. That's all it is. It's just a function.

You can have-- the codomain R is not particularly important. You could just as easily have codomain N. You could also have complex-valued RVs. But we won't be looking at those in this class. For now, you can just think of it real valued function whose domain is the sample space.

So for instance, suppose our sample space-- or suppose our experiment is we're flipping three coins. So we're going to assume for now that they are fair. They're normal physical coins. They're independent. If you flip one, it's not going to affect the outcome of any of the others.

So what is our sample space in this case? What does our sample space look like? Am I giving the right lecture? This is 6 1200, right? OK.

If we flip three coins, what are the possible outcomes? Yeah?

AUDIENCE: Each of them being heads or tails.

BRYNMOR CHAPMAN: Yeah, each of them being heads or tails. So if we flip three coins, our sample space is going to be-- so each coin can take either-- it can be either heads or tails. There are three of them. So we're going to be considering ordered triples of either heads or tails.

And maybe we've got this function f that maps an outcome omega to the number of heads in omega. So that's one possible RV. We're taking an outcome, and we're mapping it to some number, the number of heads.

Another one could be maybe instead, we just look at is the first coin flip heads or tails? Oh, maybe I should not write it as that just yet. So 1, if and only if first flip is heads, and it's 0 otherwise.

So these are two examples of RVs. They are functions that map outcomes to real numbers. Does this make sense to everybody?

We could have one more that's h of omega is 1 if all three coins match, 0 otherwise. So these are just examples of RVs that we could have.

So g and h in particular have a special property. They are either 1 or 0. So this is what we call an indicator. So all that an indicator is, is just an RV whose codomain is 0 and 1. That's all it is. It's just a binary RV.

And RVs, and in particular, indicator RVs, naturally give rise to events. Does anybody see how? Yeah?

AUDIENCE: You can define a subset of outcomes where the indicator RV is 1, and that is equivalent to an event, and it's 0 when it's not in that event.

BRYNMOR CHAPMAN: Yeah, exactly. So the answer was-- now, in particular, if you have an indicator RV, you can look at the set of outcomes where it's 1. That's an event. Or its complement is the set of outcomes where it's 0.

So more generally, if we have an RV f and a value x, we can define the event. So slightly weird notation here-- f equals x by the set of outcomes omega, where f of omega equals x.

Or you may also have seen the notation f inverse of x, inverse here meaning inverse as a relation, not as a bijection. But if you haven't seen this notation, don't worry about it. This is what we call it in terms of probability.

And vice versa-- if you have an event, you can define the indicator of that event. So if we have an event A, corresponds to the event, this is the notation I was trying to use before, blackboard bold 1 of A. This notation here just means the indicator that is 1 precisely for the outcomes in A, so if omega is in A, and 0 otherwise.

Does that make sense? Does this correspondence make sense? Yeah, question?

AUDIENCE: [INAUDIBLE]

BRYNMOR CHAPMAN: Explain the notation again, OK. Yeah, so basically, we are taking an event and using it to define an indicator RV. This notation is just how we write it. This is called the indicator of A.

And all it means is just the indicator RV, which takes the value 1 on all of the events in A, and it takes the value 0 on all of the events that-- or sorry, all of the outcomes that aren't an A. Does that make sense? Cool. Any other questions? OK.

Or actually, maybe we should keep that visible for now. We can also define a different kind of event, f greater than or equal to x. So that one obviously doesn't make sense for complex numbers. But for real numbers or any subset of the reals, this is perfectly fine.

As you might expect, this is going to be the event that contains all of the outcomes omega, such that f of omega is greater than or equal to x, so analogous to what we just defined. And more generally, we could say, f in some set S-- or maybe T, because S is our sample space, this is going to be omega such that f of omega is in T. Does that make sense to everybody?

So as you might expect, we can condition on these events, just as we can condition on any other events. We can also have independence of random variables, just as we have independence of events. Sorry.

So, for instance, now we could talk about the probability that-- what did we call them-- f equals 2, conditioned on h equals 1. So what is this probability? What is it saying?

Yeah? Oh, is that a hand or just-- OK, sorry. Yeah?

AUDIENCE: That the probability [INAUDIBLE] given that all the [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, exactly. So we're looking at the probability that we have exactly two heads, conditioned on all three coins being the same. So we could think of these as two events, A and B. If we actually write out what this means, this is the probability that f equals 2 and h equals 1, divided by probability that h equals 1. And this probability in particular is zero.

