ALBERT MEYER: Today's topic is random variables. Random variables are an absolutely fundamental concept in probability theory. But before we get into officially defining them, let's start off with an example that in fact, is a game because that's a fun way to start.

So we're going to play the bigger number game and here's how it works. There are two teams, and Team 1 has the task of picking two different integers between 0 and 7 inclusive, and they write one integer on one piece of paper and the other integer on the other piece of paper. They turn the two pieces of paper face down so the numbers are not visible, and the other team then sees these two pieces of paper whose other side has different numbers written on them sitting on the table.

What Team 2 then does is picks one of the pieces of paper and turns it over and looks at the number on it. And then, based on what that number is, they make a decision, stick with the number they have or switch to the other unknown number on the face down piece of paper. And that'll be their final number. And the game is that Team 2 wins if they wind up with the larger number.

So they're going to look at the number on the paper that they expose and they're going to try to decide whether it looks like a big number or little number. If it looks like a big number, they'll stick with it. If it looks like a little number, they'll switch to the other one that they hope is larger.

So which team do you think has an advantage here? Course, if you've read the notes, you know. But if you haven't been exposed to this before, it's not really so obvious. And what we encourage and what we used to do when we ran this in real time in classes that we would have students in teams, split their team in half, one would be Team 1 and the other would be Team 2, and they'd play the game few times, see if they could figure out who had the advantage.

And if you have the opportunity, this might be a good moment to stop this video and try playing the game with some friends if they're around. Otherwise, let's just proceed and see how it all works.

So this is the strategy Team 2 is going to adopt. They're going to take this idea about big and small that I mentioned and act on it in a methodical way. So they're going to pick a paper to expose, giving each paper equal probability. So that guarantees that they have a 50/50 chance of picking the big number and a 50/50 chance of picking the little number. Whatever ingenuity Team 1 tried to do on which piece of paper was on the left and which was on the right, it doesn't really matter if Team 2 simply picks a piece of paper at random. There's no way that Team 1 can try to fake out Team 2 on where they put the number. OK.

The next step is that Team 2 is going to decide whether the number that they can see, the exposed number, is small. And if so, would they switch? And otherwise they stick. So that is, they're going to define some threshold Z where being less than or equal to Z means small, and being greater than Z means large. The question is, how do they choose Z?

Well, a naive thing to do would be to choose Z to be in the middle of the interval from 0 to 7. Let's say, you choose Z equals 3. So there would be four numbers less than or equal to Z and four numbers greater than Z. But of course, as soon as Team 1 knew that that was your Z, what would they do?

Well, they would make sure that both numbers were on the same side of Z. If you're Z was 3, they would always choose their numbers to be, say, 0 and 1. And that way, when you were switching, your Z would tell you that you had a small number, you should switch to the other one. And you'd only have a 50/50 chance of winning. So if you fixed that value of Z, Team 2 has a way of ensuring that you have no advantage.

You can only win with probability 50/50. And that's true no matter what Z you take. If Team 1 knew what your Z was, they would just make sure to pick their two numbers on the same side of your Z. And then your Z wouldn't really tell you anything. You'd switch or stick in both cases, and you'd only have a 50/50 chance of picking the right number.

So what you do-- and this is where probability comes in-- is you pick Z in a way that can't be predicted or made use of by Team 1. You pick Z at random, to be any number from 0 to 7, not including 7 including 0. That is, your number is either 0, 1, 2, up through 6. And being less than or equal to Z means small, and being greater than Z means large. And when you see a small number, you'll switch and when you see a large number, you'll stick. But what's going to be large and what's going to be small is going to vary each time you play the game, depending on what random number, Z, comes out to be.

So let's analyze your probability if you're Team 2. What's the probability that you're going to win now? Well, let's suppose that Team 1 picks these two numbers. We don't know what they are, but they have to pick a low number that's less than a high number. So these two numbers are at least 1 apart, they can't have the same number on both pieces of paper. Otherwise, it's clear that you are not going to be able to pick the large one, that would be cheating.

OK, so there's two different numbers. So one of them has to be less than the other. We don't know how much less, might be a lot less, might be only one less, but low is less than high. OK, now we can consider three cases of what happens with your strategy.

