

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 view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](https://ocw.mit.edu).

**PROFESSOR:**

And today it's me, back again. And we'll study continuous types of stochastic processes. So far we were discussing discrete time processes. We studied the basics like variance expectation, all this stuff-- moments, moment generating function, and some important concepts for Markov chains, and Martingales.

So I'm sure a lot of you would have forgot about what Martingale and Markov chains were, but try to review this before the next few lectures. Because starting next week when we start discussing continuous types of stochastic processes-- not from me. You're not going to hear Martingale from me that much. But from people-- say, outside speakers-- they're going to use this Martingale concept to do pricing.

So I will give you some easy exercises. You will have some problems on Martingales. Just refer back to the notes that I had like a month ago, and just review. It won't be difficult problems, but try to make the concept comfortable. OK.

And then Peter taught some time series analysis. Time series is just the same as discrete time process. And regression analysis, this was all done on discrete time. That means the underlying space was  $x_1, x_2, x_3, \dots, x_t$ .

But now we're going to talk about continuous time processes. What are they? They're just a collection of random variables indexed by time. But now the time is a real variable. Here, time was just in integer values. Here, we have real variable. So a stochastic process develops over time, and the time variable is continuous now.

It doesn't necessarily mean that the process to solve this continuous-- it may as well look like these jumps. It may as well have a lot of jumps like this. It just means that the underlying time variable is continuous. Whereas when it was discrete time, you were only looking at specific observations at some times. I'll draw it here. Discrete

time looks more like that. OK.

So the first difficulty when you try to understand continuous time stochastic processes when you look at it is, how do you describe the probability distribution? How to describe the probability distribution? So let's go back to discrete time processes.

So the universal example was a simple random walk. And if you remember how we described it was  $x_t$  minus  $x_{t-1}$ , was either 1 or minus 1, probability half each. This was how we described it. And if you think about it, this is a slightly indirect way of describing the process.

You're not describing the probability of this process following this path cycle, cycle path. Instead what you're doing is, you're describing the probability of this event happening. From time  $t$  to  $t+1$ , what is the probability that it will go down? And at each step you describe the probability altogether, when you combine them, you get the probability distribution over the process. But you can't do it for continuous time, right?

The time variable is continuous so you can't just take intervals  $t$  and interval  $t'$  and describe the difference. If you want to do that, you have to do it infinitely amount of times. You have to do it for all possible values. That's the first difficulty. Actually, that's the main difficulty.

And how can we handle this? It's not an easy question. And you'll see a very indirect way to handle it. It's somewhat in the spirit of this thing. But it's not like you draw some path to describe a probability density of this path. That's the  $\omega$ .

What is the probability density at  $\omega$ ? Of course, it's not a discrete variable so you have a probability density function, not a probability mass function. In fact, can we even write it down? You'll later see that we won't even be able to write this down. So just have this in mind and you'll see what I was trying to say.

So finally, I get to talk about Brownian processes, Brownian motion. Some outside speakers already started talking about it. I wish I already was able to cover it before they talked about it, but you'll see a lot more from now. And let's see what it actually

is.

So it's described as the following, it actually follows from a theorem. There exists a probability distribution over the set of continuous functions from real positive reals to the reals such that first,  $B_0$  is always 0. So probability of  $B_0$  is equal to  $B_0$  is 1.

Number two-- we call this stationary. For all  $s$  and  $t$ ,  $B_t$  minus  $B_s$  has normal distribution with mean 0 and variance  $t$  minus  $s$ . And the third-- independent increment. That means if intervals  $s_i$   $t_i$  are not overlapping then  $B_{t_i}$  minus  $B_{s_i}$  are independent.

So it's actually a theorem saying that there is some strange probability distribution over the continuous functions from positive reals-- non-negative reals-- to the reals. So if you look at some continuous function, this theorem gives you a probability distribution. It describes the probability of this path happening. It doesn't really describe it. It just says that there exists some distribution such that it always starts at 0 and it's continuous.

Second, the distribution for all fixed  $s$  and  $t$ , the distribution of this difference is normally distributed with mean 0 and variance  $t$  minus  $s$ , which scales according to the time. And then third, independent increment means what happened between this interval--  $s_1, t_1$  and  $s_2, t_2$ . This part and this part is independent as long as intervals do not overlap.

