PROFESSOR: So far in our discussion of probability, we've focused our attention on the probability that some kind of event occurs-- the probability you win the Monty Hall game, the probability that you get a heads when you flip a coin, the probability we all have different birthdays in the room-- that kind of thing. In each case, the event happens or it doesn't. It's a zero-one kind of thing.

For the rest of the course, we're going to start talking about more complicated situations. Instead of talking about zero-one events, like winning or losing, we're going to start talking about how much you win. Instead of talking about whether or not you get a heads in a coin flip, we're going to talk about flipping lots of coins and asking how many heads did you get?

Now, this involves the notion of a random variable to count how much or how many in a random situation. And actually, the name is sort of weird. It's not a variable at all. In fact, a random variable is a function from the sample space to the real numbers.

See random variable R is a function R from the sample space to the reals. So this is the random variable, and we'll often denote that by rv. This is the sample space of all possible outcomes. And this is just the reals.

All right, so in other words, a random variable is just a way of assigning a real number to each possible outcome. For example, say that we toss three coins. We could let R equal the number of heads among the three coins. And R is thus a random variable. For example, for the outcome where the first coin is heads, the second is tails, and the third is heads, R would be the real number 2, indicating I got two heads in that outcome.

We could also define another value called M, which we'll say is 1 if all three coins match, if they're all the same, and 0 otherwise, if they're not all the same. For example, M on the outcome heads, heads, tails is 0. M on the outcome tails, tails, tails would be 1, because they all are the same.

Is M a random variable? Yeah, it assigns a value, a real number, to every outcome. And that

value is either 0 or 1, depending on if all the coins match. In fact, it's a very special kind of random variable that's called an indicator random variable, also known as Bernoulli or characteristic random variable, because it's zero-one.

An indicator also known as a Bernoulli or characteristic random variable, is a random variable whose range or possible values is just 0 and 1. So it's a random variable where the real numbers are 0, 1 that you assign to every sample point. And it's called a characteristic or indicator random variable because it indicates which sample points have a certain property.

In this case, all the coins were the same. Or it indicates that a sample point has a certain characteristic, which is why it's called a characteristic random variable. And we're going to be talking a lot about these kinds of random variables over the next couple of weeks.

Now, in general, a random variable is equivalent to, or it can define a partition on the sample space. For an indicator random variable, it defines a partition with two blocks. For example, say we look at the sample space with three coin tosses there. There's eight outcomes. There's heads, heads, heads; heads, tails, heads; heads, heads, tails; tail, heads, heads; and the reverse of those.

This is the sample space when I toss three coins. There's eight possible outcomes. The random variable M defines this partition. Up top, you have M equals 1, and down below you have M equals 0. Because these are the outcomes where all the coins match. These are the ones that don't.

The random variable R defines a different partition. In that case, the partition has four blocks. Here is R equals 0, no heads. This is R equals 1, to one head. R equals 2, and R equals 3. So a random variable just sort of organizes your sample space, partitions it into blocks, defined by all the elements in the block, the outcomes in the block have the same value for the random variable.

And, in fact, every block is really an event. In particular, we can say that, if you look at the outcomes for which the random variable on that outcome equals some value x, this is simply the event that R equals x. The random variable is x. And, of course, an event is just a subset of the sample space. It's a collection of outcomes that define the event.

And since we know we could talk about probabilities of events, we can now say things like this. We can say the probability that a random variable equals x is simply the probability of that

event happening, which is the sum over all outcomes for which the random variable equals x for that outcome, summing its probability. So the probability that the random variable equals a value is the sum of the probabilities of the sample points for which R has that value.

All right, this is all really basic stuff, but important to get down. So, for example, what's the probability that R equals 2 in our case with the three coins? You flip three coins-- let's say the coins are fair and mutually independent. We probably need to know that. What is the probability R equals 2?

AUDIENCE: 3/8?

PROFESSOR: 3/8, because you've got three outcomes, each with a probability 1/8. All right, so by this definition, that would be the same as the probability of heads, heads, tails; heads, tails, heads; tails, heads, heads; that's 3/8.

What's the probability that M equals 1? What's the probability that M equals 1? M is 1 if all the coins match. 1/4 because you've got two cases there-- heads, heads, heads-- whoops-- or tails, tails, tails; and that's 2/8, or 1/4.

Now we can also talk about the probability that a random variable is in some range. For example, I could ask, what's the probability that R is at least 2? All right, that would mean the sum i equals 2 to infinity, the probability R equals i. And what is that? What's the probability? R is at least 2. 1/2. It's the probability that R equals 2, plus the probability that R equals 3. There's four outcomes there, each with 1/8, so that's 1/2.

And finally, we can talk about the probability that R is in some set of values. So we could say, for example, for any subset of the real numbers, we define the probability that R attains a value in that set is simply the sum over all possible values in the set, that the probability R equals that value. So we could ask, what's the probability that, say, we take A being the set 1, 3. I could ask for the probability R is A when it's 1, 3, which is the same as the probability R is odd. Because Our could only be 0, 1, 2, or 3.

What's the probability R is odd in this example? Flip three coins, we're asking the probability you get an odd number of heads-- one or three heads. 1/2. Again, you've got four sample points-- here and there. So that's 1/2. All right, any questions about the definition?

