

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](http://ocw.mit.edu).

**JOHN GUTTAG:** Welcome to Lecture 6. As usual, I want to start by posting some relevant reading. For those who don't know, this lovely picture is of the Casino at Monte Carlo, and shortly you'll see why we're talking about casinos and gambling today. Not because I want to encourage you to gamble your life savings away.

A little history about Monte Carlo simulation, which is the topic of today's lecture. The concept was invented by the Polish American mathematician, Stanislaw Ulam. Probably more well known for his work on thermonuclear weapons than on mathematics, but he did do a lot of very important mathematics earlier in his life.

The story here starts that he was ill, recovering from some serious illness, and was home and was bored and was playing a lot of games of solitaire, a game I suspect you've all played. Being a mathematician, he naturally wondered, what's the probability of my winning this stupid game which I keep losing?

And so he actually spent quite a lot of time trying to work out the combinatorics, so that he could actually compute the probability. And despite being a really amazing mathematician, he failed. The combinatorics were just too complicated.

So he thought, well suppose I just play lots of hands and count the number I win, divide by the number of hands I played. Well then he thought about it and said, well, I've already played a lot of hands and I haven't won yet. So it probably will take me years to play enough hands to actually get a good estimate, and I don't want to do that.

So he said, well, suppose instead of playing the game, I just simulate the game on a computer. He had no idea how to use a computer, but he had friends in high places. And actually talked to John von Neumann, who is often viewed as the inventor of the stored program computer. And said, John, could you do this on your fancy new ENIAC machine?

And on the lower right here, you'll see a picture of the ENIAC. It was a very large machine. It

filled a room.

And von Neumann said, sure, we could probably do it in only a few hours of computation.

Today we would think of a few microseconds, but those machines were slow. Hence was born Monte Carlo simulation, and then they actually used it in the design of the hydrogen bomb. So it turned out to be not just useful for cards.

So what is Monte Carlo simulation? It's a method of estimating the values of an unknown quantity using what is called inferential statistics. And we've been using inferential statistics for the last several lectures. The key concepts-- and I want to be careful about these things will be coming back to them-- are the population.

So think of the population as the universe of possible examples. So in the case of solitaire, it's a universe of all possible games of solitaire that you could possibly play. I have no idea how big that is, but it's really big,

Then we take that universe, that population, and we sample it by drawing a proper subset. Proper means not the whole thing. Usually more than one sample to be useful. Certainly more than 0.

And then we make an inference about the population based upon some set of statistics we do on the sample. So the population is typically a very large set of examples, and the sample is a smaller set of examples.

And the key fact that makes them work is that if we choose the sample at random, the sample will tend to exhibit the same properties as the population from which it is drawn. And that's exactly what we did with the random walk, right?

There were a very large number of different random walks you could take of say, 10,000 steps. We didn't look at all possible random walks of 10,000 steps. We drew a small sample of, say 100 such walks, computed the mean of those 100, and said, we think that's probably a good expectation of what the mean would be of all the possible walks of 10,000 steps. So we were depending upon this principle.

And of course the key fact here is that the sample has to be random. If you start drawing the sample and it's not random, then there's no reason to expect it to have the same properties as that of the population. And we'll go on throughout the term, and talk about the various ways you can get fooled and think of a random sample when exactly you don't.

All right, let's look at a very simple example. People like to use flipping coins because coins are easy. So let's assume we have some coin. All right, so I bought two coins slightly larger than the usual coin. And I can flip it. Flip it once, and let's consider one flip, and let's assume it came out heads. I have to say the coin I flipped is not actually a \$20 gold piece, in case any of you were thinking of stealing it.

All right, so we've got one flip, and it came up heads. And now I can ask you the question-- if I were to flip the same coin an infinite number of times, how confident would you be about answering that all infinite flips would be heads? Or even if I were to flip it once more, how confident would you be that the next flip would be heads? And the answer is not very.