It sounds very similar to the simple random walk. But the reason we have to do this very complicated process is because the time is continuous. You can't really describe at each time what's happening. Instead, what you're describing is over all possible intervals what's happening.

When you have a fixed interval, it describes the probability distribution. And then when you have several intervals, as long as they don't overlap, they're independent. OK? And then by this theorem, we call this probability distribution a Brownian motion.

So probability distribution, the definition, distribution given by this theorem is called the Brownian motion. That's why I'm saying it's indirect. I'm not saying Brownian motion is this probability distribution. It satisfies these conditions, but we are reversing it.

Actually, we have these properties in mind. We're not sure if such a probability distribution even exists or not. And actually this theorem is very, very difficult. I don't know how to prove it right now. I have to go through a book. And even graduate probability courses usually don't cover it because it's really technical.

That means this just shows how continuous time stochastic processes can be so much more complicated than discrete time. Then why are you why are we studying continuous time processes when it's so complicated? Well, you'll see in the next few lectures. Any questions? OK.

So let's go through this a little bit more.

**AUDIENCE:** Excuse me.

**PROFESSOR:** Yes.

**AUDIENCE:** So when you talk about the probability distribution, what's the underlying space? Is it the space of--

**PROFESSOR:** Yes, that's a very good question. The space is the space of all functions. That means it's a space of all possible paths, if you want to think about it this way. Just think about all possible ways your variable can evolve over time. And for some fixed drawing for this path, there's some probability that this path will happen.

It's not the probability spaces that you have been looking at. It's not one point-- well, a point is now a path. And your probability distribution is given over paths, not for a fixed point. And that's also a reason why it makes it so complicated. Other questions?

So the main thing you have to remember-- well, intuitively you will just know it. But one thing you want to try to remember is this property. As your time scales, what

happens between that interval is it's like a normal variable.

So this is a collection of a bunch of normal variables. And the mean is always 0, but the variance is determined by the length of your interval. Exactly that will be the variance. So try to remember this property.

A few more things, it is a lot of different names. It's also called minor process. And let's see, there was one more. Is there another name for it? I thought I had one more name in mind, but maybe not.

**AUDIENCE:** [INAUDIBLE] was an MIT professor.

**PROFESSOR:** Oh, yeah. That's important.

**AUDIENCE:** Of course.

**PROFESSOR:** Yeah, a professor at MIT. But apparently he wasn't the first person who discovered this process. I was some other person in 1900. And actually, in the first paper that appeared, of course, they didn't know about each other's result. In that paper the reason he studied this was to evaluate stock prices and auction prices.

And here's another slightly different description, maybe a more intuitive description of the Brownian motion. So here is this philosophy. Philosophy is that Brownian motion is the limit of simple random walks. The limit-- it's a very vague concept. You'll see what I mean by this.

So fix a time interval of 0 up to 1 and slice it into very small pieces. So I'll say, into  $n$  pieces.  $1$  over  $n$ ,  $2$  over  $n$ ,  $3$  over  $n$ , dot dot dot, to minus  $1$  over  $n$ . And consider a simple random walk,  $n$  step simple random walk. So from time 0 you go up or down, up or down. Then you get something like that. OK?

So let me be a little bit more precise. Let  $y_0, y_1, \dots, y_n$ , be a simple random walk, and let  $z$  be the function such that at time  $t$  over  $n$ , we let it to be  $y$  of  $t$ . That's exactly just written down in formula what it means. So this process is  $z$ . I take a simple random walk and scale it so that it goes from time 0 to time 1.

And then in the intermediate values-- for values that are not this, just linearly extended-- linearly extend in intermediate values. It's a complicated way of saying just connect the dots. And take  $n$  to infinity. Then the resulting distribution is a Brownian motion.

So mathematically, that's just saying the limit of simple random walks is a Brownian motion. But it's more than that. That means if you have some suspicion that some physical quantity follows a Brownian motion, and then you observe the variable at discrete times at very, very fine scales-- so you observe it really, really often, like a million times in one second.

Then once you see-- if you see that and take it to the limit, it looks like a Brownian motion. Then now you can conclude that it's a Brownian motion. What I'm trying to say is this continuous time process, whatever the strange thing is, it follows from something from a discrete world. It's not something new. It's the limit of these objects that you already now.