All right, now, we're only going to worry about the case when random variables are discrete. Namely, there's a finite number of values. If you have continuous random variables, which

they'll deal with in other courses, instead of doing sums and stuff like that, you're working with integrals. But for this course, it's going to be all countable sets, and usually often finite sets.

Now, conditional probability carries over very nicely to random variables as well. For example, we could write down the probability that R equals 2 given that M equals 1, where R and M are the same random variables we've been looking at when you flip the three coins. What is that probability? 0, the probability of getting exactly two heads when they're all the same, can't happen. It's 0. You can either have three heads or no heads. So that's 0.

The notion of independence also carries over to random variables, but you've got to be a little careful here. Two random variables, R1 and R2, are said to be independent if-- and this is a little complicated-- for all possible values, x1 and x2 in the real numbers, the probability that R1 is x1, given that R2 is x2, is the same as the probability of R1 equals x1 not knowing anything about R2. Or there's a special case when this can't happen. Namely, the probability that R2 equals x2 is 0.

So, for two random variables to be independent, it needs to be the case that, no matter what happens out here with R2, and whatever you know about it, it can't influence the probability for R1 to equal anything. So no value of R2 influences any value of R1. So it's the strongest possible thing with independence for it to apply to random variables.

And there's an equivalent definition, and we're going to use these interchangeably, just as there were two definitions for independence of events, there's the product for them. So the equivalent definitions says for all x1 and x2 in the real numbers, the probability that R1 equals x1 and R2 equals x2 is simply the product of those probabilities. Probability R1 equals x1 times the probability R2 equals x2.

And you don't have to worry about that case where R2 being x2 is 0. So we can use both of those as definitions of independence of random variables.

All right, so let's see an example. We've got R and M. We've talked about two random variables. Are they independent random variables? M is the indicated random variable for all the coins matching. R is the random variable that tells us how many heads there were in three coins. Are they independent?

AUDIENCE: No.

PROFESSOR: No, why not?

AUDIENCE: Because information with what one is narrows down [INAUDIBLE].

PROFESSOR: Yes, that's true. Information about what one value is influences the probability that the other guy has a certain value. In particular, we can find a case of x1 and x2 where this fails. One case would be the probability of what we've already done, that R equals 2 and M equals 1. What was this probability? 0. If everything matches, you can't have two heads. That's 0. That does not equal the probability of two heads times the probability they all match. Because this probability, two heads is 3/8, probability they all match is 1/4. And 3/8 times a 1/4 is not 0. So R and M are not independent.

OK, now in general to show two random variables are not independent, all you gotta do is find one pair of [? out ?] values for which there's dependence. To show their independent, you've got to deal with all possible pairs of values. So it's harder show independence in general.

All right, let's do another example, get a little practice with this. Say we have two fair independent six-sided die, normal dice. And the outcome of the first one is D1, and the outcome of the second roll is D2. And by independent, I mean here that knowing any information about what the second die is does not give you any change in your probability that the first die has any value.

And now we define another value, S, to be D1 plus D2, the sum of the dice. Is S a random variable? Yeah, for any outcome-- an outcome being a pair of dice values-- it maps that outcome to a real number, a number between 2 and 12, OK?

Let's do another one. Let's let T be 1 if S is 7, namely the sum of the dice is 7, and 0 otherwise. T is also a random variable. In fact, what kind is it? Yeah, indicator, characteristic, whatever. It's telling you if the sum of the dice is 7 or not.

All right, now, there's four random variables here. Each die, which we already have told you, we're assuming are independent. The sum and the indicator if the sum is 7. All right, what about these two values? Are D1, the first die, and the sum independent?

AUDIENCE: No? No?

PROFESSOR: No, intuitively they're not. Because if I know something about the first die, I probably know something about the sum. Now to really nail that down, you've got to give me a value for this,

which influences the probability that obtains some value, right? That's how you would convince me that, in fact, they're dependent.

Can anybody give me a value for this that changes the probability this equals something?

AUDIENCE: Guess? If it's like 1, then [INAUDIBLE] has to be 6.

PROFESSOR: If this is 1, what value of this could be influenced?

AUDIENCE: You know that in order for-- the only set of values S can take now, it can only be from 1 to-- from 2 to 7.

PROFESSOR: From 2 to 7. In particular now, this can't be 12. So what I could do is say the probability that S equals 12 and the first die was 1, what's that probability?

AUDIENCE: 0.

PROFESSOR: 0. And that does not equal the probability S is 12 times the probability the first die was 1. Or we could plug it into the other definition. The probability that S equals 12 given the first die is 1 is not equal to the probability S equals 12. And, in fact, probability the first die is 1 is not 0. So either definition, of course, works to show that they are dependent. So S and D1 are dependent. All you need is one possible pair of values to show dependence. That's usually easy to do.

All right, what about D1, the first die, and T, the indicator for getting a 7? Are D1 and T independent random variables? All right, somebody's shaking their head no. They seem to be dependent because knowing the first die seems like it should tell you something about the probability of getting a 7 for the sum. That's a good first intuition. In fact, you always want to assume things are dependent unless you convince yourself otherwise. It's always a good rule.

Now, in fact, in this case, it is independent. The probability of getting a 7 is not, it turns out, influenced by the first die. Anything else would be, but 7 is not. Let's see why that is. And to do that, we'd actually have to check, I think, 12 cases, all possible values for D1 and T, but you'll get the idea pretty quick.

