

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.

PROFESSOR: Let's begin. Today we're going to continue the discussion on Ito calculus. I briefly introduced you to Ito's lemma last time, but let's begin by reviewing it and stating it in a slightly more general form.

Last time what we did was we did the quadratic variation of Brownian motion, Brownian process. We defined the Brownian process, Brownian motion, and then showed that it has quadratic variation, which can be written in this form-- $d B^2$ is equal to dt . And then we used that to show the simple form of Ito's lemma, which says that if f is a function on the Brownian motion, then $d f$ is equal to f' of $d B_t$ plus f'' of dt .

This additional term was a characteristic of Ito calculus. In classical calculus we only have this term, but we have this additional term. And if you remember, this happened exactly because of this quadratic variation. Let's review it, and let's do it in a slightly more general form.

As you know, we have a function f depending on two variables, t and x . Now we're interested in-- we want to evaluate our information on the function $f(t, B_t)$. The second coordinate, we're planning to put in the Brownian motion there. Then again, let's do the same analysis. Can we describe $d f$ in terms of these differentiations?

To do that, deflect this, let me start from Taylor expansion. f at a point $t + \Delta t$, $x + \Delta x$ by Taylor expansion for two variables is $f(t, x) + \frac{\partial f}{\partial t} \Delta t + \frac{\partial f}{\partial x} \Delta x$. That's the first order of terms.

Then we have the second order of terms. Then the third order of terms, and so on. That's just Taylor expansion.

If you look at it, we have a function f . We want to look at the difference of f when we change to the first variable a little bit and the second variable a little bit. We start from f of t of x . In the first order of terms, you take the partial derivative, so take $\frac{\partial f}{\partial t}$, and then multiply by the t difference. Second term, you take the partial derivative with respect to the second variable-- $\frac{\partial f}{\partial x}$ -- and then multiply by Δx .

That much is enough for classical calculus. But then, as we have seen before, we ought to look at the second order of term. So let's first write down what it is. That's exactly what happened in Taylor expansion, if you remember. If you don't remember, just believe me. This $\frac{1}{2}$ times takes the second derivative partial.

Let's write it in terms of-- yes?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Oh, yeah, you're right. Thank you. Is it good enough?

Let's write it as Δt all these deltas. I'll just write like that. I'll just not write down t of x . And what we have is f plus $\frac{\partial f}{\partial t} \Delta t$ plus $\frac{\partial f}{\partial x} \Delta x$ plus the second order of terms.

The only important terms-- first of all, these terms are important. But then, if you want to use x equals B of t -- so if you're now interested in $f(t, B(t))$. Or more generally, if you're interested in $f(t, B(t))$ plus Δt , $f(B(t))$ plus d of $B(t)$, then these terms are important.

If you subtract f of t of $B(t)$, what you get is these two terms. $\frac{\partial f}{\partial t} \Delta t$ plus $\frac{\partial f}{\partial x} \Delta x$ -- I'm just writing this as a second variable differentiation-- $\Delta B(t)$. And then the second order of terms.

Instead of writing it all down, Δt^2 is insignificant, and $\Delta t \Delta B(t)$ also is insignificant. But the only thing that matters will be this one. This is $d B(t)^2$, which you saw is equal to Δt .

From the second order of term, you'll have this term surviving. $\frac{1}{2}$ over 2 partial f over partial x second derivative of dt . That's it. If you rearrange it, what we get is partial f over partial t plus $\frac{1}{2}$ this plus-- and that's the additional term.

If you ask me why these terms are not important and this term is important, I can't really say it rigorously. But if you think about $d B_t$ square equals dt , then d times B_t is kind of like square root of dt . It's not a good notation, but if you do that-- these two terms are significantly smaller than dt because you're taking a power of it. dt square becomes a lot smaller than dt , dt [INAUDIBLE] is a lot smaller than dt . But this one survives because it's equal to dt here. That's just the high level description.