So this tells you that it might be a reasonable model for stock prices because for stock prices, no matter how-- there's only a finite amount of time scale that you can observe the prices. But still, if you observe it infinitely as much as you can, and the distribution looks like a Brownian motion, then you can use a Brownian motion to model it. So it's not only the theoretical observation. It also has implication when you want to use Brownian motion as a physical model for some quantity.

It also tells you why Brownian motion might appear in some situations. So here's an example. Here's a completely different context where Brownian motion was discovered, and why it has the name Brownian motion. So a botanist-- I don't know if I'm pronouncing it correctly-- named Brown in the 1800s, what he did was he observed a pollen particle in water.

So you have a couple of water and there's some pollen. Of course you have gravity that pulls the pollen down. And pollen is heavier than water so eventually it will go down, eventually. But that only explains the vertical action, it will only go down.

But in fact, if you observe what's happening, it just bounces back and forth crazily until it finally reaches down the bottom of your cup. And this motion, if you just look at a two-dimension picture, it's a Brownian motion to the left and right. So it moves as according to Brownian motion.

Well, first of all, I should say a little bit more. What Brown did was he observed it. He wasn't able to explain the horizontal actions because he only understood gravity, but then people tried to explain it.

They suspected that it was the water molecules that caused this action, but weren't able to really explain it. But the first person to actually rigorously explain it was, surprisingly, Einstein, that relativity guy, that famous guy. So I was really surprised. He's really smart, apparently.

And why? So why will this follow a Brownian motion? Why is it a reasonable model? And this gives you a fairly good reason for that. This description, where it's the limit of simple random walks. Because if you think about it, what's happening is there is a big molecule that you can observe, this big particle. But inside there's tiny water molecules, tiny ones that don't really see, but it's filling the space. And they're just moving crazily.

Even though the water looks still, what's really happening is these water molecules are just crazily moving inside the cup. And each water molecule, when they collide with the pollen, it will change the action of the pollen a little bit, by a tiny amount. So if you think about each collision as one step, then each step will either push this pollen to the left or to the right by some tiny amount. And it just accumulates over time.

So you're looking at a very, very fine time scale. Of course, the times will differ a little bit, but let's just forget about it, assume that it's uniform. And at each time it just pushes to the left or right by a tiny amount. And you look at what accumulates, as we saw, the limit of a simple random walk is a Brownian motion. And that tells you why we should get something like a Brownian motion here.

So the action of pollen particle is determined by infinitesimal-- I don't know if that's the right word-- but just, quote, "infinitesimal" interactions with water molecules. That explains, at least intuitively, why it follows Brownian motion.

And the second example is-- any questions here-- is stock prices. At least to give you some reasonable reason, some reason that Brownian motion is not so bad a model for stock prices. Because if you look at a stock price,  $S$ , the price is determined by buying actions or selling actions.

Each action kind of pulls down the price or pulls up the price, pushes down the price or pulls up the price. And if look at very, very tiny scales, what's happening is at a very tiny amount they will go up or down. Of course, it doesn't go up and down by a uniform amount, but just forget about that technicality. It just bounces back and forth infinitely often, and then you're taking these tiny scales to be tinier, so very, very small.

So again, you see this limiting picture. Where you have a discrete-- something looking like a random walk, and you take  $t$  to infinity. So if that's the only action causing the price, then Brownian motion will be the right model to use. Of course, there are many other things involved which makes this deviate from Brownian motions, but at least, theoretically, it's a good starting point. Any questions? OK.

So you saw Brownian motion. You already know that it's used in the financial market a lot. It's also being used in science and other fields like that. And really big names, like Einstein, is involved. So it's a really, really important theoretical thing.

Now that you've learned it, it's time to get used to it. So I'll tell you some properties, and actually prove a little bit-- just some propositions to show you some properties. Some of them are quite surprising if you never saw it before.

OK. So here are some properties. Crosses the  $x$ -axis infinitely often, or I should say the  $t$ -axis. Because you start from 0, it will never go to infinity, or get to negative infinity. It will always go balanced positive and negative infinitely often.

And the second, it does not deviate too much from  $t$  equals  $y$  squared. We'll call this



y. Now, this is a very vague statement. What I'm trying to say is to draw this curve is this. If you start at time 0, at some time  $t_0$ , the probability distribution here is given as a normal random variable with mean 0 and variance  $t_0$ . And because of that, the standard deviation is square root  $t_0$ .