The probability that T equals 1, namely you got a 7, given that D1 equals 1. What's that? The probability of getting 7 for the sum given your first die is a 1, what's that probability?

AUDIENCE: 1/6?

PROFESSOR: 1/6, because the second one better be a 6. All right, and what's the probability T equals 1? What's the probability of rolling a 7? Yeah?

AUDIENCE: 6/36.

PROFESSOR: 6/36. There's six ways to do it. 1 and 6, 2 and 5, 3 and 4, 4 and 3, 5 and 2, and 6 and 1, out of 36 possibilities equally likely. So that worked here. Now probability of getting a 7 given the second die-- sorry-- the first die is a 2. What's that? Probability of getting a 7 given that your first die is a 2. 1/6, because you've got to have the second die be a 5, and that happens 1/6. And that equals the right thing.

In fact, I can keep on going here. Probability T equals 1, given D1 is anything, 6. Well, for every value of D1, there's exactly one value of D2 that adds to a 7. So it's still 1/6, which is the probability of getting the 7 in the first place. Now we've also got to check the probability T equals 0, given all these cases for D1. D1 equals 1.

What's this? The probability of not getting a 7 given the first die is a 1. What's that?

AUDIENCE: 5/6.

PROFESSOR: 5/6, it's the opposite of this case. They pair up. That's got to be 5/6, and that is, of course, the probability of T equals 0. And the same is true for all the other five cases. The probability T equals 0 given any value of D is 5/6, which equals, the probability T equals 0. So we check all possible values of T1 and all possible ways of D1, and lo and behold, it works every time.

So knowing the first die does not change the probability of getting a 7. Would it change the probability of getting a 6? Yeah, because if the first die were a 6, you know you're going to get more than a 6. So it just holds true because I picked it to be an indicator for 7. Any questions about that?

So if you're ever asked to show things are independent, you've got to go through all the cases. If you're asked to show their dependent, just find one case that blows it up, and you're done. OK, you can also talk about independence for many random variables. And you have a notion of mutual independence. And this is a little hairy. It's a natural thing. R1, R2, up to Rn are mutually independent if for all the values of all the random variables, so x1, x2, xn. And we're going to give the product form, because it's the simpler way to do it here.

The probability that R1 is x1, and R2 is x2, and Rn is xn equals the product of the individual probabilities. It's the natural generalization for two variables. And there's also the equivalent form in conditional probabilities, but that's a little harrier to write down, and we generally don't work with that. Any questions about mutual independence? Yeah.

AUDIENCE: [INAUDIBLE].

PROFESSOR: So any subset what?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yes, in the conditional form you check any subset, here you don't have to worry about the subsets. You take any possible value for all of them, so there's no subset notation in this version, so you can do without it. That would be equivalent, there's another form where you look at all the possible subsets as well, and conditioning on that. But this is good enough, and this is probably the simplest way to do the mutual independence, is like this. So every variable is included here.

And that will imply the same thing for any subset. You don't have to check every subset here. And then you could do that by just summing over Rn being everything, and that cancels away to get rid of items. But you don't have to worry about that.

OK, so we're going to change gears a little bit now, and talk about the probability distribution function, which is just a way a sort of writing down or characterizing the probabilities associated with each value being attained. So given a random variable, R, the probability also known as the point, distribution function, also denoted pdf, probability distribution function, for R is, well, it's very simple. It's just the function f of x, which equals the probability that R is x. Very simple. But now we're characterizing it as a function.

And there's also a notion of the cumulative distribution function, which is going to be the probability that R is less than or equal to x. Let's write that down. The cumulative distribution function F for random variable R is simply f of x is the probability that R is at most x, which is just the sum over all y less than or equal to x the probability R equals y.

So the distribution functions characterize the probability a random variable takes on any value, and there's certain common ones that just come up all the time. And so people have done a lot of analysis about them because they occur so frequently. And we're going to talk about three of them today. The first is really simple, and that is the Bernoulli random variable, or

indicator random variable.

All right, so for an indicator random variable, f 0 is p, and f 1 is 1 minus p, for sum p. The probability could be half, if you're flipping a coin. The probability of heads could be a half, probability of tails would be a half. It's just two values, 0 and 1. And then the cumulative function is very simple. F 0 is p, big F of 1 is 1. So that is about as simple of functions as you can get.

The next simplest, and also very common, is called a uniform random variable. Let's define that. For a uniform random variable-- and we have to define what it's defined on-- on say the integers from 1 to n, every value is equally likely. All the integers from 1 to n are equally likely to occur. And so in this case, n is a parameter of the function. fn of K is simply 1/n. Each integer from 1 to n is equally likely. So it's 1 and n.

All right, what is the cumulative distribution function for this? What is big F of n of K for the uniform? The probability that the random variable takes a value that's at most K?

AUDIENCE: [INAUDIBLE]?

PROFESSOR: Close, K/n. And you could think of K being 1. You have a least the 1 over n probability. So you have K chances. It could be 1, 2, 3, 4 up to K, each has a 1/n chance. Because we've got, the definition is less than or equal, not less than. Uniform distributions come up all the time, rolling a fair die is uniform on 1 through 6 for values. If I pick a random student in the class, the chance I pick you is 1 in the size of the class. We have uniform sample spaces. Each outcome occurs equally likely. All right, any questions about uniform? [INAUDIBLE].