So if all three coins are the same, it's impossible for exactly two of them to be heads. Does that make sense to everybody? We're basically doing exactly the same thing with these weirdly written events as we did with any other events.

And we can also extend this to independence. But I should point out here that things are going to diverge a little bit. So we're going to say that two RVs are independent if and only if for all possible x and y, the probability that f equals x and g equals y equals the probability that f equals x times the probability that g equals y.

So it's similar to what we had for events. How is it different? Well, for events, we didn't have this quantifier. We just said that two events are independent if the probability of their intersection is the same as the product of their probabilities.

Now we have to quantify everything. It's not just enough to say the event that each of them takes value 0 is independent. We kind of need that for all possible pairs of values that they could take. Does that at least intuitively make sense? Yes, no? Some nods, tentative nods, at least. OK.

Or just as with normal events, we could equivalently say, for all x, y, either probability that g equals y equals 0-- whoops, not of, or-- probability that f equals x, conditioned on g equals y, which is the same as the unconditional probability. So if you prefer, you could think of it in terms of conditional probability.

So we're basically saying that two random-- two RVs are independent if and only if all of the events that they generate-- the events f equals x and g equals y-- those are all independent events. And intuitively, it's saying that if we know something about the value that g takes, for instance, it doesn't tell us anything about what value f takes.

So example-- where were they-- f, g, and h, are those independent? So are f and g independent? Who thinks f and g, as I've defined them over there, are independent? Show of hands? Yeah?

Who thinks they're not independent? A few hands. Who's still fast asleep? Oh, that's unfortunate. I'm sorry. Me, too.

OK, so does anybody want to give me an explanation one way or the other? Presumably, somebody who is not asleep. I don't need an explanation for why you're asleep. Anybody else? No? Yeah?

AUDIENCE: I'm not sure, but maybe if g is 0, then f can be--

BRYNMOR CHAPMAN: Yeah, exactly. So if g is 0, that tells us something about the value that f can take. It cannot be 3. So the events g is 0 and f is 3 are not independent events. So the RVs are not independent.

Or perhaps even more obvious, if f is 0, that tells us exactly what g is. g must also be 0.

What about f and h? So f is the number of heads. h is 1 if and only if all three coins match. Who thinks those are independent? No?

Who thinks they're not? A few hands? OK, why? Yeah?

AUDIENCE: [INAUDIBLE]

BRYNMOR CHAPMAN: Yeah.

AUDIENCE: [INAUDIBLE]

BRYNMOR CHAPMAN: Yeah. So f basically tells us everything we need to know about h. h is just a function of f. And, well, neither of them is trivial.

So if we know f, no matter what it is, we know what h is. So these are also not independent.

What about g and h? Who thinks g and h are independent? Any hands? Who thinks they're not? Oh, nobody. OK, why? Yeah?

AUDIENCE: [INAUDIBLE]

BRYNMOR CHAPMAN: Sorry?

AUDIENCE: The first one doesn't tell us anything about the other two.

BRYNMOR CHAPMAN: Yeah. Intuitively, that's what's going on. The first flip doesn't tell us anything about the other two. And, in particular, it doesn't tell us anything about whether the other two match each other and the first flip.

So we could put it all into here and figure it out. But I will leave that as an exercise to the reader. Basically, no matter what happens, you've got a 1 in 4 chance that the other two coins are going to match the first one. And that matches the probability that they all match to begin with.

OK, does anybody have any questions about independence before we move on? People reasonably happy?

So the next topic is distributions. So for those of you who have seen RVs before, this is probably the context-- or this is probably the context in which you've seen them. Historically, this is where they started, and more or less why they got their name.

But basically, the distribution is-- it's basically a fancy way of talking about the RV without the underlying sample space. It's talking about the probabilities that it takes each individual value in its range. So there are kind of two equivalent ways to talk about it, at least for real-valued RVs.

So for an RV f, we define the Probability Mass Function, often abbreviated as PMF, of omega-- oh, sorry, x. And this is just the probability that f equals x.

So does that definition make sense to everybody? The PMF of an RV is just talking about the probability that that RV takes each individual value in its range. So you could think of this as a probability space over R. So that's how RVs originated, talking about probability spaces over R.

Equivalently, you could also define the Cumulative Distribution Function, CDF, as CDF sub f x. This is just the probability that f is less than or equal to x, which is the sum y less than x of the probability that f equals x.