So the typical value will be around the standard deviation. And it won't deviate. It can be 100 times this. It won't really be a million times that or something. So most likely it will look something like that. So it plays around this curve a lot, but it crosses the axis infinitely often. It goes back and forth. What else?

The third one is quite really interesting. It's more theoretical interest, but it also has real life implications. It's not differentiable anywhere. It's nowhere differentiable. So this curve, whatever that curve is, it's a continuous path, but it's nowhere differentiable, really surprising.

It's hard to imagine even one such path. What it's saying is if you take one path according to this probability distribution, then more than likely you'll obtain a path which is nowhere differentiable. That just sounds nice, but why it does it matter?

It matters because we can't use calculus anymore. Because all the theory of calculus is based on differentiation. However, our paths have some nice things, it's universal, and it appears in very different contexts. But if you want to do analysis on it, it's just not differentiable.

So the standard tools of calculus can't be used here, which is quite unfortunate if you think about it. You have this nice model, which can describe many things, you can't really do analysis on it. We'll later see that actually there is a variant, a different calculus that works. And I'm sure many of you would have heard about it. It's called Ito's calculus.

So we have this nice object. Unfortunately, it's not differentiable, so the standard calculus does not work here. However, there is a modified version of calculus called Ito's calculus, which extends the classical calculus to this setting. And it's really powerful and it's really cool. But unfortunately, we don't have that much time to

cover it.

I will only be able to tell you really basic properties and basic computations of it. And you'll see how this calculus is being used in the financial world in the coming up lectures. But before going into Ito's calculus, let's talk about the property of Brownian motion a little bit because we have to get used to it.

Suppose I'm using it as a model of a stock price. So I'm using-- use Brownian motion as a model for stock price-- say, daily stock price. The market opens at 9:30 AM. It closes at 4:00 PM. It starts at some price, and then moves according to the Brownian motion.

And then you want to obtain the distribution of the min value and the max value for the stock. So these are very useful statistics. So a daily stock price, what will the minimum and the maximum-- what will the distribution of those be? So let's compute it. We can actually compute it.

What we want to do is-- I'll just compute the maximum. I want to compute this thing over  $s$  smaller than  $t$  of the Brownian motion. So I define this new process from the Brownian motion, and I want to compute the distribution of this new stochastic process. And here's the theorem. So for all  $t$ , the probability that you have  $M_t$  greater than  $a$  and positive  $a$  is equal to 2 times the probability that you have the Brownian motion greater than  $a$ .

It's quite surprising. If you just look at this, there's no reason to expect that such a nice formula should exist at all. And notice that maximum is always at least 0, so we don't have to worry about negative values. It starts at 0.

How do we prove it? Proof. Take this  $\tau$ . It's a stopping time, if you remember what it is. It's a minimum value of  $t$ . So the Brownian motion at time  $t$  is equal to  $a$ . That's a complicated way of saying, just record the first time you hit the line  $a$ . Line  $a$ , with some Brownian motion, and you record this time. That will be your  $\tau$  of  $a$ .

So now here's some strange thing. The probability that  $B_t, B_{\tau a}$ , given this-- OK. So what this is saying is, if you're interested at time  $t$ , if your  $\tau a$  happened before

time  $t$ , so if your Brownian motion hit the line  $a$  before time  $t$ , then afterwards you have the same probability of ending up above  $a$  and ending up below  $a$ .

The reason is because you can just reflect the path. Whatever path that ends over  $a$ , you can reflect it to obtain a path that ends below  $a$ . And by symmetry, you just have this property. Well, it's not obvious how you'll use this right now.

And then we're almost done. The probability that maximum at time  $t$  is greater than  $a$  that's equal to the probability that you're stopping time is less than  $t$ , just by definition. And that's equal to the probability that  $B_t - B_{\tau_a}$  is positive given  $\tau_a$ .

Because if you know that  $\tau_a$  is less than  $t$ , there's only two possible ways. You can either go up afterwards, or you can go down afterwards. But these two are the same probability. What you obtain is 2 times the probability that-- and that's just equal to 2 times the probability that  $B_t$  is greater than  $a$ .

What happened? Some magic happened. First of all, these two are the same because of this property by symmetry. Then from here to here,  $B_{\tau_a}$  is always equal to  $a$ , as long as  $\tau_a$  is less than  $t$ . This is just-- I rewrote this as  $a$ , and I got this thing. And then I can just remove this because if I already know that  $\tau_a$  is less than  $t$ -- order is reversed.