Well, suppose I flip the coin twice, and both times it came up heads. And I'll ask you the same question-- do you think that the next flip is likely to be heads? Well, maybe you would be more inclined to say yes and having only seen one flip, but you wouldn't really jump to say, sure.

On the other hand, if I flipped it 100 times and all 100 flips came up heads, well, you might be suspicious that my coin only has a head on both sides, for example. Or is weighted in some funny way that it mostly comes up heads. And so a lot of people, maybe even me, if you said, I flipped it 100 times and it came up heads. What do you think the next one will be? My best guess would be probably heads.

How about this one? So here I've simulated 100 flips, and we have 50 heads here, two heads here, And 48 tails. And now if I said, do you think that the probability of the next flip coming up heads-- is it 52 out of 100?

Well, if you had to guess, that should be the guess you make. Based upon the available evidence, that's the best guess you should probably make. You have no reason to believe it's a fair coin. It could well be weighted. We don't see it with coins, but we see weighted dice all the time. We shouldn't, but they exist. You can buy them on the internet.

So typically our best guess is what we've seen, but we really shouldn't have very much confidence in that guess. Because well, could've just been an accident. Highly unlikely even if the coin is fair that you'd get 50-50, right?

So why when we see 100 samples and they all come up heads do we feel better about guessing heads for the 101st than we did when we saw two samples? And why don't we feel

so good about guessing 52 out of 100 when we've seen a hundred flips that came out 52 and 48? And the answer is something called variance.

When I had all heads, there was no variability in my answer. I got the same answer all the time. And so there was no variability, and that intuitively-- and in fact, mathematically-- should make us feel confident that, OK, maybe that's really the way the world is.

On the other hand, when almost half are heads and almost half are tails, there's a lot of variance. Right, it's hard to predict what the next one will be. And so we should have very little confidence that it isn't an accident that it happened to be 52-48 in one direction. So as the variance grows, we need larger samples to have the same amount of confidence.

All right, let's look at that with a detailed example. We'll look at roulette in keeping with the theme of Monte Carlo simulation. This is a roulette wheel that could well be at Monte Carlo.

There's no need to simulate roulette, by the way. It's a very simple game, but as we've seen with our earlier examples, it's nice when we're learning about simulations to simulate things where we actually can know what the actual answer is so that we can then understand our simulation better.

For those of you who don't know how roulette is played-- is there anyone here who doesn't know how roulette is played? Good for you. You grew up virtuous. All right, so-- well all right. Maybe I won't go there.

So you have a wheel that spins around, and in the middle are a bunch of pockets. Each pocket has a number and a color. You bet in advance on what number you think is going to come up, or what color you think is going to come up. Then somebody drops a ball in that wheel, gives it a spin. And through centrifugal force, the ball stays on the outside for a while. But as the wheel slows down and heads towards the middle, and eventually settles in one of those pockets. And you win or you lose.

Now you can bet on it, and so let's look at an example of that. So here is a roulette game. I've called it fair roulette, because it's set up in such a way that in principle, if you bet, your expected value should be 0. You'll win some, you'll lose some, but it's fair in the sense that it's not either a negative or positive sum game.

So as always, we have an underbar underbar in it. Well we're setting up the wheel with 36 pockets on it, so you can bet on the numbers 1 through 36. That's way range work, you'll

recall. Initially, we don't know where the ball is, so we'll say it's none. And here's the key thing is, if you make a bet, this tells you what your odds are. That if you bet on a pocket and you win, you get [? len ?] of pockets minus 1.

So This is why it's a fair game, right? You bet \$1. If you win, you get \$36, your dollar plus \$35 back. If you lose, you lose.

All right, self dot spin will be random dot choice among the pockets. And then there is simply bet, where you just can choose an amount to bet and the pocket you want to bet on. I've simplified it. I'm not allowing you to bet here on colors.

