[SQUEAKING]

[RUSTLING]

[CLICKING]

PETER SHOR: So my plan for today is to start off by showing you how to derive the generating function for Catalan numbers. And then I will show you how to find the formula for Catalan numbers using the generating function. And, finally, I will answer questions for-- because we're having a Thursday, and this Catalan number generating function is not going to take the entire hour and a half, I will throw the floor open to questions about what to-- about material for the test.

And so the test covers everything we've done so far. I don't think there are going to be any Catalan number-generating functions on the test. But seeing the material for today is not going to hurt you on the test because it's more about generating functions.

So recall that the Catalan numbers counted a whole bunch of different things. And Catalan numbers count binary trees, or maybe I should say rooted binary trees. So here are the five rooted binary trees on seven nodes. There we go. So C sub 3 equals 5, and they count Dyck paths.

So Dyck walks were walks that never went below 0 and go from 0 to 2n. So for C3, there should be five of them, and there are. Wait. I want to go up twice, down, and then over. And down here, let's do up, down, up, down, down, and up, up, up, down, down, down. So how would you get a generating function for Catalan numbers?

Well, you want some kind of recurrence. Well, first, let's start out by looking at binary trees. So a binary tree, well, unless it's a binary tree on one node, it starts with a vertex. And there are two branches because, in our definition of binary tree, every vertex has two branches coming out of it. And now, on this branch, we have T let's call it a. And this branch, we have Tb.

And suppose we want to, say, count C sub n. Let's say C sub n, which is the number of binary trees on 2n minus 3 nodes, I guess-- no, 2n plus 1 nodes-- is going to be equal to sum j equals 0 through n minus 1 C sub j C n minus J minus 1.

Why? Well, the number of possibilities for this tree is C sub j because it's just a smaller Catalan number. And if this side has 2j minus 1 nodes, this side must have 2n minus j minus 1 nodes. And so this is number of possibilities is 2-- sorry-- is C sub n minus j minus 1. And when you combine these two trees and add a node, you get a Catalan number C sub n. So C sub n equals summation j equals 0 n minus 1 Cj C n minus j.

And we can check that if you want with the Catalan number 3-- I'm sorry 5, 13, 14-- sorry. I cannot remember my Catalan numbers. So here, 14 should equal 5 times 1 plus 2 times 1 plus 1 times 2 plus 1 times 5. And it does. So this recurrence checks out.

So now you know this recurrence. How would you write down a formula, or an equation, with a C of x? i equals 0 to infinity x to the n C sub n. So what's a equation for C of x? Anybody? Yeah?

AUDIENCE: [INAUDIBLE]

PETER SHOR: Yeah, exactly. C of x is equal to C of x squared. Actually, this isn't quite right because if this is j and this is n minus j minus 1, this is n, so we need to add another x here to make the degrees right. And this isn't quite right, either, because C of x starts 1 plus x plus 2x squared, et cetera. And this side is divisible by x, so this side has 1 in it. So we need to add 1 for the initial condition.

And now we have an equation. It's a quadratic equation. We know how to solve quadratic equations, so let's do it. C of x is equal to-- well, let's put this quadratic equation in the standard order-- C of x squared minus C of x plus 1 equals 0.

And we have 1-- that's minus b-- plus or minus the square root of 1 minus 4aC-- that's 4x-- over 2a. That's 2x. So there's two possibilities for this. How would you figure out which one is the correct one Any suggestions

OK, so what is C of-- what do you get when you substitute x equals 0 nC? Well, you get 1 plus x plus 2x squared plus dot, dot, dot. You get 1. So if you substitute x equals 0 in here and you have the plus thing, you get 2 divided by 0, which is infinity, so that's not right.

If you substitute the minus term, you get 1 minus 1 over 0, which might be right. And you can easily check with calculus that limit as x goes to 0 1 minus square root of 1 minus 4x over 2x is equal to 1. So the right value is minus.

So how do you evaluate this? Well, to get the Taylor expansion of this, you use the binomial theorem-- binomial theorem. Use binomial theorem. So what the binomial theorem says is that, I guess, 1 plus a to the k is equal to 1 plus k choose 1 a plus k choose 2 a squared plus k choose 3 a cubed plus dot, dot, dot.

And it's also true for fractional values of k. So 1 plus a to the 1/2 is equal to, I guess, 1 plus 1/2 choose 1 a plus 1/2 choose 2 a squared plus dot, dot, dot. And, of course, you have to take this and subtract it from 1 and divide by 2x.