If I already know that  $B$  at time  $t$  is greater than  $a$ , then I know that  $\tau_a$  is less than  $t$ . Because if you want to reach  $a$  because of continuity, if you want to go over  $a$ , you have to reach  $a$  at some point. That means you hit  $a$  before time  $t$ .

So that event is already inside that event. And you just get rid of it. Sorry, all this should be-- something looks weird. Not conditioned. OK. That makes more sense. Just the intersection of two properties. Any questions here?

So again, you just want to compute the probability that the maximum is greater than  $a$  at time  $t$ . In other words, just by definition of  $\tau_a$ , that's equal to the problem that  $\tau_a$  is less than  $t$ . And if  $\tau_a$  is less than  $t$ , afterwards, depending on afterwards what happens, it increases or decreases. So there's only two possibilities. It

increases or it decreases.

But these two events have the same probability because of this property. Here's a bar and that's an intersection. But it doesn't matter because if you have the  $B$  of  $x_1$  bar  $y$  equals  $B$  of  $x_2$  bar  $y$  then probability of  $x_1$  intersection  $y$  over probability of  $y$  is equal to-- these two cancel.

So this bar can just be replaced by intersection. That means these two events have the same probability. So you can just take one. What I'm going to take is one that goes above 0. So after  $\tau$  a it accumulates more value.

And if you rewrite it, what that means is just  $B_t$  is greater than  $a$  given that how  $\tau$  a is less than  $t$ . But now that just became redundant. Because if you already know that  $B_t$  is greater than  $a$ ,  $\tau$  a has to be less than  $t$ . And that's just the conclusion. And it's just some nice result about the maximum over some time interval. And actually, I think Peter uses distribution in your lecture, right?

**AUDIENCE:** Yes. [INAUDIBLE] is that the distribution of the max minus the movement of the Brownian motion. And use that range of the process as a scaling for [INAUDIBLE] and get more precise measures of volatility than just using, say, the close price [INAUDIBLE].

**PROFESSOR:** Yeah. That was one property. And another property is-- and that's what I already told you, but I'm going to prove this. So at each time the Brownian motion is not differentiable is that time with probability equal to 1. Well, not very strictly, but I will use this theorem to prove it. OK?

Suppose the Brownian motion has a differentiation at time  $t$  and it's equal to  $a$ . Then what you just see is that the Brownian motion at time  $t$  plus  $\epsilon$ , minus Brownian motion at time  $t$ , has to be less than or equal to  $\epsilon$  times  $a$ . Not precisely, so I'll say just almost, to make it mathematically rigorous.

But what I'm trying to say here is by-- is it mean value theorem? So from  $t$  to  $t$  plus  $\epsilon$ , you expect to gain  $a$  times  $\epsilon$ . That's-- OK? You should have this--

then. In fact, for all  $\epsilon$ . Greater than  $\epsilon'$ . Let's write it like that.

So in other words, the maximum in this interval  $B_t$  plus  $\epsilon$  minus  $t$ , this distribution is the same as the maximum at  $\epsilon'$ . That has to be less than  $\epsilon$  times  $a$ . So what I'm trying to say is if this differentiable, depending on the slope, your Brownian motion should have always been inside this cone from  $t$  up to time  $t$  plus  $\epsilon$ .

If you draw this slope, it must have been inside this cone. I'm trying to say that this cannot happen. From here to here, it should have passed this line at some point. OK?

So to do that I'm looking at the distribution of the maximum value over this time interval. And I want to say that it's even greater than that. So if your maximum is greater than that, you definitely can't have this control.

So if differentiable, then maximum of  $\epsilon'$ -- the maximum of  $\epsilon$ , actually, and just compute it. So the probability that  $M_\epsilon$  is less than  $\epsilon a$  is equal to 2 times the probability of that, the Brownian motion at  $\epsilon$  is less than or equal to  $a$ . This has normal distribution. And if you normalize it to  $N(0, 1)$ , divide by the standard deviation until you get the square root of  $\epsilon a$ .

As  $\epsilon$  goes to 0, this goes to 0. That means this goes to half. The whole thing goes to 1. What am I missing? I did something wrong. I flipped it. This is greater. Now, if you combine it, if it was differentiable, your maximum should have been than  $\epsilon a$ . But what we saw here is your maximum is always greater than that  $\epsilon$  times  $a$ . With probability 1, you take  $\epsilon$ . It goes to 0.