OK, so while we're talking about uniform, I want to play a game, whose optimal strategy is going to turn out to be related to uniform distributions. Now the game works as follows. I'm going to have two envelopes here. And inside each envelope, I've written a number. And the number in this case is a number between 0 and 100 inclusive. It's not a random number, because I wrote it, and maybe I'm not such a nice guy. I might have picked nasty numbers to write here. They are different, though. The numbers are different in the envelopes.

Now I'm going to pick a volunteer from the class, and their job is to pick the envelope with the bigger number. And if they get the one with the bigger number, then they get one of these guys for ice cream. And if they don't get the one with the bigger number, they get the nerd pride pocket protector.

Now, since you don't know which number I put in which, 50-50 if you get the prize, the $10 gift certificate. So to give you a little bit of an edge, I'm going to let you open one of the envelopes and see what number's there. Then you have to decide, do you want the number you got, or do you want the other envelope? OK? Is the game clear, what we're going to do?

OK, so need a couple of volunteers. We're going to do it twice. All right, somebody way in the back up there. Any other volunteers? All right, come on down. Oh, you've already played before. (LAUGHTER) You already got. Who hasn't played before, wants to-- all right, come on up. Now you want to be thinking about, boy, is there a strategy here? Or is this just dumb luck, 50-50?

OK, what's your name?

AUDIENCE: Sean.

PROFESSOR: Sean. All right, I got two envelopes, Sean. They are numbers between 0 and 100. And you can pick one, and we'll reveal the number, and then you decide. Do you want the number you got, or do you want the other one? The goal is to get the bigger number. So take one and open it. What did you get?

AUDIENCE: 6. He got

PROFESSOR: A 6. The numbers go from 0 to 100. What do you think Sean should do?

AUDIENCE: Switch. [INAUDIBLE].

PROFESSOR: What do you think you should do?

AUDIENCE: [INAUDIBLE].

AUDIENCE: They've got to check the other one.

PROFESSOR: No, no, you can't look at the other one. And unfortunately, you're going to play with different envelopes. What should you do? 6, there's a lot of numbers bigger than 6, Sean. But they might not be in that envelope. Might be a 0 in that envelope.

AUDIENCE: That would suck.

PROFESSOR: Yeah.

AUDIENCE: I'll stay.

PROFESSOR: You're going to stay with 6?

AUDIENCE: Yup.

PROFESSOR: All right, he picked 6. What do you think he should have done? How many people think he should have switched? Ooh, Sean. How many people like Sean's choice? Not so good. All right, let's see if you won. Here we go. 5. Sean wins the ice cream. Good work. Now Sean, what was your thinking here?

AUDIENCE: I was thinking [INAUDIBLE] in the other envelope, knowing that 6 doesn't tell me anything about what's in the other one, so my chances of seeing [INAUDIBLE].

PROFESSOR: Totally 50-50. How many people buy that argument? It's 50-50. He really has no idea what's in the other envelope. How many people think there's a better way? Not too many. All right, we're going to try it again. Well done. I've got different envelopes here. And tell me your name again?

AUDIENCE: Drew.

PROFESSOR: Drew, OK Drew. Two envelopes. Which one would you like to open?

AUDIENCE: They're both labeled B.

PROFESSOR: They're both labeled B. That won't help you there. All right, he's got one. Let's see what you got. 92. Oh my goodness, and we've seen what 5 and 6. What are you going to do? 92. What should he do, guys?

AUDIENCE: Stay.

PROFESSOR: Stay. Oh he's got a big number there. You're going to switch? You're giving up a 92? Ooh, you're sure? You don't want the 92? How many people think Drew is going to win?

AUDIENCE: He's trying to lose.

PROFESSOR: A couple people. Let's see.

AUDIENCE: 91. It's either 91 or 93.

PROFESSOR: 93, how did you know there was a bigger number? Wow, very good. There you go. What was your strategy?

AUDIENCE: I figured you'd probably pick them within 1.

PROFESSOR: Yeah, that's good.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Actually, they're both [INAUDIBLE] to 93. Would you still switch?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Still switch. Hmm. All right, so is it 50-50 then? Because you see a 92 a 93, you're switching either way.

AUDIENCE: Yeah.

PROFESSOR: 50-50. All right, how many people still like 50-50? He can't beat 50-50 here. Any ideas? Can you beat 50-50? Yeah.

AUDIENCE: I just pay attention to your face.

PROFESSOR: I guess I gave it away. I think I've lost every bet so far in the course here, even on Monty Hall, every time. So I guess I have a tell or something here. Now, in fact, you can beat 50-50. The information is helpful. Now, just to be clear, are these numbers that I wrote down random? No. No, I'm trying not to give away-- my ice cream bill is going through the roof here. I'm trying to make it hard. They're very much not random. There were two smalls and two bigs. What is random here?

AUDIENCE: Which one he picked is.

PROFESSOR: Which one he saw, that's 50-50, effectively. So that's random there. Now his strategy could also be random, but it wasn't. His strategy was, if he sees a big number, he's swapping, which is odd. Most people see the big number, they want to keep it. You know, I'd have done well today if I had a really big number and a really small number, because then I would have won both times.