So I should warn you that if you see-- in many other probability classes, a PMF or something analogous to a PMF would often be called a PDF, for a Probability Density Function instead of Probability Mass Function. And I believe, in particular, the text does this. So for the purposes of our class, those are the same. If you see PDF, think to yourselves PMF.

We prefer this notation because this is what's used for discrete distributions. But if you have a continuous distribution like-- well, it's kind of hard to talk about the probability that a uniform random variable over the interval from 0 to 1 takes exactly a particular value. If you throw a dart at a dart board, you're never going to hit one particular point, essentially. You have null events, so it doesn't really make sense to talk about the probability that you hit a particular point, because that's always 0. So instead, you look at the density function instead.

But for discrete distributions, it does make sense to talk about the probability that you hit exactly one particular value. Does that make sense to everybody? So for this class, PMF and PDF are the same. If you see continuous probability, they won't be.

So, often, it is useful to talk about RVs in terms of their distributions, in terms of their PMFs or CDFs. And just kind of forget about the underlying sample space. So a couple of examples-- indicator-- indicators-- what does the PMF of an indicator look like?

Well, it can only take two possible values. So it's going to be 0 everywhere, except at 0 and 1. And at one of those, it will take the value p. At the other one, it'll take the value 1 minus p. And we call these-- we say that these have a Bernoulli distribution.

So PMFs basically look like this. You'll have two values. Everywhere else, it's 0. And the CDF is going to look like this. Oh, wait, no. It should be at 1.

Another useful one is a uniform random variable. So a uniform RV, it can take any value with equal probability. So in particular, let's look at an RV on the set up to n. So here, the PDF of x is going to be 1/n if and only if x is in this set, and 0 otherwise. And CDF of x, what is that going to be?

Well you're basically going to have a piecewise function. So it's going to be 0 up until 1. Then at 1, it'll go up to 1/n. At 2, it'll go up to 2/n, et cetera. So you get something that looks like that. Ah, yeah, question?

AUDIENCE: [INAUDIBLE]

BRYNMOR CHAPMAN: Oh, explain again how I got to this graph. Well, the CDF is-- you could think of it as a sum of the PDF-- or, sorry, sum of the PMF. So the CDF at 1, it's the probability that this uniform RV is no greater than 1. How can that happen?

Well, only if you take the value 1. That happens with probability 1/n. So at 1, we should get probability 1/n. Anything between 1 and 2 has that same probability. The CDF of 1 and 1/2, the probability that we're no greater than 1 and 1/2 is the same as the probability that we're no greater than 1 because we can't take any value between 1 and 1 and 1/2.

So all the way up to 2, it's going to be constant. And then at 2, it's going to jump up by another 1/n, because at that point, the probability that we're no greater than 2 is the probability that we take either of these two values, which happens with probability 2/n. So we get this piecewise step function that at every integer, it increases by 1/n. Does that make sense?

Is there any way that you can beat random chance? Is there any way that you can do better, like win with probability greater than 1/2? It's not really obvious.

So let's simplify the game a bit. So suppose I guarantee that one of the boxes has more than 5 candies, and the other one doesn't? [LAUGHS] Can you then beat random chance? I'm seeing some nods. Yeah, how? How do you beat random chance in that case? Yeah?

AUDIENCE: The boxes [INAUDIBLE] boxes [INAUDIBLE] less than 5 [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, exactly. You pick either box. You open it. If it has more than 5, that's the box with more. You stick with that.

If it doesn't have more than 5, well, the other one does. So the other one has more. So if you know that one of the boxes has more than 5, and the other one doesn't, then you can win with certainty. It's still not probabilistic. Well, I suppose you could think of it as probabilistic, but you can definitely do better than 1/2.

Does this give you any ideas as to how you might win with probability greater than 1/2 even if you don't know that one of the boxes has more than 5 candies, and the other one doesn't? Any ideas? Yeah?

AUDIENCE: [INAUDIBLE] anyways to [INAUDIBLE] better than 1/2?

BRYNMOR CHAPMAN: OK. So your colleague's proposal was, do it anyway. If you open up a box with less than 5 candies, you switch and hope that the other one has at least more than what you found. If you open up a box that has more than 5 candies, you stick with it and hope that the other one has fewer.

Does anybody see a problem with this? What is the probability of winning? Yeah?