Any questions? OK. So those are some interesting things, properties of Brownian motion that I want to talk about. I have one final thing, and this one it's really important theoretically. And also, it will be the main lemma for Ito's calculus.

So the theorem is called quadratic variation. And it's something that doesn't happen that often. So let 0-- let me write it down even more clear. Now that's something strange. Let me just first parse it before proving it. Think about it as just a function,

function  $f$ . What is this quantity?

This quantity means that from 0 up to time  $t$ , you chop it up into  $n$  pieces. You get  $t$  over  $n$ ,  $2t$  over  $n$ ,  $3t$  over  $n$ , and you look at the function. The difference between each consecutive points record these differences and then square it.

And you sum it as  $n$  goes to infinity. So you take smaller and smaller scales take it to infinity. What the theorem says is for Brownian motion this goes to  $T$ , the limit. Why is this something strange?

Assume  $f$  is a lot better function. Assume  $f$  is continuously differentiable. That means it's differentiable, and its differentiation is continuous. Derivative is continuous. Then let's compute the exact same property, exact same thing. I'll just call this-- maybe it will be better.

This time  $t_i$  and time  $t_{i-1}$ , then the sum over  $i$  of  $f(t_i) - f(t_{i-1})$ . If you square it, this is at most sum from  $i=1$  to  $n$ ,  $(f(t_i) - f(t_{i-1}))^2$ . By mean value theorem--  $f'(s_i)$ .

So by mean value theorem, there exists a point  $s_i$  such that  $f(t_i) - f(t_{i-1})$  is equal to  $f'(s_i) \cdot (t_i - t_{i-1})$ . Yes. And then you take this term out. You take the maximum from 0 up to  $t$  of  $f'(s)^2$ , times  $i=1$  to  $n$ ,  $(t_i - t_{i-1})^2$ . This thing is  $t$  over  $n$  because we chopped it up into  $n$  intervals. Each consecutive difference is  $t$  over  $n$ .

If you square it, that's equal to  $t^2$  over  $n^2$ . If you had  $n$  of them, you get  $t^2$  over  $n$ . So you get whatever that maximum is times  $t^2$  over  $n$ . If you take  $n$  times infinity, that goes to 0. So if you have a reasonable function, which is differentiable, this variation, this is called a quadratic variation. Quadratic variation is 0.

So all these classical functions that you've been studying will not even have this quadratic variation. But for Brownian motion, what's happening is it just bounced back and forth too much. Even if you scale it smaller and smaller, the variation is big enough to accumulate. They won't disappear like if it was a differential function.

And that pretty much-- it's a slightly stronger version than this that it's not differentiable. We saw that it's not differentiable. And this a different way of saying that it's not differential. It has very important implications.

And another way to write it is-- so here's a difference of  $B$ , it's  $dB$  squared is equal to  $dt$ . So if you take the differential-- whatever that means-- if you take the infinitesimal difference of each side, this part is just  $dB$  squared. The Brownian motion difference squared, this part is  $d$  of  $t$ . And that we'll see again. But before that, let's just prove this theorem.

So we're looking at the sum of  $B$  of  $t_i$  plus 1, minus  $B$  of  $t_i$ , squared. Where  $t_i$  is  $i$  over  $n$  times the time. From 1 to  $n$ , 0 to  $n$  minus 1. OK. What's the distribution of this?

**AUDIENCE:** Normal.

**PROFESSOR:** Normal, meaning 0, variance  $t_i$  plus 1 minus  $t_i$ . But that one's just  $t$  over  $n$ . It's the distribution. So I'll write it like this. Your sum from  $i$  equal 1 to  $n$  minus 1,  $x_i$  squared for  $x_i$  is normal variable. OK?

And what's the expectation of  $x_i$  squared? It's  $t$  squared over  $n$  squared. OK. So maybe it's better to write it like this. So I'll just write it again-- the sum from  $i$  equals 0 to  $n$  minus 1 of random variable  $y_i$ . [INAUDIBLE] expectation of  $y_i$ .

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** Did I make a mistake somewhere?