OK, so to see why there might be a winning strategy, or better than 50-50, imagine that I had been nice and put a very small number in one envelope and a very big number of the other

envelope. Say I had a 5 in one and a 92 in the other. Can you win then, if you know that? OK, how do you win if you know there's one less than 10 and one bigger than 90? Yeah?

AUDIENCE: If you get the one less than 10, switch.

PROFESSOR: Yeah, and you're going to win with what probability? 1. It's a certainty. All right, well, but I'm not that nice, say. Say, though, there is a threshold x, such that one of the numbers is less than x, and one is bigger than x, and you know x. Can you win now? Say you know that one number is less than 47 and 1/2, and one is bigger than 47 and 1/2. Can you win? Yeah, because if you get the one less than 47 and 1/2, you switch. And otherwise you'll stay. So again, you win with certainty. All right, that's good.

Is there always such an x? You may not know it, but is there always such an x? Yeah, because the numbers are different. Just pick up a real number in between them. You know, with 92 and 93, there's 92 and 1/2, is such an x. Now the only problem is you don't know it. And there's no way to figure it out that's within the rules of the game.

Now in life, if you don't know something, and you can't figure it out, what can you do?

AUDIENCE: Guess.

PROFESSOR: Guess it. All right, now this turns out to be a really good thing to do, and especially good in a lot of computer science situations. If you don't know it, and you can't know it, well you could try to guess and you might be right. Now, if you guess x and you are right, what's your probability of winning? 1. If you don't guess x, what's your probability of winning? Say you guess wrong, but you follow the rule that if it's less than what you guessed, you swap it, otherwise you don't. What's your probability of winning that? P? Well, yeah, what's P? It's a nice value of P.

You know, say you guessed-- well, say you weren't paying attention and you guessed 200. And your rule is, if you're less than what you guessed, you're less than 200, you swap, which means you're just going to swap. And you started with a random one, so what's your chance of winning? A half. So if you guessed wrong, you didn't lose anything. You still win with probability a half. If you guessed right, you win with probability 1, and there's some chance you guessed right, so now you've got a strategy that beats 50-50, because you guessed.

OK, and this is a whole field in computer science where you get randomized algorithms, where this strategy depends on a coin flip, on guessing a value. And it leads to getting potentially a

better outcome. All right, so let's prove that now and see what the probability is of winning with the guess strategy and what's a good way to guess.

So we're going to first formalize our random protocol. All right, so if this is the winning strategy, well first the envelopes contain y and z, and they're in the interval 0 to n. And y is going to be less than z. That's how we're going to define them. Now you don't know y and z, but y is the smaller number in the envelope, z is the larger. And in our example, n was 100.

Now the player chooses x randomly, uniformly among all the possible half integers. So he might pick 1/2, he might pick 1 and 1/2, 2 and 1/2, all the way out to n minus 1/2. You know, because if the player picks anything else, it's useless. He won't have a chance to win. But he might, in the case of 5 and 6, might have picked 5 and 1/2.

And we're going to make our random guess be equally likely among those n values, because well he doesn't really know what the numbers are I put in the envelopes. If he sees 6, he doesn't know if it's 5 or it's 7, OK? So that's why we're going to pick them all equally likely and it's uniform.

Now the player is hoping that he picked a value that splits y and z, OK? Because then he's going to win. If he picked an x between the numbers in the envelopes, and follows the swap, if he got the small number less than x, he's going to win, all right? Then the player opens a random envelope, 50-50, to reveal a number r, which is either y or z, but the player doesn't know which one. So one of the numbers gets revealed. And then the last step is the player swaps if the number he revealed is less than the guessed split number.

That's the strategy, and it's a strategy that depends on a random guess or a random number. So let's figure out the probability of winning with that strategy. And we're going to use the tree method. Well, the first branch in the tree method is whether or not-- where the guess wound up. And there's three cases for how the guess did.

You might have guessed too low, in which case, x is less than y, and y is less than z. You might have guessed perfectly, in which case x is between y and z, equals not possible. Or you might have guessed too high, in which case y is less than z is less than x. All right, so here's low, here's high, and here's an OK guess, a good guess.

Now what's the probability you guessed low, that you picked an x less than y? y is integers, x is the half integers. What's the probability to assign to that chance, that you guessed low? You

could use the value of y in your answer. You don't know y, but we can write it on the tree. What if y was 1? The smallest number in the envelopes is one, what's the probability you guessed below 1?

Not quite 0, one of your possible end guesses would be too low, namely, a half. If the smallest number y were 2, what's the probability you guess below 2? 2/n. And in general, if y is the smallest value, you've got y possible guesses that it would be too low, namely all the half integers less than y. So the probability you guess low is y/n, because each one is 1 in n chance. And there's y that are too low.

What's the probability your guess is good, that you split y and z? z minus y over n, because they're-- between these integers, z and y, there's z minus y half integers. Each could have been guessed with probability 1/n. And what's the probability you guessed high? n minus z over n, because there's n minus g half integers between z and n being the last possible one, and minus the half there.

All right, so we've got the first branch. Now the next branch is the revealed value. Did you open the smaller one or the larger one? r equals y means you opened the smaller one. r equals z means you saw the bigger one.

All right, if I've gone down this branch where I guessed too low, but I don't know that, yet, but say I've guessed too low, what's the probability I opened the smaller envelope? One half, because you're just picking a random envelope and opening up. So these each happen with probability a half. And that's true no matter what. They're all one half.