So I'm going to do 1 minus square root of 1 minus 4x over 2x-- so that's the generating function for Catalan numbers-- is equal to 1 over 2x. And there's a 1 here as the first term, but it gets canceled by this 1. So we don't have a constant term, which is good because we don't want a 1 over 2x in our expansion. So the next term is 1/2 times minus 4x plus 1.

So 1/2 choose 1 equals 1/2. 1/2 choose 2 equals 1/2 times minus 1/2 divided by 2 factorial. 1/2 choose 3 is 1/2 times minus 1/2 times minus 3/2 divided by 3 factorial, et cetera. So this is 1/2 times minus 1/2 divided by 2 factorial minus 4x squared-- OK, I have forgotten the minus sign here because of the minus there-- plus 1/2 minus 1/2 minus 3/2 over 3 factorial. And there's a 4X cubed, so let's put that here, and the minus sign.

So the first thing you can check is that all of these terms are positive because this minus cancels this minus. This minus is squared, so this minus cancels this minus. This minus is cubed, but there are two minuses here, so that one cancels this minus, et cetera. So we can ignore the minus signs.

So I would like to take the term for x equals 3. Well, I want to claim that that is 7/2 times 5/2 times 3/2 times 1/2 times 1/2-- actually, this is minus 1/2-- divided by-- so I guess that is 5 factorial. There's a 4 to the fourth here. There's a 2 here and a 2 to the fourth here. Well, no. Well, I hope I haven't forgotten any terms.

So what is this? The top is 7 times 5 times 3 divided by 2 to the fifth times 2 times 5 times 4 times 3 times 2. And there's a 4 to the fourth up here. So there are five, six, seven 2's on the bottom and eight 2's on the top, so that's 2 to the eighth over 2 to the seventh. And the 3 and the 5 cancel out. And you get 7 equals 14. So we did this right.

And I want to claim just by-- well, just by generalization, that this is going to be 2 and minus 1 C sub n equals 2n minus 1 double factorial. So double factorial means 2n minus 1, 2n minus 3, 2n minus 5 all the way down to 1. So this is 7 double factorial. That's 7 times 5 times 3 times 1 4 to the n plus 1 divided by 2 times 2 to the n plus 1 times-- should be n plus 1 factorial.

And wait. This is x equals 4, not x equals 3. And so this is 4 to the fourth times x cubed over 2 times 5 factorial. So this is indeed n plus 1 factorial. And this is 2. Well, it is 2 to the n plus 1. 2's here and another 2. And we have n plus 1 factorial.

So now, we need to manipulate this into some kind of-- oh, I should say x to the n here. We need to manipulate this into a better form. And the way we do that is we say 2k minus 1 factorial is equal to 2k factorial over 2k times 2k minus 2 all the way down to 2k minus 4.

So this is the odd terms between 1 and 2k minus 1. So this is all the terms between 1 and 2k. And we divide by the even terms, and we get the odd terms. But this is just 2k factorial over k factorial times 2 to the k. So that says that C sub n is equal to 2n factorial divided by n factorial.

And over here, we have an n plus 1 factorial. There's a 4 to the n plus 1 on top. And there are 2 times 2 to the n plus 1 times 2 to the n on the bottom. This 2 to the n comes from this one. This 2 to the n plus 1 comes from that guy. And the 2 comes from this 2x.

And everything cancels out except-- all the factors of 2 cancel out, and we're left with 2n factorial over n factorial n plus 1 factorial, which is 1 over n plus 1 2n choose n, which you should recognize as the formula for Catalan numbers. Yes?

AUDIENCE: So I have [INAUDIBLE] question. How did you get 1 plus x plus 2 [INAUDIBLE]?

PETER SHOR: Oh. Well, the Catalan numbers are 1, 1, 2, 5, 14. And this is just summation C sub n x to the n is a generating function. And that's 1 plus x plus 2x squared plus 5x to the fourth-- 5x cubed-- plus 14x to the fourth plus dot, dot, dot.

So the expansion. Here are the first few terms of the expansion. The first term, you have a minus 4x and a minus 1 over 2x because the minus 1 over 2x term here comes from this minus sign. The second term, you have 1/2 times minus 1/2 divided by 2 factorial times minus 4x squared. So this is positive. This is minus. And this is minus.

The third term, you have this minus, this minus, and this minus. And multiplying them all together, you get a minus sign. And it cancels out this minus sign, so it's positive. So all the terms are positive. And, really-- yeah.

So now that we know all the terms are positive, we can ignore the minus sign on this, although if you're multiplying by-- yeah. So this is another way of getting the formula for C sub n.

And I want to say that there's a completely different way of deriving this formula for the Catalan number. We derive this equation in a completely different way. And I figured I might as well show it because, well, because it illustrates a neat technique of generating functions that I had some questions about in my office hours yesterday. So Catalan numbers also correspond to Dyck paths.

