

This tutorial is going to teach you about simulation. Now we can use simulation to figure out complicated problems like this one. Suppose a family has five children. What's the probability that at least two of those five are girls? Well, we could figure this out in a tree diagram, but that would be a pretty unwieldy tree diagram.

So one way to do it is to approximate the probability using a simulation. What we're going to do is we're going to get our dice out and let the boys and girls be equally likely, letting the faces one through three represent girls, and four through six represent boys. So here's our dice. We have five dice.

And we're going to-- like I said-- let one through three represent girls, and four through six represent boys. So this situation has two girls in it. The one and the two. Let's roll it again. This situation had four girls in it-- the two, the three, the three, and the one-- were four girls. We can redo this many, many times, and the relative frequency of getting two or more girls will approximate the probability.

So let's do this 20 times. So we've already done 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20. So now we're done with those dice, let's analyze the results.

Here's the first trial where we got a 6, 4, 1, 2, and 6. And we had two girls. These are all the rest of the dice rolls which are being tabulated behind the scenes. It looks like we had one situation where we didn't get any girls in our group of five.

This is the histogram of the number of girls per family. Once, we got no girls. I find it a little strange-- this is pretty unusual-- that in none of our simulations did we get exactly one girl. Kind of unusual. I wouldn't expect that to happen again.

The most common was two girls, then three and four girls was also very common. I wouldn't expect us to get four girls as often as we did if we had run this simulation again. And in none of the families were all five of the children girls.

But we got at least two girls in 19 of our 20 simulations. So our guess for the probability of having at least two girls in a family is 19 out of 20, or 95%. Our best guess is that we have a 95% probability of getting at least two girls in a family.

Now that's not in fact the right answer. If we had done more simulations, we would have gotten a more correct answer because things would have started to even out. The right answer is in fact 0.8125, about 81% of the time. Now the thing is, if we simulated more than 20 times with the dice, we would have gotten a more accurate response.

This graph shows an approximation based on 10,000 simulations. I won't show you all the simulations that we did, but just trust me that this is what the curve is supposed to look like. Our sample of 20 was not super representative. Apparently, four is fairly uncommon. Two and three are very common.

But in our sample of 10,000 families, 8,127 of the 10,000 had two girls or more. And so our best guess now as to the probability that a family has two or more girls is 0.8127, which is a lot closer to the 0.8125 that the correct answer is.

And so to recap, it's possible to solve these complicated problems through a process called simulation. This is also called the Monte Carlo method, named after Monte Carlo Casino because it uses physical objects like dice.

First, you come up with a way to simulate that has the same frequencies of what you're trying to predict. Like here, we had boys and girls. So we had to pick something that had a  $1/2$  probability, so we used three faces of the dice. Second, run many trials of your simulation. And third, answer the question based on the frequencies from your simulated results.

And so, we used simulation, also known as the Monte Carlo method, in order to obtain a guess at a probability for a complicated problem. Good luck. And we'll see you next time.