All right, so then we can play it. So here is play roulette. I've made game the class a parameter, because later we'll look at other kinds of roulette games.

You tell it how many spins. What pocket you want to bet on. For simplicity, I'm going to bet on this same pocket all the time. Pick your favorite lucky number and how much you want to bet, and then we'll have a simulation just like the ones we've already looked at.

So the number you get right starts at 0. For I and range number of spins, we'll do a spin. And then tote pocket plus equal game dot that pocket. And it will come back either 0 if you've lost, or 35 if you've won. And then we'll just print the results. So we can do it. In fact, let's run it.

So here it is. I guess I'm doing a million games here, so quite a few. Actually I'm going to do two. What happens when you spin it 100 times? What happens when you spin it a million times? And we'll see what we get.

So what we see here is that we do 100 spins. The first time I did it my expected return was minus 100%. I lost everything I bet. Not so unlikely, given that the odds are pretty long that you could do 100 times without winning. Next time I did a 100, my return was a positive 44%, and then a positive 28%.

So you can see, for 100 spins it's highly variable what the expected return is. That's one of the things that makes gambling attractive to people. If you go to a casino, 100 spins would be a pretty long night at the table. And maybe you'd won 44%, and you'd feel pretty good about it.

What about a million spins? Well people aren't interested in that, but the casino is, right? They don't really care what happens with 100 spins. They care what happens with a million spins. What happens when everybody comes every night to play.

And there what we see is-- you'll notice much less variance. Happens to be minus 0.04 plus 0.6 plus 0.79. So it's still not 0, but it's certainly, these are all closer to 0 than any of these are. We know it should be 0, but it doesn't happen to be in these examples.

But not only are they closer to 0, they're closer together. There is much less variance in the results, right? So here I show you these three numbers, and ask what do you expect to happen? You have no clue, right? So I don't know, maybe I'll win a lot. Maybe I'll lose everything. I show you these three numbers, you're going to look at it and say, well you know, I'm going to be somewhere between around 0 and maybe 1%.

But you're never going to guess it's going to be radically different from that. And if I were to change this number to be even higher, it would go even closer to 0. But we won't bother.

OK, so these are the numbers we just looked at, because I said the seed to be the same. So what's going on here is something called the law of large numbers, or sometimes Bernoulli's law. This is a picture of Bernoulli on the stamp. It's one of the two most important theorems in all of statistics, and we'll come to the second most important theorem in the next lecture.

Here it says, "in repeated independent tests with the same actual probability, the chance that the fraction of times the outcome differs from p converges to 0 as the number of trials goes to infinity." So this says if I were to spin this fair roulette wheel an infinite number of times, the expected-- the return would be 0. The real true probability from the mathematics.

Well, infinite is a lot, but a million is getting closer to infinite. And what this says is the closer I get to infinite, the closer it will be to the true probability. So that's why we did better with a million than with a hundred. And if I did a 100 million, we'd do way better than I did with a million.

I want to take a minute to talk about a way this law is often misunderstood. This is something called the gambler's fallacy. And all you have to do is say, let's go watch a sporting event. And you'll watch a batter strike out for the sixth consecutive time. The next time they come to the plate, the idiot announcer says, well he struck out six times in a row. He's due for a hit this time, because he's usually a pretty good hitter.

Well that's nonsense. It says, people somehow believe that if deviations from expected occur, they'll be evened out in the future. And we'll see something similar to this that is true, but this is

not true.

And there is a great story about it. This is told in a book by [INAUDIBLE] and [INAUDIBLE].

And this truly happened in Monte Carlo, with Roulette. And you could either bet on black or red. Black came up 26 times in a row. Highly unlikely, right? 2 to the 26th is a giant number.

And what happened is, word got out on the casino floor that black had kept coming up way too often. And people more or less panicked to rush to the table to bet on red, saying, well it can't keep coming up black. Surely the next one will be red.