**AUDIENCE:** The expected value of  $x_i$  squared is the variance.

**PROFESSOR:** It's  $t$  over  $n$ . Oh, yeah, you're right. Thank you. OK. So divide by  $n$  and multiply by  $n$ . What is this? What will this go to?

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** No. Remember strong law of large numbers. You have a bunch of random

variables, which are independent, identically distributed, and mean  $t$  over  $n$ . You sum  $n$  of them and divide by  $n$ . You know that it just converges to  $t$  over  $n$ , just this one number.

It doesn't-- it's a distribution, but most of the time it's just  $t$  over  $n$ . OK? If you take that equal to  $t$ , because these are random variables accumulating these squared terms. That's what's happened. Just a nice application of strong law of large numbers, or just law of large numbers. To be precise, you'll have to use strong law of large numbers.

OK. So I think that's enough for Brownian motion. And final question? OK. Now, let's move on--

**AUDIENCE:** I have a question.

**PROFESSOR:** Yes.

**AUDIENCE:** So this [INAUDIBLE], is it for all Brownian motion speed?

**PROFESSOR:** Oh, yeah. That's a good question. This is what happens with probability one. So always-- I'll just say always. It's not a very strict sense. But if you take one path according to the Brownian motion, in that path you'll have this. No matter what path you get, it always happens.

**AUDIENCE:** With probability one.

**PROFESSOR:** With probability one. So there's a hiding statement-- with probability. And you'll see why you need this with probability one is because we're using this probability statement here. But for all practical means, like with probability one, it just means always.

Now, I want to motivate Ito's calculus. First of all, this. So now, I was saying that Brownian motion, at least, is not so bad a model for stock prices. But if you remember what I said before, and what people are actually doing, a better way to describe it is instead of the differences being a normal distribution, what we want is



the percentile difference.

So for stock prices we want the percentile difference to be normally distributed. In other words, you want to find the distribution of  $f_t$ , such that the difference of  $f_t$  divided by  $st$ , is a normal distribution. So it's like a Brownian motion. That's the differential equation for it.

So the percentile difference follows Brownian motion. That's what it's saying.

Question, is  $st$  equal to  $e^{-Bt}$ ? Because in classical calculus this is not a very absurd thing to say. If you differentiate each side, what you get is  $df_t$  equals  $e^{-Bt}$  times  $dB_t$ . That's  $st$  times  $dB_t$ . It doesn't look that wrong. Actually, it looks right, but it's wrong. For reasons that you don't know yet, OK?

So this is wrong and you'll see why. First of all, Brownian motion is not differentiable. So what does it even mean to say that? And then that means if you want to solve this equation, or in other words, if you want to model this thing, you need something else. And that's where Ito's calculus comes in.

OK. I'll try not to rush too much. So suppose-- now we're talking about Ito's calculus-- you want to compute. So here is a motivation. You have a function  $f$ . I will call it a very smooth function  $f$ . Just think about the best function you can imagine, like an exponential function.

Then you have a Brownian motion, and then you apply this function. As an input, you put the Brownian motion inside the input. And you want to estimate the outcome. More precisely, you want to estimate infinitesimal differences.

Why will we want to do that? For example,  $f$  can be the price of an option. More precisely, let  $f$  be this thing. OK. You have some  $s_0$ . Up to  $s_0$ , the value of  $f$  is equal to 0. After  $s_0$ , it's just a line with slope 1. Then  $f$  of Brownian motion is just the price exercise-- what is it-- value of the option at the expiration.  $t$  is the expiration time. It's a call option. That's the call option.

So if you're stuck at time  $t$ , go with over  $s_0$ , you make that much. If it's below  $s_0$ , you'll lose that much. More precisely, you have to put it below like that. Let's just do

it like that. And it looks like that.

So that's like a financial derivative. You have an underlying stock and then some function applies to it. And then what you have, the financial asset you have, actually can be described as this function. A function of an underlying stock, that's called financial derivatives. And then in the mathematical world, it's just a function applied to the underlying financial asset.

And then, of course, what you want to do is understand the difference of the value, in terms of the difference of the underlying asset. If  $B_t$  was a very nice function as well. If  $B_t$  was differentiable, then the classical world calculus tells us that  $d$  of  $f$  is equal to  $d$  of  $B_t$  over  $d$  of  $t$  times  $dt$ . Yes.