All right, well, now we can compute the probability of each outcome. This is y/n times 1/2, is y over 2n. This is y over 2n also. This is z minus y over 2n. z minus y over 2n. n minus z over 2n. All right, I got all the sample points. I got all the probabilities.

Let's figure out if you win now. And to do that, the first thing we have to figure out is, do you swap? Do you swap here? Well, what happened? I revealed y, which is bigger than my split value x, my guessed value. And so I am guessing I've got the biggest value. So I don't swap. So there's no swap.

What about here? Do I swap here? I open z, I saw z. z is bigger than what I think the midpoint is, so I wouldn't swap. I only guess if the value I reveal is less than my guessed midpoint. If I got a value I see bigger than my midpoint, I think I've got the big one. So there's no swap.

What about here? Do I swap here? Yes, here I swap because I open up something that's less than my guessed midpoint, so I think I've got the small one so I'm going to swap.

What about here? Do I swap here? No swap, because I opened up something that's bigger than midpoint. What about here? I swap. And what about here? Swap on both of them, because both of them are smaller than my guessed midpoint. All right, now let's figure out if you won or you lost.

So here I did not swap and I started with a smaller value. So what happened? Did I win or lose? Lose. I opened he smaller value and did not swap. What happened here? I win. I open the bigger value and did not swap. Here what happened? Win, I opened the small value when I swapped, that's a win. Here? I opened the big value, did not swap. It's a win. Here I opened a small value and swap, that's a win. And here I opened a big value and swap, that's a lose.

All right, now we can compute the probability of winning. The probability of a win is the sum of these four sample points-- y over 2n plus z minus y over 2n plus z minus y-- whoops-- over 2n plus n minus z over 2n. That equals-- z cancels the z, and a y cancels the y, so I've got n, one left, n left, one z left, and one negative y left over 2n. And I can simplify that. N over 2n is 1/2, plus now what's left over is z minus y over 2n. And we know the z and y are different by at least 1.

So this is at least 1/2 plus 1 over 2n. And so if n is 100, you've got a 50 and 1/2% chance of winning. If n is 10, the numbers are from 0 to 10, you've got at least a 55% chance to win, which is pretty high, OK? Any questions here? You see what's going on? So here's the zone where you guessed right. You get the win either way. Here you guess low, and it doesn't make any difference. You're 50-50. You guessed high and you're 50-50 again. So guessing helps. Any questions about that? Yeah.

AUDIENCE: So aren't you assuming that-- because you say, have a range, a greater range of numbers, that's there's a greater chance that the number is going fall in that range.

PROFESSOR: I was the nasty guy here. I picked z and y to be consecutive numbers, which minimized your chance of doing well with this strategy.

AUDIENCE: Because it seems like we don't know anything about [INAUDIBLE].

PROFESSOR: The distribution of what?

AUDIENCE: Of possible other numbers in envelopes [INAUDIBLE].

PROFESSOR: Right, you know nothing about the distribution because there is none. There's no randomness there. I picked worst case numbers here.

AUDIENCE: But it's still sort of like a [INAUDIBLE]. If it's 10, you still have a better chance to swap than if you [INAUDIBLE] below that.

PROFESSOR: No. No, any deterministic strategy you pick, if you don't guess that random x, you will not do better than 50-50. The only way you'd beat 50-50 is to, in your mind, make a random number 13 and 1/2 and flip if you see less than 13 and 1/2. If you come in with a strategy that hey, 10 is a small number, out of 100, and so I'm going to swap if I see a 10, well, no, because first it could be-- I might have done 9 or 11, either one. In fact, if I know that's your strategy, I'm going to make a 9. And then you're doomed in seeing a 10.

AUDIENCE: So is there a best way to pick that?

PROFESSOR: Yes, I also have an optimal strategy. In fact, there's two interesting things about this game which we won't prove. The first thing we'll prove, we will prove but it's true, is this is the optimal strategy for you. And my optimal strategy is to pick a random value of y between 0 and n minus 1, and then to make z be one more. And the way to think about this, and there's whole classes you can take on game theory, is that suppose I do pick my guys randomly by picking a random y uniformly from 0 to n minus 1. So my smaller number is anything between 0 and 99, equally likely, and then the bigger number's just one more.

So if I picked 92 for the random one, the next one's going to be 93. If that's my strategy, even if I tell you that, you cannot do any better than getting 1/2 plus 1 over 2n probability of winning, no matter what you do in the whole world, OK? And if you use this strategy here, no matter what I do, even knowing that's your strategy, I can't keep you from getting this much. It's called a minimax solution.

So my optimal strategy is to pick uniformly in 0 to 99, and the next one's one more. Your optimal strategy is this. Pick a number uniformly in the half integers there, and swap if you see something smaller. And there's no better strategy. In fact, any deterministic strategy, you don't do better than 50-50. The optimal strategies require randomness for each of us. OK, any more questions about what happened here, this game? Yeah.

AUDIENCE: [INAUDIBLE] when you see a small number you're going to swap, does that just mean

[INAUDIBLE]?