That's a slightly more sophisticated form of Ito's lemma. Let me write it down here. And let's just fix it now. If f is t of B_t -- that's d of f is equal to-- Any questions? Just remember, from the classical calculus term, we're only adding this one term there. Yes?

AUDIENCE: Why do we have x there?

PROFESSOR: Because the second variable is supposed to be x . I don't want to write down partial derivative with respect to a Brownian motion here because it doesn't look good. It just means, take the partial derivative with respect to the second term. So just view this as a function f of t of x , evaluate it, and then plug in x equals B_t in the end, because I don't want to write down partial B_t here. Other questions?

Consider a stochastic process X of t such that d of x is equal to μ times d of t plus σ times d of B_t . This is almost like a Brownian motion, but you have this additional term. This is called a drift term. Basically, this happens if X_t is equal to μt plus σ of B_t . μ and σ are constants.

From now on, what we're going to study is stochastic process of this type, whose difference can be written in terms of drift term and the Brownian motion term. We want to do a slightly more general form of Ito's lemma, where we want f of t of X_t here. That will be the main object of study.

I'll finally state the strongest Ito's lemma that we're going to use. f is some smooth

function and X_t is a stochastic process like that. X_t satisfies where B_t is the Brownian motion. Then df of t , X_t can be expressed as-- it's just getting more and more complicated.

But it's based on this one simple principle, really. It all happened because of quadratic variation. Now I'll show you why this form deviates from this form when we replace B to x .

Remember here all other terms didn't matter, that the only term that mattered was partial square of f of dx square. To prove this, note that df is partial f over partial t dt plus partial f over partial x d of X_t plus $1/2$ of d of x squared. Just exactly the same, but I've place the $d B_t$ previously, what we had $d B_t$ I'm replacing to dX_t .

Now what changes is dX_t can be written like that. If you just plug it in, to get here is partial f over partial x μdt plus σ of $d B_t$. Then what you get here is $1/2$ of partials and then μ plus σ $d B_t$ square.

Out of those three terms here we get μ square dt square plus 2 times μ σ d μdB plus σ square $d B_t$ square. Only this was survives, just as before. These ones disappear.

And then you just collect the terms. So dt -- there's one dt here. There's μ times that here, and that one will become a dt . It's $1/2$ of σ square f square of dt . And there's only one $d B_t$ term here. σ -- I made a mistake, σ .

This will be a form that you'll use the most, because you want to evaluate some stochastic process-- some function that depends on time and that stochastic process. You want to understand the difference, df . The X would have been written in terms of a Brownian motion and a drift term, and then that's the Ito lemma for you.

But if you want to just-- if you just see this for the first time, it just looks too complicated. You don't understand where all the terms are coming from. But in reality, what it's really doing is just take this Taylor expansion. Remember these two classical terms, and remember that there's one more term here. You can derive it if

you want to.

Really try to know where it all comes from. It all started from this one fact, quadratic variation, because that made some of the second derivative survive, and because of those, you get these kind of complicated terms. Questions? Let's do some examples. That's too much. Sorry, I'm going to use it a lot, so let me record it.

Example number one. Let f of x be equal to x square, and then you want to compute d of f at B_t . I'll give you three minutes just to try a practice. Did you manage to do this? It's a very simple example.

Assume it's just the function of two variables, but it doesn't depend on t . You don't have to do that, but let me just do that. $\partial f / \partial t$ is 0. $\partial f / \partial x$ is equal to $2x$, and the second derivative equal to 2 at $t x$. Now we just plug in t comma B_t , and what you have is μ equals 0, σ equals 1, if you want to write it in this formula.

What you're going to have is 2 times B_t of $d B_t$ plus 1 over 2 times $2 dt$. You should write it down. You can either use these parameters and just plug in each of them to figure it out. Or a different way to do it is really write down, remember the proof.

This is $\partial f / \partial t dt$ plus $\partial f / \partial x dx$ plus $1/2$ -- remember this one. And x is $d B_t$ here. That one is 0, that one was $2x$, so $2 B_t d B_t$. Use it one more time, so you get dt . Make sense?