So a Dyck path can be broken into a sequence of paths with plus and minus 1 steps and which touch 0 only at their endpoints. So let's do it for this Dyck path. The first Dyck path is this guy. The second one is this guy-- 1, 2. The third one is this guy, and the fourth one will be this guy. So it's pretty clear that every Dyck path can be broken into a sequence of this kind of paths.

So if D of x is the generating function for paths that-- well, for paths that only touch 0 at their own points, well, we can use the theorem about generating functions from sequences to get the generating function for Catalan numbers, or generating function for Dyck paths.

So C of x is equal to 1 over 1 minus D of x, right? But D of x, well, you have-- everything in D of x looks like this, where this is a Dyck path, and you have an upstep on the left and a downstep on the right. So every one of this kind of paths, it just comes from a Dyck path where you add two steps to it.

So this says that D of x is equal to x times C of x because x to the k was-- in this generating function, x to the k, k was twice the number of steps in a Dyck path. And to get the raised Dyck paths, you add two steps to a Dyck path, so you get x times C of x. So we have C of x equals 1 over 1 minus x C of x.

And you should be able to multiply this equation on both sides to get-- so we have C of x minus x C of x squared minus 1 equals 0, which is the same equation we had earlier. And you get the same quadratic equation. And you can do everything exactly the same way as we did earlier. Yeah?

AUDIENCE: Can you repeat the D of x [INAUDIBLE]?

PETER SHOR: Right, OK. So D of x are paths that start with an upstep and with a downstep and never go below 1. S equals sum n equals 0 to infinity d sub n x to the n, d paths of length 2n, which start up, end with a downstep, and never go below 1.

So here, we have let's call this 0, call this 1. And we have a path up here that never goes below 1 until the last step, in which case it goes down to 0.

Now, if you take off the last step and the first step here, you get this thing is a Dyck path. So d sub n is equal to the number of these raised Dyck paths is equal to C sub n minus 1 because we have subtracted two edges from d sub n to get a Dyck path.

So D of x equals sum d sub n x to the n is equal to sum C sub n x-- I'm sorry, C n minus 1 x to the n is equal to sum-- oh, gosh-- is equal to sum-- I'm going to claim this is C sub n minus 1 x to the n minus 1 times x.

And maybe we should-- well, there's no d paths of length 0. The shortest one is of length 2, so this summation starts at n equals 1. And we can just say n equals 1 to infinity. And this is n equals 1 to infinity. And you can recognize this as just equal to the generating function for Catalan numbers, which is x C of x.

So I said I would hold a question and answer session in the last, I guess, 30 minutes of this question. So is there anything that people want to-- you have a question?

AUDIENCE: I had a question. How do you get [INAUDIBLE]?

PETER SHOR: Oh, OK. Well, that just comes from plugging this equation into this equation.

AUDIENCE: How do you get the top equation?

PETER SHOR: We got the top equation because of this theorem about sequences for generating functions, which maybe you used on your homework yesterday, although-- so let's go over this theorem about sequences.

So suppose a generating function-- suppose a generating function counts let's call them objects. There are a n objects of size n where size might be the length of a Dyck path, or, I guess, size might be half the length of a Dyck path, but it could be anything else, anyway. You can think of this as an abstract quantity.

So we ask, how many sequences are there of let's call these objects alpha-- of objects alpha 1, alpha 2, alpha 3, through alpha k. And these sequences can be any length. And the size of the sequence is equal to sum i equals 1 through k alpha sub i-- I'm sorry, is equal to the sum of the sizes of these objects.

Then the generating function for the sequences is 1 over 1 minus the generating function for the objects. I should write down the generating functions for the objects.

A of x equals sum n equals, I guess, 1 to infinity alpha sub n x to the n because you're really not allowed to have objects of size 0 in this theorem because then there would be an infinite number of sequences of size k, which doesn't make sense because if you could add arbitrarily many objects of size 0, you can get arbitrary-many objects of any finite size-- or arbitrary-many sequences of any finite size. And that gives you infinities.

So how does this work for Dyck paths? Well, any Dyck path is just a sequence of raised Dyck paths. So you have a-- so here-- actually, here, the Dyck path is the empty path. So the raised Dyck path has length 2 and size 1. Here, the Dyck path is up, down, up, down, which has size 2.

And the raised Dyck path has size 1 more-- up, up, down, up, down, down. And here, the Dyck path is up, down. The raised Dyck path has size 2. And here, we have Dyck path of size 3, so the raised Dyck path has size 4.