PROFESSOR: Yeah, basically, yeah, that's what it means. Effectively, that 50 is your random number. If you think OK, if I see less than 50, I'm going to swap, bigger than 50 I don't. Now you have to decide what happens at 50, if you saw exactly 50. So you'd probably go in at 49 and 1/2 and 50 and 1/2. And so that could be construed as that way. But you didn't pick it randomly, and that sort of human intuition, to swap so I could use that now to design my strategy, knowing that your random numbers could be 50 and 1/2, OK?

No, I'll be 50-50 because what I'll do, if I know that's your strategy, my numbers will be 1 and 2. Well, when you do these analyses, you sort of, if it's declared that's it, I can assume that. And I might actually guess that. In fact, that's why I did pick two very small numbers and two very big numbers. But given the perverse nature here, when you saw the big number, you got rid of it, and when you saw the small number, you kept it. So you sort of did reverse psychology and it worked out, by chance, that it worked out that. Normally it would go the other way, I think. Of course, you've all seen enough games now, you know to do the non-standard thing, I think.

Now, this kind of thinking comes up all the time in computer science algorithms. You know, the protocol that's used to communicate across a network, ethernet, shared bus is randomized. Each entity that wants to use the bus flips a random number, flips a random coin. You could think of it that way. And it broadcasts with that probability. And if there's a collision, it backs off and chooses a smaller probability next time. And if it gets through, then it tries to cram a bunch of stuff through, until you get a collision again.

The best algorithms for sorting a list of numbers are randomized, quick sort, which you'll see in 6046, is a randomized algorithm. In fact, it's not dissimilar. You guess a random value and split all the numbers based on the random value. Who's bigger, who's smaller, and then you [INAUDIBLE]. You do things like that.

OK, that's all I'm going to say about uniform distributions. Next I want to talk about the binomial distribution. And this is the most important distribution in computer science, and probably the most important in all of discrete calculations in the sciences. In continuous distributions, you get the normal distribution. But, for discrete problems, the binomial is the most important.

And there's two versions. There's the unbiased binomial distribution and then there's the

biased one. Now this one's a little more complicated. You've got a parameter n again, and the distribution function on k, probability of getting k is n choose k times 2 to the minus n. And n is at least 1, and k is between 0 and n.

And then for the general binomial distribution, we have fn of k, the probability of getting k is n choose k times-- actually, it's fnp, there's another parameter p here. p to the k times 1 minus p to the n minus k power. And p is a value between 0 to 1 typically. It's a probability of something happening.

So in the general case, you've got nnp. The unbiased case corresponds to when p is 1/2. Because if p is 1/2, you get 2 to the minus k and 2 the minus n minus k, and that's just 2 to the minus n. So you can think of this as the case when p is 1/2.

All right, to give you an example why this is so important, why it comes up all the time. Imagine that you've got a system with n components. And that each component fails with probability p. And you want to know the probability of having some number of failures. So for example, there's n components and each fails independently-- in fact, mutually independently-- with probability p. And p will be between 0 and 1.

And now we're interested in the number of failures, so we'll make R be a random variable that tells us the number of failed components. And it turns out the answer is simply that function, which we'll prove is a theorem. The probability that R equals k is simply fnp of k [INAUDIBLE]. So the general binomial distribution gives you the distribution on the number of failures in a system with n components where they fail with probability p. So let's prove that.

To do that, we're going to construct our tree again. It's a little big because you've got n components. But you look at the first component, and that can fail or not. And it fails a probability p. It's OK with probability 1 minus p. And you have the second component. It can fail or not. Again p and 1 minus p. And you keep on going until you get to the nth component. And that can fail or not, the probability p or 1 minus p, in general down here.

And now we look at all the sample points out here. And it's all length and vectors of failure or not. So the top sample point here would be n good components. All right, the next one would be the first n minus 1 are good, the last one failed. All the way down to the very bottom, you can have all n fail.

All right, so how many sample points are there in the sample space with n components? 2 to

the n, all right? Because you've got n positions. There's two choices for each value there. Now how many of the sample points have exactly k failed components? Out of the 2 to the n, how many correspond to k of the components fail? n choose k-- now this goes back to counting. Remember the binomial coefficient.

So there are n choose k sample points have k failed components. All right, what's the probability of a sample point with k failed components? Well, I took k branches with a failure. Each of those gives me a p, p to the k. And I took n minus k branches with no failure. Each one of those multiplies by a 1 minus p. So no matter how I got my k failures, the probability for a particular sample point with k failures is just that, because I had p get factored in k times, in the failures, 1 minus p get factored in n minus k times for the situations where it worked out all right.

So I've got this many sample points. Each of them has probability p to the k times 1 minus p n minus k. Any questions about that, why that's the case? All right, so now I can compute the probability there are k failures. The probability that R equals k is simply n choose k, number of sample points times their probability, since they're all the same for k failures. And that, of course, is just the formula for the general binomial distribution. So that equals fnp of k, which is what we are trying to prove. OK. Any questions about that?

So that's why it's important, because there's a lot of situations where you're interested in the number of-- if you had n possible things, what's the chance k of them happen? And it's just given by this formula. Now you can calculate this, but, if I just told you, for example, maybe I'm looking at n is 100 and even p is 1/2, you know, what's the probability that we'd get 25-- say, what's the probably if getting 50 failed components?