So if you can differentiate it over the time difference, over a small time scale. All we have to do is understand the differentiation. Unfortunately, we can't do that. We cannot do this. Because we don't know what-- we don't even have this differentiation. OK.

Try one, take one failed, take two. Second try, OK? This is not differentiable, but still I understand the minuscule difference of  $dB_t$ . So what about this?  $df$ -- maybe I didn't write something,  $f'$ -- is equal to just  $dB_t$  of  $f'$ . OK?

What is this? We can't differentiate Brownian motion, but still we understand the minuscule and infinitesimal difference of the Brownian motion. So I just gave up trying to compute the differentiation. But instead, I'm going to just compute how much the Brownian motion changed over this small time scale, this difference, and describe the change of our function in terms of the differentiator of our function  $f$ .  $f$  is a very good function, so it's differentiable.

So we know this. This is computable. This is computable. It's the difference of Brownian motion over a very small time scale. So that at least now is reasonable. We can expect it. It might be true. Here, it didn't make sense at all. Here, it at least make sense, but it's wrong.

And why is it wrong? It's precisely because of this. The reason it's wrong, the

reason it is not valid is because of the fact  $dB$  squared equals  $dt$ . And let's see how this comes into play, this factor. I think that will be the last thing that we'll cover today.

OK. So if you remember where you got this formula from, you probably won't remember. But from calculus, this follows from Taylor's expansion.  $f$  of  $t$  plus  $x$ , I'll say, is equal to  $f$  of  $t$  plus  $f$  prime of  $t$  times  $x$ , plus  $f$  double prime of  $t$ , over  $2$  times  $x$  squared plus-- over  $3$  factorial  $x$  cubed plus--  $df$  is just this difference. Over a very small time increase, we want to understand the difference of the function. That's equal to  $f$  prime  $t$  times  $x$ .

OK. In classical calculus we were able to ignore all these terms. So in the classical world  $f(t + \Delta t) - f(t)$  was about  $f'(t) \Delta t$ . And that's precisely this formula. But if you use Brownian motion here-- so what I'm trying to say is if  $B$  at some time  $t + x$ , minus Brownian motion  $B$  at time  $t$ , then let's just write down the Taylor formula.

We get  $f'$  at  $B_t$ .  $x$  will be this difference,  $B$  at  $t + x$  minus  $B$  of  $t$ . That's like the difference in  $B_t$ . So up to this much we see this formula. And the next term, we get the second derivative of this function over  $2$  and  $x$  squared,  $x$  plus this difference. So what we get is  $dB_t$  squared. OK?

But as you saw, this is no longer ignorable. That is like a  $dt$ , as we deduced. And that comes into play. So the correct-- then by Taylor expansion, the right way to do it is  $df$  is equal to the first derivative term,  $dB_t$ , plus the second derivative term, double prime over  $2 dt$ .

This is called Ito's lemma. And now let's say if you want to remember one thing from the math part, try to make it this one. This had great impact. If you follow the logic it makes sense.

It's really amazing how somebody came up with for the first time because it all makes sense. It all fits together if you think about it for a long time. But actually, I once saw that Ito's lemma is one of the most cited lemmas, like most cited paper.

The paper that's containing this thing. Because people think it's nontrivial.

Of course, there are facts that are being used more than these classical facts, like trigonometric functions, exponential functions. They are being used a lot more than this, but people think that's trivial so they don't site it in their research and paper. But this, people respect the result. It's a highly nontrivial result.

And it's really amazing how just by adding this term, all this theory of calculus all now fit together. Without this-- maybe it's a too strong statemeng-- but really Brownian motion becomes much more rich because of this fact. Now we can do calculus with it.

So there's two things to remember. Well, if you want to remember one thing, that's Ito's lemma. If you want to remember two things, it's just quadratic variation,  $dB_t$  squared is equal to  $dt$ .

And I remember that's exactly because  $B_t$  is like a normal variable with  $0_t$ , and time scale  $B_t$  is like a normal [INAUDIBLE] interfering with  $t$ .  $dB_t$  squared is like the variance of it. So it's  $t$ , and if you differentiate it, you get  $dt$ . That was exactly how we computed it.

So, yeah, I'll just quickly go over it again next time just to try to make it stick in to your head. But please, think about it. This is really cool stuff. Of course, because of that computation calculus using Brownian motion, it becomes a lot more complicated. Anyway, so I'll see you on Thursday. Any last minute questions? Great.