And as it happened when the casino totaled up its winnings, it was a record night for the casino. Millions of francs got bet, because people were sure it would have to even out. Well if we think about it, probability of 26 consecutive reds is that. A pretty small number. But the probability of 26 consecutive reds when the previous 25 rolls were red is what? No, that.

**AUDIENCE:** Oh, I thought you meant [INAUDIBLE].

**JOHN GUTTAG:** No, if you had 25 reds and then you spun the wheel once more, the probability of it having 26 reds is now 0.5, because these are independent events. Unless of course the wheel is rigged, and we're assuming it's not.

People have a hard time accepting this, and I know it seems funny. But I guarantee there will be some point in the next month or so when you will find yourself thinking this way, that something has to even out. I did so badly on the midterm, I will have to do better on the final. That was mean, I'm sorry.

All right, speaking of means-- see? Professor [? Grimm's ?] not the only one who can make bad jokes. There is something-- it's not the gambler's fallacy-- that's often confused with it, and that's called regression to the mean. This term was coined in 1885 by Francis Galton in a paper, of which I've shown you a page from it here.

And the basic conclusion here was-- what this table says is if somebody's parents are both taller than average, it's likely that the child will be smaller than the parents. Conversely, if the parents are shorter than average, it's likely that the child will be taller than average.

Now you can think about this in terms of genetics and stuff. That's not what he did. He just looked at a bunch of data, and the data actually supported this. And this led him to this notion of regression to the mean. And here's what it is, and here's the way in which it is subtly

different from the gambler's fallacy.

What he said here is, following an extreme event-- parents being unusually tall-- the next random event is likely to be less extreme. He didn't know much about genetics, and he kind of assumed the height of people were random. But we'll ignore that. OK, but the idea is here that it will be less extreme.

So let's look at it in roulette. If I spin a fair roulette wheel 10 times and get 10 reds, that's an extreme event. Right, here's a probability of basically 1.1024. Now the gambler's fallacy says, if I were to spin it another 10 times, it would need to even out. As in I should get more blacks than you would usually get to make up for these excess reds.

What regression to the mean says is different. It says, it's likely that in the next 10 spins, you will get fewer than 10 reds. You will get a less extreme event. Now it doesn't have to be 10. If I'd gotten 7 reds instead of 5, you'd consider that extreme, and you would bet that the next 10 would have fewer than 7. But you wouldn't bet that it would have fewer than 5.

Because of this, if you now look at the average of the 20 spins, it will be closer to the mean of 50% reds than you got from the extreme first spins. So that's why it's called regression to the mean. The more samples you take, the more likely you'll get to the mean. Yes?

**AUDIENCE:** So, roulette wheel spins are supposed to be independent.

**JOHN GUTTAG:** Yes.

**AUDIENCE:** So it seems like the second 10--

**JOHN GUTTAG:** Pardon?

**AUDIENCE:** It seems like the second 10 times that you spin it. that shouldn't have to [INAUDIBLE].

**JOHN GUTTAG:** Has nothing to do with the first one.

**AUDIENCE:** But you said it's likely [INAUDIBLE].

**JOHN GUTTAG:** Right, because you have an extreme event, which was unlikely. And now if you have another event, it's likely to be closer to the average than the extreme was to the average. Precisely because it is independent. That makes sense to everybody? Yeah?

**AUDIENCE:** Isn't that the same as the gambler's fallacy, then? By saying that, because this was super

unlikely, the next one [INAUDIBLE].

**JOHN GUTTAG:** No, the gambler's fallacy here-- and it's a good question, and indeed people often do get these things confused. The gambler's fallacy would say that the second 10 spins would-- we would expect to have fewer than 5 reds, because you're trying to even out the unusual number of reds in the first Spin

Whereas here we're not saying we would have fewer than 5. We're saying we'd probably have fewer than 10. That it'll be closer to the mean, not that it would be below the mean. Whereas the gambler's fallacy would say it should be below that mean to quote, even out, the first 10. Does that makes sense? OK, great questions. Thank you.