AUDIENCE: Whoever gave us [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah. So let me try and rephrase that. I haven't given you any probabilities of what the probability is that I've put a certain number of candies in either box. So it kind of depends on what I've done.

There are two cases. If I have split them, like greater than 5 and less, then you win with certainty. If I haven't, then you win with probability exactly 1/2. If they're both greater than 5, you win if you picked the bigger one to begin with. And if they're both less than 5, then you win if you picked the smaller one to begin with.

So you're still not doing worse than 1/2. But you're not doing better, or you're not guaranteeing that you're doing better because I made no guarantees as to how I was splitting up the candies. I may put fewer than 5 in both of them. Or, as I did this time, I may put more than 5 in both.

But can you use a similar idea to get greater than 1/2 probability, no matter what I've done? What if instead, you know that I've put more than 2 candies in one box and not the other? Can you win in that case?

Anyone? If you know that I've put more than 2 candies in one box and not the other one, how do you win? It's not a trick question, I promise. [LAUGHS]

Just as what your colleague said with 5, you can open up one of the boxes and see whether or not it has more than 2 candies. If you know that one of the boxes has more than 2 candies and the other one doesn't, you can identify which is which by looking at one of them.

So why don't I start writing some of this down? So if one box-- so exactly one box has at least one candy, then we open up our box, we see whether or not it has any candies, and if it does, we take it. So we can win.

So if we know that two boxes-- sorry, if one box has greater than or equal to 2 candies. If we know this, we can win. If we know that one of them has greater than or equal to 3 candies, we can still win.

The problem is, we don't actually know any of these. So what can we do about that lack of knowledge? Yeah?

AUDIENCE: [INAUDIBLE]

BRYNMOR CHAPMAN: Yeah, exactly. So we don't know, but we can guess. One of these 10 things has to be true. I've told you that I've put different numbers of candies in each box.

So there is some divider. One of these-- well, at least one is true. Maybe I've put 0 in one and 10 in the other, in which case they're all true. But at least one of them has to be true. So if you randomly guess one of them and just say, pretend that you know it, and adopt the strategy we were just talking about, if you guessed correctly, then you win. And if you didn't guess correctly, as your colleague in the back pointed out, you still don't lose, at least. You still win with probability 1/2. Does that make sense to everybody?

So our strat is guess a threshold, and then assume it's true. So what exactly is the probability of winning if we do this? Well, it's going to be we can use the law of total probability. It's going to be the probability that we win conditioned on being correct, times the probability that we are actually correct, plus the probability that we win, conditioned on being wrong, times the probability that we're wrong. So what are each of these probabilities?

What's the probability that we win if we made the correct guess-- or a correct guess, rather? One. Yeah, exactly.

What's the probability that we made a correct guess? Yeah?

AUDIENCE: 1/10.

BRYNMOR CHAPMAN: 1/10. Well, it might be greater, but it's at least 1/10. So it's at least 1 times 1/10 plus, what's the probability that we win conditioned on being wrong? Yeah?

AUDIENCE: 1/2.

BRYNMOR CHAPMAN: 1/2, yeah. So if we guess wrong, say without loss of generality, our threshold is too low. We're going to open up a box, and we're going to stick with it regardless of what it is. So we win if and only if we happen to pick the bigger one. So we win with probability 1/2.

And what's the probability that we made the wrong guess? Yeah?

AUDIENCE: Is it 9/10?

BRYNMOR CHAPMAN: 9/10, yeah. So what is this going to be? This is going to be what, 55%? So it's not that much better than 1/2, but it is better.

So does that make sense to everybody? So we've-- oh, yeah?

AUDIENCE: [INAUDIBLE]

BRYNMOR CHAPMAN: Yeah, so we don't actually know that. So it's possible that I could have a large difference in the number of candies in each box. We do know that there's at least a 1/10 probability that we pick the right threshold. It could be more than that.

But if there are, say, two correct thresholds, then instead of this, you have 1 by 2/10 plus 1/2 by 8/10.

AUDIENCE: Why is 1/10 [INAUDIBLE]?

BRYNMOR CHAPMAN: Oh, because these are the thresholds that we're picking from. So we do know that there are between 0 and 10 candies in each box. So we're going to pick a threshold that's in that range. But yeah, if I have put-- if I have allowed you to pick more than one correct threshold, then that just increases by 5% per candy difference between the boxes.