And now what we do is we use the formula for the number of-- we say a Dyck path was just a sequence of raised Dyck paths. So that means C of x equals 1 over 1 minus D of x. But raised Dyck paths are just Dyck paths with two extra edges. So D of x equals x times C of x. And you get C of x equals 1 over 1 minus x C of x.

I believe there was a generating function problem on the homework with tiling things of length 3 with 2 by 1 tiles. And you can use the theorem about generating functions for sequences to solve that one, too.

Well, I have I have prepared another problem, which is very similar to that problem, so why don't I present the similar problem?

And I actually think that I will-- how many tilings of 2 by n strip with two shapes of tiles, which you can rotate and reflect if you want.

So these are the two shapes. Use the sequence theorem for generating functions.

And let A of-- let's use a lowercase a-- a sub n b number of tilings of 2 by n strip which cannot be separated into-- cannot be separated into a 2 by k and 2 by n minus k strip.

So in some sense, this is the number of, I don't know, fundamental tilings of a 2 by n strip which cannot be assembled out of other tilings. Let's think-- let's try to figure out what these numbers are.

So the smallest tiling is clearly this. And this is-- so this is a sub 3 because this has length 3. And there are two of these tilings because you can reflect it along a horizontal line. So a sub 3 equals 2. a sub 4 equals 0. I want to say there are no ways to tile a 2 by 4 strip with those tiles. a sub 5 equals 2 because-- a sub 6 equals 0.

So let's call the number of tilings let B sub n be number of tilings. So b sub n is going to be 2, 0, 2. But I want to claim that there is a tiling of b sub n, but it's not elementary tiling because it's made up of two subtiles. I'm doing a bad job of drawing it.

I want to claim that there are four of these because you could reflect the first half, or you could reflect the second half. And that's just this squared because you take two a3's and put them next to each other. And if there are two tilings of length 3, this gives you four tilings of length 6.

a7 equals 2. And

AUDIENCE: Wait. What is bn? [INAUDIBLE]?

PETER SHOR: b sub n is the number of tilings where you don't require that they're fundamental tilings. So b sub n can be made up of-- the b's can be made up of two tilings which are just stuck next to each other, whereas the a's, we want to require that there are no vertical-- you cannot find a vertical line that splits this tiling into two.

AUDIENCE: So the vertical line is defined by separating into k and minus k?

PETER SHOR: That's right. And I want to claim that a sub 8 equals a5 times a3 plus a3 times a5 is equal to 8 because you can take this tile and put it next to this tile-- or you take this tile and put it before this tile, or you can take this tile and put it before that tile.

So thinking about this, if you don't want any vertical lines, the only way to get an a tile is to start with an L and put all these Z tiles next to it and then end it with an L. So a sub 2k plus 1 is equal to 2.

So the odd a's, there are two ways of giving a fundamental tile for odd numbers and 0 for even numbers. So that means A of x is equal to summation k equals 0-- or k equals 1 through infinity x to the 2k plus 1 times 2 equals x cubed 1 over 1 minus x squared.

So now, let's compute this generating function. B of x is equal to 1 over x cubed 1 over 1 minus x squared-- also, 1 minus that.

Oh, wait. I forgot. There's a 2 here. So 1 minus twice that is equal to-- we can multiply 1 minus x squared on the top and bottom, and we get 1 minus x squared over 1 minus x squared minus 2x cubed.

So we have a generating function for equals sum b sub n x to the n. So we have a generating function for b. Now, I want to say that you can read off a recurrence from a generating function.

AUDIENCE: I'm sorry. It's kind of a minor thing. Is that the b 8 of the top right corner on this blackboard?

PETER SHOR: Oh, yeah, right. b8 equals 8. Thank you.

AUDIENCE: So it's not a8?

PETER SHOR: a8 equals 0. OK, so this is b8 equals a5 times a3 plus a3 times a5. I'm sorry about that. So we have--

AUDIENCE: A question. What did you do to get to simplify [INAUDIBLE] 3 sub ax when you did the [? xq ?] thing before you do the sequence theorem?

PETER SHOR: Well, it's just twice times the sum of x to all odd numbers. And maybe I should do this more slowly. This is x cubed times sum 2x to the 2n n equals 0 through infinity. Because the smallest one of these is x cubed, so you can multiply by that. And then you have all the even ones after that.

And now, this, you should know the formula for a geometric series is equal to 2x cubed over 1 minus x squared because sum n equals 0 to infinity of x to the 2n is 1 minus x squared. And, actually, that is very easy to see because 1 plus x squared plus x fourth plus x sixth plus dot, dot, dot times 1 minus x squared is equal to 1.