All right, now you may not know this, but casinos are not in the business of being fair. And the way they don't do that is in Europe, they're not all red and black. They sneak in one green. And so now if you bet red, well sometimes it isn't always red or black. And furthermore, there is this 0. They index from 0 rather than from one, and so you don't get a full payoff.

In American roulette, they manage to sneak in two greens. They have a 0 in a double 0. Tilting the odds even more in favor of the casino. So we can do that in our simulation. We'll look at European roulette as a subclass of fair roulette. I've just added this extra pocket, 0.

And notice I have not changed the odds. So what you get if you get your number is no higher, but you're a little bit less likely to get it because we snuck in that 0. Than American roulette is a subclass of European roulette in which I add yet another pocket.

All right, we can simulate those. Again, nice thing about simulations, we can play these games. So I've simulated 20 trials of 1,000 spins, 10,000 spins, 100,000, and a million. And what do we see as we look at this? Well, right away we can see that fair roulette is usually a much better bet than either of the other two. That even with only 1,000 spins the return is negative.

And as we get more and more as I got to a million, it starts to look much more like closer to 0. And these, we have reason to believe at least, are much closer to true expectation saying that, while you break even in fair roulette, you'll lose 2.7% in Europe and over 5% in Las Vegas, or soon in Massachusetts.

All right, we're sampling, right? That's why the results will change, and if I ran a different simulation with a different seed I'd get different numbers. Whenever you're sampling, you can't

be guaranteed to get perfect accuracy. It's always possible you get a weird sample.

That's not to say that you won't get exactly the right answer. I might have spun the wheel twice and happened to get the exact right answer of the return. Actually not twice, because the math doesn't work out, but 35 times and gotten exactly the right answer.

But that's not the point. We need to be able to differentiate between what happens to be true and what we actually know, in a rigorous sense, is true. Or maybe don't know it, but have real good reason to believe it's true. So it's not just a question of faith.

And that gets us to what's in some sense the fundamental question of all computational statistics, is how many samples do we need to look at before we can have real, justifiable confidence in our answer? As we've just seen-- not just, a few minutes ago-- with the coins, our intuition tells us that it depends upon the variability in the underlying possibilities.

So let's look at that more carefully. We have to look at the variation in the data. So let's look at first something called variance.

So this is variance of  $x$ . Think of  $x$  as just a list of data examples, data items. And the variance is we first compute the average of value, that's  $\mu$ . So  $\mu$  is for the mean. For each little  $x$  and big  $X$ , we compare the difference of that and the mean. How far is it from the mean? And square of the difference, and then we just sum them.

So this takes, how far is everything from the mean? We just add them all up. And then we end up dividing by the size of the set, the number of examples.

Why do we have to do this division? Well, because we don't want to say something has high variance just because it has many members, right? So this sort of normalizes is by the number of members, and this just sums how different the members are from the mean.

So if everything is the same value, what's the variance going to be? If I have a set of 1,000 6's, what's the variance? Yes?

**AUDIENCE:** 0.

**JOHN GUTTAG:** 0. You think this is going to be hard, but I came prepared. I was hoping this would happen. Look out, I don't know where this is going to go. [FIRES SLINGSHOT]

**AUDIENCE:** [LAUGHTER]

**JOHN GUTTAG:** All right, maybe it isn't the best technology. I'll go home and practice. And then the thing you're more familiar with is the standard deviation. And if you look at the standard deviation is, it's simply the square root of the variance.

Now, let's understand this a little bit and first ask, why am I squaring this here, especially because later on I'm just going to take a square root anyway? Well squaring it has one virtue, which is that it means I don't care whether the difference is positive or negative. And I shouldn't, right? I don't care which side of the mean it's on, I just care it's not near the mean.