Are people reasonably happy with this? So this is an example of a problem where we've taken something that isn't exactly a probability problem to begin with, but we've turned it into a probability problem. And we've given what we call a probabilistic algorithm for solving it.

So yeah, if you do it deterministically, there's not much you can do. You pick a box, and if I happen to feed you the correct one, then you win. But you can do much better if you introduce randomness into your algorithm and make random choices.

So you will see a lot more of this in-- oh, sorry, 6 1210, and even more of that in 1220. And, Eric, do you remember? 5210, 5220? There are other randomized algorithms classes. There are some higher-level ones whose numbers I forget that are entirely about randomized algorithms. So it's a very powerful tool that is frequently used in computer science to make your algorithms more efficient or make them better in various ways.

So now that we've finished with our candy boxes, unless anybody has any questions-- questions about candy boxes before we move on? No? OK.

So let's give another example of a very common distribution called the binomial distribution. So this is a distribution that arises very frequently in computer science. And, in fact, we've already been dealing with it implicitly in this class. We just haven't named it.

It's parameterized by two values, so bin of n, comma, p. And basically, this models the following experiment. So we have n coins. These are all-- they're coins. They're independent. Flipping one doesn't affect any of the others. They're all mutually independent.

And they're biased. They're not necessarily fair coins. Each comes up heads with probability p.

And then the RV x, which is the number of coins that come up heads, is or has the binomial distribution. Oops, I ran out of room, sorry. OK.

So we've already seen this in particular with p equals 1/2. Flip several fair coins and count the number of heads. We did that at the very beginning with n equals 3. But more generally, you can flip n coins, and they can be biased towards heads with probability p or away from heads, as the case may be.

You could also model different things, like maybe you've got a very complicated device. Each component may or may not fail. They're all independent. They all fail with probability p. Then you could also count the number of failures. This would also follow the binomial distribution.

So can anybody tell me the PMF of a binomial distribution?

AUDIENCE: n choose x times p to the x [INAUDIBLE]?

BRYNMOR CHAPMAN: So, sorry, could you repeat that?

AUDIENCE: n choose x plus p to the x [INAUDIBLE]

BRYNMOR CHAPMAN: Nice. OK, why?

AUDIENCE: [INAUDIBLE] with x and y [INAUDIBLE] the probability of [INAUDIBLE] p to the x [INAUDIBLE], and 1 minus p [INAUDIBLE] failures.

BRYNMOR CHAPMAN: Yeah. So if you look at sequences of these coin flips, there are n choose x sequences that have exactly x heads. And if you have a sequence with exactly x heads, each index which is a heads, that's heads with probability p. And each index that's tails, of which there are n minus x of them, is tails with probability n minus-- sorry, 1 minus p.

So product rule says that we can multiply all of those together. We get p to the x, 1 minus p to the n minus x. And now each of those sequences, the distinct outcomes, each occurs with the same probability. So we just count them and then multiply by that probability. Does that make sense to everybody?

What about the CDF? Can anybody tell me what the CDF of the binomial distribution is? Yeah?

AUDIENCE: Isn't it just summing up the p of f? So summation of n to the x, p to the x, 1 minus e, and [INAUDIBLE].

BRYNMOR CHAPMAN: Yeah, exactly. So let's say from i equals 0 to floor of x of PMF, of i. So that is frequently how you want to compute a CDF in practice. It is often easy to figure out the PMF. And then you can just compute sums to give you the CDF.

There are other cases where it's easier to go in reverse, like maybe the CDF is easier to compute, and then you can compute finite differences to get back to the PMF. But the fact that both of these are equivalent ways of talking about the distribution means that you can use either one of them, and the other one can be computed fairly easily from the first.

So unfortunately, there are summations involved, though. So that can be a little bit unwieldy. So it is often easier, rather than dealing with the CDF exactly, to approximate it. And this might be good enough for whatever purpose you have in mind.

So let's see how we can do that. Or actually, for the sake of notation, let's abbreviate the PMF. Let's just call it f. And big F can be the CDF. So how could we approximate f of, say, alpha times n? Any ideas?

Well, first, why don't we write out what exactly this says in terms of factorials? So this is going to be n factorial, divided by alpha n factorial, times 1 minus alpha times n factorial, times p to the alpha n, 1 minus p to the 1 minus alpha n. Now, who remembers Stirling? Oh, hand, yeah.