And now we have a minus x squared plus x squared, a minus x to the fourth plus x to the fourth, a minus x to the sixth plus x to the sixth. So all the other terms cancel out. So this is a formula for a geometric series.

So what did we have? We have-- yeah. I want to say that you can read off a recurrence from a generating function. So suppose you have p of x over 1 minus x squared minus 2x cubed p of x gives initial conditions. And 1 minus x squared minus 2x cubed gives recurrence.

So this recurrence says b sub n is equal to b sub n minus 2 plus 2 b sub n minus 3. I guess we can explain this.

So summation B sub i x to the i is equal to p of x over 1 minus x squared minus x cubed. So what this means is that-- I want to multiply both sides here.

So we're going to get b sub n x to the n minus b n minus 2 x to the n minus 2 times x squared-- oh, wait, this had a 2-- minus b n minus 3 x to the n minus 3-- and there's a 2 here-- equals x to the n term of p of x.

So multiplying just this by both sides and pulling out the x to the n term-- oh, times x cubed-- gives you that. So this says that bn minus bn minus 2 minus bn minus 3 equals 0 unless there is an x to the n term in p of x. And p of x is a polynomial with only a finite number of terms, so after this finite number of terms, you have this recurrence.

So bn is bn-- OK, let's go back here. So we have b sub n equals b n minus 2 plus 2b n minus 3. b0 equals 1. b3 equals 2. b5 equals 2. b6 equals 4. b7 equals 2. b8 equals 8, et cetera.

So let's see if this works. So b2 should be equal to b0 plus 2b minus 1. But b2 is 0 and not 1. But that is because we had the B of x equals 1 minus x squared over 1 minus x squared minus 2x cubed.

That's because this term was minus 1, which says that b2 b2 equals b0 minus 1 when you work it all out. But after the second term, that recurrence works great. b3 is equal to 2b0.

b4 equals 2b2 plus-- I mean, b2 plus 3b1. b5 is b3. b6 is 2b3. b7 is b5. b8 is 2b6 plus b8 is-- yeah, b8 is b6 plus 2b3, which is 8, et cetera.

So you can just read off the recurrence from the bottom of the generating function, and the top numerator tells you what the initial conditions are. Any other questions?

I also could say that you could have figured this out, this recurrence out, without using generating functions because, basically, if you have a-- oh, wait-- tiling that looks like this, there are two ways to extend it.

You can extend it-- use the same tiling in the first part and add one of these Z tiles and add that. Or you can extend it.

So this is-- you've extended this by 2. So that gives you the formula bn minus 2. Here, you've extended this by 3, but there were two ways to do it because you could have added this block, or you could have added this block reflected around its horizontal axis. So that gives you 2 bn minus 3.

And you cannot extend the empty tile in this second way, which explains why b2 is 0 and not 2-- which explains why b2 is 0 and not 1. So you could have gotten this recurrence without using generating functions.

So any other questions? Yeah?

AUDIENCE: So this problem, intuitively, is that a n is like counting the components. So bn is like a sequence of individual components?

PETER SHOR: Exactly. bn is a sequence of these a n components. So you get b of x equals 1 over 1 minus a of x. Yes?

AUDIENCE: Can you show how you got the bn recurrence?

PETER SHOR: How I got the bn recurrence? Oh.

So it's something-- if you go through our proof about how to construct generating functions from recurrences, if a generating function has, I guess, summation f sub n x to the n equals p of x over 1 minus a1x minus a2x squared minus a3x cubed-- and, actually, this is more general than that-- p of x gives initial conditions.

And we have f sub n equals a1f n minus 1 plus a2 f n minus 2 plus a3f sub n minus 3. e of g if p of x equals 1 plus 2x squared, then I want to say f sub 0 equals 1, f sub 1 equals 0, and f sub 2 equals 2.

Well, actually, that's not quite true because I should probably use this as the example. We have B of x equals 1 minus x squared over 1 minus x squared minus x cubed.

So 1 over 1 minus x squared minus 2x cubed would give you f0 equals 1 and fn equals fn minus 2 plus 2f n minus 3. So that would give you 1, 0, 1, 2, and this is-- f4 would be f2-- 1 again. f5 would be 2 plus 2, which is 4.

So this is 1 over 1 minus x squared minus 2x cubed. But we also have minus x squared over 1 minus x squared minus 2x cubed. And this gives you, basically, this, except you have to-- this will give you a minus 1 here and then 0 minus 1.

The next one was minus 2 and minus 1 minus 4. And then, when you add them up, you actually get the sequence of f sub i. So that gives you 1, 0, 0, 2, 0, 2, et cetera. And we have 1, 0, 0, 2, 0, 2, et cetera.