Or if I flip 100 coins and they're fair. So probability half of getting a heads. What's the chance I actually get 50 heads in 100 coins? Looking at that, it's not so clear what the answer is. In fact, let's test intuition here, for this. We're going to take a little vote to see how good you are looking at that formula, or what your intuition is about the probability of getting exactly 50 heads. It could be between 1 and 1/2, a half and a tenth, a tenth and a hundredth, a hundredth and one thousandth, a thousandth and a millionth, and 0.

I want to know what's the probability, when you flip 100 mutually independent coins, you get exactly 50 heads? How many people think it's at least a half, that you get half heads? Nobody likes that. How many people think it's between a half and a tenth? One vote. A tenth and a

hundredth? More. Doesn't it have to be at least a hundredth? That's sort of the most likely outcome. There's only-- how many think a hundredth and a thousandth? All right. A thousandth and a millionth? Whoa, so you've got to believe it is not likely to get exactly 50 heads when I flip 100 coins.

All right, well the answer is actually between a tenth and a hundredth here. It's about 8%. And we're going to compute that. But here's another one. I flip 100 coins. What's the probability of getting at most 25 heads? So here, doesn't have to be exactly 25. It could be 25, 24, 23, 22, all the way down to no heads. So a lot of chances, all right? To get at most 25 heads. How many people think it's here? Some, yeah, you've got a lot of chances. How about here? Yeah, OK. What about between a tenth and a hundredth? One person, two people left. Between a hundredth and a thousandth? Nobody left? Anybody here? Thousandth and a millionth? One person is left, less than a million. 1 in a million? There's the contrarian.

And you're right, the chance of getting 25 or less heads is less than 1 in a million. So if somebody does it, his name better be Persi Diaconis. To get it consistently to 70, [INAUDIBLE] 25 or fewer heads. You get 75 tails or more is extremely unlikely, all right? So to see why that's the case, we've got to do a little work on this formula. Because we're saying that if we sum this up for k ending 100 and k being 0 to 25, and p being a half, it's an incredibly tiny number. So let's see why that is.

And this phenomenon is course going to be important in computer science because it's going to enable you to tell almost exactly how many components are going to fail in a system, if they're mutually independent. All right, now I'm not going to drag you through the math. It's bad enough I'm going to write it on the board. There's a bunch of this in the text. And instead of k, I'm going to represent by a parameter alpha times n to be the integer k. All right, this will replace k with a new parameter alpha.

It's at most, and also it's asymptotically equivalent using tilde-- both of those are true-- to this nasty looking expression, 2 to the alpha log p over alpha plus 1 minus alpha log 1 minus p over 1 minus alpha times and n-- that's a big number, it's 100-- over square root 2 pi alpha 1 minus alpha n. And this, of course, alpha is between 0 and 1.

Now when you do the derivative on this thing, it turns out its maximum value is when alpha equals p. And when you have alpha equals p, you get log of 1 is 0, log of 1 is 0. All this messy stuff goes away and we just get that. And so in that case, fnp of pn is at most, and also

asymptotically equal to 1 over square root 2 pi p 1 minus p n.

And now you can plug in values and compute things. For example, if I flip 100 coins, and I want to know the probability of getting 50-- that means alpha's a half, and they're fair, which means p is a half, the answer is 8%. So n equals 100 coins, p is a half, then the probability of 50 heads. I just plug in p is a half here, I get 1 over square root 50 pi, which equals 0.080 and so forth.

So there's an 8% chance of getting exactly 50 heads. Now let's look at the probability of getting 25 or fewer. And we'll see why that is so surprisingly small. OK, so for n equal 100 and p equal a half an alpha equal a quarter, because we want 25 heads, I'll get exactly 25 heads first. Probability of 25 heads is at most, well that square root thing comes out to about 0.09. Then I get 2 to something. I get 2 to the minus, all those logs come out the 0.1887 times n, which is the kicker, that's 100. So I get a 2 to the minus 18th here. And that makes this thing be smaller than or equal about to 1.913 times 10 to the minus 7. So about 1 in 5 million.

The reason this gets so small is because you get the n in that exponent up there. And of course, this is all computed using Stirling's formula, if you actually want to go through the calculations. You go start with n choose k, which is your n factorial, times over k factorial n minus k factorial. Plug in Stirling's formula and you do a bunch of messy stuff, and these things pop out. And so it gets exponentially small, exponentially fast.

In fact, if you were to plot this thing, if you plot the binomial distribution function, it looks something like this. All right, so I have 0 to n, and then I have pn, which is the maximum. Here's the probability. It goes 0 to 1. And your maximum value is here at 8% in the case of 100 coins, and it just zooms down exponential small. You can't even draw because it gets so small so fast. This height here is about 1 over root n. And this width here of the hump, is about root n. And these things zoom down to 0, exponentially fast.

And so that's what the binomial distribution looks like. And so it says that you are very likely to be very close to pn heads, or pn things happening. And I won't go through the math now-- probably do it in recitation tomorrow-- of computing at most 25 heads, which is the cumulative distribution function.

And this comes up in all sorts of places. Like you have a noisy communications channel, and every bit as dropped with 1% chance. If you have 10,000 bits, you'd like to know, what's the probability I lost 2% of them when I have 1% failure rate? The chance of losing 2% out of

10,000 is like 2 to the minus 60 if they're mutually independent. So no chance of losing 2%. All right, very good. We'll do more this in recitation tomorrow.