But if that's all I wanted to do I could take the absolute value. The other thing we see with squaring is it gives the outliers extra emphasis, because I'm squaring that distance. Now you can think that's good or bad, but it's worth knowing it's a fact.

The more important thing to think about is standard deviation all by itself is a meaningless number. You always have to think about it in the context of the mean. If I tell you the standard deviation is 100, you then say, well-- and I ask you whether it's big or small, you have no idea.

If the mean is 100 and the standard deviation is 100, it's pretty big. If the mean is a billion and the standard deviation is 100, it's pretty small. So you should never want to look at just the standard deviation.

All right, here is just some code to compute those, easy enough. Why am I doing this? Because we're now getting to the punch line. We often try and estimate values just by giving the mean. So we might report on an exam that the mean grade was 80.

It's better instead of trying to describe an unknown value by it-- an unknown parameter by a single value, say the expected return on betting a roulette wheel, to provide a confidence interval. So what a confidence interval is is a range that's likely to contain the unknown value, and a confidence that the unknown value is within that range.

So I might say on a fair roulette wheel I expect that your return will be between minus 1% and plus 1%, and I expect that to be true 95% of the time you play the game if you play 100 rolls, spins. If you take 100 spins of the roulette wheel, I expect that 95% of the time your return will be between this and that.

So here, we're saying the return on betting a pocket 10 times, 10,000 times in European roulette is minus 3.3%. I think that was the number we just saw. And now I'm going to add to

that this margin of error, which is plus or minus 3.5% with a 95% level of confidence.

What does this mean? If I were to conduct an infinite number of trials of 10,000 bets each, my expected average return would indeed be minus 3.3%, and it would be between these values 95% of the time. I've just subtracted and added this 3.5, saying nothing about what would happen in the other 5% of the time. How far away I might be from this, this is totally silent on that subject. Yes?

**AUDIENCE:** I think you want 0.2 not 9.2.

**JOHN GUTTAG:** Oh, let's see. Yep, I do. Thank you. We'll fix it on the spot. This is why you have to come to lecture rather than just reading the slides, because I make mistakes. Thank you, Eric.

All right, so it's telling me that, and that's all it means. And it's amazing how often people don't quite know what this means. For example, when they look at a political poll and they see how many votes somebody is expected to get. And they see this confidence interval and say, what does that really mean? Most people don't know. But it does have a very precise meaning, and this is it.

How do we compute confidence intervals? Most of the time we compute them using something called the empirical rule. Under some assumptions, which I'll get to a little bit later, the empirical rule says that if I take the data, find the mean, compute the standard deviation as we've just seen, 68% of the data will be within one standard deviation in front of or behind the mean. Within one standard deviation of the mean.

95% will be within 1.96 standard deviations. And that's what people usually use. Usually when people talk about confidence intervals, they're talking about the 95% confidence interval. And they use this 1.6 number.

And 99.7% of the data will be within three standard deviations. So you can see if you are outside the third standard deviation, you are a pretty rare bird, for better or worse depending upon which side.

All right, so let's apply the empirical rule to our roulette game. So I've got my three roulette games as before. I'm going to run a simple simulation. And the key thing to notice is really this print statement here.

Right, that I'll print the mean, which I'm rounding. And then I'm going to give the confidence

intervals, plus or minus, and I'll just take the standard deviation times 1.6 times 100, y times 100, because I'm showing you percentages.

All right so again, very straightforward code. Just simulation, just like the ones we've been looking at. And well, I'm just going-- I don't think I'll bother running it for you in the interest of time. You can run it yourself. But here's what I got when I ran it.

So when I simulated betting a pocket for 20 trials, we see that the-- of 1,000 spins each, for 1,000 spins the expected return for fair roulette happened to be 3.68%. A bit high. But you'll notice the confidence interval plus or minus 27 includes the actual answer, which is 0. And we have very large confidence intervals for the other two games.