The most interesting case is the middle case. That is, when your Z, which was chosen at random, happens to fall in the interval between low and high. That is, your Z is strictly less than high and greater than or equal to low. And then in that case, your Z is really guiding you correctly on what to do.

If you turn over the low card, then it's going to look low because it's less than or equal to Z so you'll switch to the high card and win. If you turn over the high card, it's going to be greater than Z so it'll look high and you'll know to stick with it. So in this case, you're guaranteed to win. If you were lucky enough to guess the right threshold between low and high, you're going to win. And so the probability that you win, given the middle case occurs, is 1.

Now, what about the middle case? How often does that happen? Well, the difference between low and high is at least 1, so there's guaranteed to be 1 chance in 7 that your Z is going to fall between them. And it could be more if low and high are further apart, but as long as they're at least one apart, there's a 1/7 chance that you're going to fall in between them. OK.

Now, in case H, that's the case where Z happens to be chosen greater than or equal to the high number that Team 1 shows. In other words, Z is bigger than both numbers than Team 1 shows and put on the pieces of paper. Well, in that case, Z just isn't telling you anything. So what's going to happen is that both numbers are going to look high to you-- sorry-- both numbers are going to look low to you because they're both less than or equal to Z. So you'll switch.

And that means that you'll win, if and only if, you happen to turn the low card over first. Well that was 50/50. So the probability that you win, given that Z-- both cards are on the low side of Z, you'll win with half the time. And symmetrically, if Z is less than the low card, that is, Z is less than both cards chosen by Team 1, then they're both going to look high, and so you'll stick. And that means that you'll stick, you'll win, if and only if, you happen to have picked the high card. There's a 50/50 chance of that.

So again, in this case that Z makes both cards look high, or Z itself is low, Team 2, you win with probability 1/2. Well, that's great because now we can apply total probability. And what total probability tells us is that Team 2 wins is the probability that they win given case M times the probability of M plus the probability that they win given not the middle case times the probability of not the middle case. But we figured out what these were.

Well, at least inequalities on them, because there's probability 1 that you'll win 1/7 of the time. And there's probability a 1/2 that you'll win the rest of the time, the other 6/7 of the time. You're going to win 4/7 of the time. The probability that you win playing your strategy is 4/7. It's better than 50/50. You have an advantage. And whether that was a priori obvious or not, I don't know. But I think it's kind of cool. OK, you win with probability 4/7.

Now, Team 2 has the advantage. And the important thing to understand is it does not matter what team does. No matter how smart Team 1 is, Team 2 has gotten control of the situation because they picked-- which piece of paper they picked at random 50/50. So it doesn't matter what strategy Team 1 used on where they placed the numbers. And they chose Z randomly, so again, it doesn't matter what numbers Team 1 shows. Team 2 is still going to have their 1/7 chance of coming out ahead, which is enough to tip the balance in their favor.

It's interesting that symmetrically, Team 1 also has a random strategy that they can use, which guarantees that no matter what Team 2 does, Team 2 wins with probability at most 4/7. So either team can force the probability that Team 2 wins to be at most 4/7 and at least 4/7. So if they both play optimally, it's going to stay at 4/7. And that's again, true no matter what Team 2 does, Team 1 can put this upper bound to 4/7 on it. So essentially we can say that the value of this game, the probability that Team 2 wins is optimally for both is 4/7.

OK, now what does this game got to do with anything, with our general topic of random variables? Well, we'll be formal in a moment. But informally, a random variable is simply a number that's produced by a random process. And just to give an example before we come up with a formal definition, the threshold variable Z was a thing that took a value from 0 to 6 inclusive, each with probability 1/7. So it was producing a number by a random process, that chose a number at random with equal probability.

If Team 2 plays properly at random picking which piece of paper to expose, then the number of the exposed card, or more precisely, whether the exposed card is high or low, will also be a random variable. And if Team 1 plays optimally, the number on the exposed card is going to be a random variable. That is, Team 1 in their optimal strategy that puts an upper bound to 4/7 is in fact, going to choose the two numbers randomly.

So the exposed card is going to wind up being another random variable, a number produced by the random process. And likewise, the number of the larger card if Team 1 picks its larger and smaller cards randomly, it's going to be another example of a number produced by a random process. And likewise, the number of the smaller card.

So that's enough examples. This little game has a whole bunch of random variables appearing in it. And in the next segment, we will look again officially, what is the definition of a random variable?