OK, how can we use Stirling to simplify this? Oh, was that not a hand? Oh, sorry. [LAUGHS]

Well, we've got a bunch of factorials here, right? If we approximate each of these factorials using Stirling, then we're going to end up with the following. So n factorial is going to be root 2 pi n.

Actually, let's split up the terms nicely-- over root 2 pi alpha n, root 2 pi 1 minus alpha n, and then multiply this by-- OK. So for the top, we have n minus-- sorry, n divided by e to the n. On the bottom, we've got alpha n divided by e to the alpha n, 1 minus alpha n divided by e to the 1 minus alpha times n. And then these p and 1 minus p terms, we can just keep as they are-- to the 1 minus alpha.

So a little bit unwieldy as it is, but we can simplify it. How can we simplify this? Well, those n's and e's over on the right, those aren't really doing anything. We've got n/e to the alpha n, n/e to the 1 minus alpha times n. And here, we've got n/e to the n, so we can get rid of those. This is just going to end up being 1 divided by alpha to the alpha n, 1 minus alpha to the 1 minus alpha times n.

And over here, we have 1 over root 2 pi alpha times 1 minus alpha times n. And then our p terms remain. OK, what next?

Well, let's try shoving these p's and 1 minus p's in here. So this carries down, and we've got p over alpha to the alpha n times 1 minus p over 1 minus alpha to the 1 minus alpha times n.

What can we do next? Well, what happens if we take the log of this and then raise 2 to that power? It's going to be the same thing, right? But as it turns out, that's going to make our life a bit easier.

So this continues to carry down. So 2 to the-- this is now going to be n times alpha log of p over alpha, plus 1 minus alpha times log of 1 minus p, divided by 1 minus alpha, and parentheses here.

All right, any of you here physicists? Proficient in physics? Does that exponent look familiar to anyone?

Well, as it turns out, this exponent, even though it looks kind of ugly, it's always negative, except for when p equals alpha. So what happens when p equals alpha?

Well, we've got a log of p over alpha here. That's 0. Log of 1 minus p over 1 minus alpha, also 0. So the exponent is 0 there. And if p is not equal to alpha, then because log is concave, this is just going to be-- it's going to be negative.

So we've got this bump at p equals alpha. And it's going to get exponentially smaller everywhere else. Does that make sense to everybody?

So if we actually put in some concrete values-- so for the sake of simplicity, let's say p equals 0.5, n equals, say, 100. Evaluating 100, choose whatever, is going to be a bit of a pain. But if we just use this approximation, then we can get, actually, some pretty good approximations. So if alpha also equals 0.5, what happens?

Well, as we just established, this exponent is 0. So this term is 1. Here, we've got 2 pi alpha times 1 minus alpha times n. So this is going to be 1 over root pi-- sorry, 1 over root 2 pi times 1/2 times 1/2 times n. So that's going to simplify to pi n over 2.

It's going to be-- f of 50 is approximately equal to 1 over square root of pi times 50. And as it turns out, this is about 0.08, so pretty small.

What about alpha equals 0.25? As it turns out, when alpha is 0.25, this doesn't change too much. But this term starts decreasing exponentially. So we actually end up with f of 25 is about 10 to the minus 7. So if you flip 100 coins, the probability that you get exactly 25 heads is minuscule.

And I think the last lecture, right, Eric, is about tail bounds? Do you remember? OK. So, yeah, last lecture, we'll discuss how to get bounds on the CDF instead of just the PMF. And, yeah, as it turns out, it's kind of funny, the probability that you get at most 24 heads, how do you think this compares with the probability that you get exactly 25?

It turns out that this is actually smaller than f of 25. Because it's decreasing exponentially, the CDF is also decreasing exponentially. So it turns out that it's more likely to get exactly 25 heads than to get fewer than 25 heads, which might be a little bit counterintuitive. For larger numbers, that's definitely not true. It's much more likely that I'll get fewer than 50 heads than exactly 50 heads, because the probability of exactly 50 heads is 8%. So the probability of fewer than 50 heads is going to be half of the complement of that.

So what's that, 47%, 46%? I can't math-- 46%. So it's much more likely that you get fewer than 50 heads than exactly 50 heads. But then because it decreases exponentially as you get close to the tail, it turns out it's more likely to get exactly 25 than fewer.

So that's all we have time for today. Or, sorry, that's all we have planned for today. We have a little bit of time left over for questions if you want to come up and ask. And otherwise, we will see you next week.