If you go way down to the bottom where I've spun, spun the wheel many more times, what we'll see is that my expected return for fair roulette is much closer to 0 than it was here. But more importantly, my confidence interval is much smaller, 0.8. So now I really have constrained it pretty well.

Similarly, for the other two games you will see-- maybe it's more accurate, maybe it's less accurate, but importantly the confidence interval is smaller. So I have good reason to believe that the mean I'm computing is close to the true mean, because my confidence interval has shrunk. So that's the really important concept here, is that we don't just guess-- compute the value in the simulation. We use, in this case, the empirical rule to tell us how much faith we should have in that value.

All right, the empirical rule doesn't always work. There are a couple of assumptions. One is that the mean estimation error is 0. What is that saying? That I'm just as likely to guess high as gas low.

In most experiments of this sort, most simulations, that's a very fair assumption. There's no reason to guess I'd be systematically off in one direction or another. It's different when you use this in a laboratory experiment, where in fact, depending upon your laboratory technique, there may be a bias in your results in one direction.

So we have to assume that there's no bias in our errors. And we have to assume that the distribution of errors is normal. And we'll come back to this in just a second. But this is a normal distribution, called the Gaussian. Under those two assumptions the empirical rule will always hold.

All right, let's talk about distributions, since I just introduced one. We've been using a probability distribution. And this captures the notion of the relative frequency with which some random variable takes on different values.

There are two kinds. , Discrete and these when the values are drawn from a finite set of values. So when I flip these coins, there are only two possible values, head or tails. And so if we look at the distribution of heads and tails, it's pretty simple. We just list the probability of heads. We list the probability of tails. We know that those two probabilities must add up to 1, and that fully describes our distribution.

Continuous random variables are a bit trickier. They're drawn from a set of reals between two numbers. For the sake of argument, let's say those two numbers are 0 and 1. Well, we can't just enumerate the probability for each number.

How many real numbers are there between 0 and 1? An infinite number, right? And so I can't say, for each of these infinite numbers, what's the probability of it occurring? Actually the probability is close to 0 for each of them. Is 0, if they're truly infinite.

So I need to do something else, and what I do that is what's called the probability density function. This is a different kind of PDF than the one Adobe sells. So there, we don't give the probability of the random variable taking on a specific value. We give the probability of it lying somewhere between two values. And then we define a curve, which shows how it works.

So let's look at an example. So we'll go back to normal distributions. This is-- for the continuous normal distribution, it's described by this function. And for those of you who don't know about the magic number e, this is one of many ways to define it.

But I really don't care whether you remember this. I don't care whether you know what e is. I don't care if you know what this is. What we really want to say is, it looks like this. In this case, the mean is 0. It doesn't have to be 0. I've [INAUDIBLE] a mean of 0 and a standard deviation of 1. This is called the so-called standard normal distribution.

But it's symmetric around the mean. And that gets back to, it's equally likely that our errors are in either direction, right? So it peaks at the mean. The peak is always at the mean. That's the most probable value, and it's symmetric about the mean.

So if we look at it, for example, and I say, what's the probability of the number being between

0 and 1? I can look at it here and say, all right, let's draw a line here, and a line here. And then I can integrate the curve under here. And that tells me the probability of this random variable being between 0 and 1.

If I want to know between minus 1 and 1. I just do this and then I integrate over that area. All right, so the area under the curve in this case defines the likelihood. Now I have to divide and normalize to actually get the answer between 0 and 1.

So the question is, what fraction of the area under the curve is between minus 1 and 1? And that will tell me the probability. So what does the empirical rule tell us? What fraction is between minus 1 and 1, roughly? Yeah? 68%, right?

So that tells me 68% of the area under this curve is between minus 1 and 1, because my standard deviation is 1, roughly 68%. And maybe your eyes will convince you that's a reasonable guess.

OK, we'll come back and look at this in a bit more detail on Monday of next week. And also look at the question of, why does this work in so many cases where we don't actually have a normal distribution to start with?