Let's do a few more examples. And you want to compute d of f at t comma B of t . Let's do it this time. Again, $\partial f / \partial t dt$ plus $\partial f / \partial x d B_t$. That's the first order of terms. The second order of term is $1/2$ partial square f over partial x square of $d B_t$ square, which is equal to dt .

Let's do it. $\partial f / \partial t$, you get μ times f . This one is just equal to μ times f . Maybe I'm going too quick. μ times e to the μt plus $dx dt$.

$\partial f / \partial x$ is σ times e to the μt plus dx , and then $d B_t$ plus-- if you take the second derivative, you do that again, what you get is $1/2$, and then σ

square times e to the μt plus $dx dt$. Yes?

AUDIENCE: In the original equation that you just wrote, isn't it $1/2$ times sigma squared, and then the second derivative? Up there.

PROFESSOR: Here?

AUDIENCE: Yes.

PROFESSOR: $1/2$?

AUDIENCE: Times sigma squared.

PROFESSOR: Oh, sigma-- OK, that's a good question. But that sigma is different. That's if you plug in X_t here. If you plug in X_t where X_t is equal to $\mu \text{prime } dt$ plus $\sigma \text{prime } d$ of B_t , then that sigma prime will become a sigma prime square here. But here the function is μ and σ , so maybe it's not a good notation. Let me use a and b here instead. The sigma here is different from here.

AUDIENCE: Yeah, that makes a lot more sense.

PROFESSOR: If you replace a and b , but I already wrote down all μ 's and σ 's. That's a good point, actually. But that's when you want to consider a general stochastic process here other than Brownian motion. But here it's just a Brownian motion, so it's the most simple form. And that's what you get.

μ plus $1/2$ sigma square-- and these are just all f itself. That's the good thing about exponential. f times dt plus sigma times d of B_t . Make sense?

And there's a reason I was covering this example. It's because-- let's come back to this question. You want to model stock price using Brownian motion, Brownian process, S of t .

But you don't want S_t to be a Brownian motion. What you want is a percentile difference to be a Brownian motion, so you want this percentile difference to behave like a Brownian motion with some variance.

The question was, is S_t equal to $e^{\sigma B_t}$ in this case? And I already told you last time that no, it's not true. We can now see why it's not true.

Take this function, S_t equals $e^{\sigma B_t}$, that's exactly where μ is equal to 0 here. What we got here was dS_t , in this case, is equal to $\mu S_t dt + \sigma S_t dB_t$, so we get $1/2 \sigma^2 S_t^2 dt + \sigma S_t dB_t$. We originally were targeting $\sigma S_t dB_t$, but we got this additional term which we didn't want in the first. In other words, we have this drift.

I wasn't really clear in the beginning, but our goal was to model stock price where the expected value is 0 at all times. Our guess was to take $e^{\sigma B_t}$, but it turns out that in this case we have a drift, if you just take natural $e^{\sigma B_t}$. To remove that drift, what you can do is subtract that term somehow. If you can get rid of that term then you can see if you add this μ to be minus $1/2 \sigma^2$, you can remove that term. That's why it doesn't work. So instead use $S_t = e^{\sigma B_t - 1/2 \sigma^2 t}$.

That's the geometric Brownian motion without drift. And the reason it has no drift is because of that. If you actually do the computation, the dt term disappears.

Question?

So far we have been discussing differentiation. Now let's talk about integration.

Yes?

AUDIENCE: Could you we do get this solution as [INAUDIBLE]. Could you also describe what it means? What does it mean, this solution of B_t ? Does that mean if we have a sample B_t , then we could get a sample B_t [INAUDIBLE]?

PROFESSOR: Oh, what this means, yes. Whenever you have the B_t value, just at each time take the exponential value. Because why we want to express this in terms of a Brownian motion is, for Brownian motion we have a pretty good understanding. It's a really good process you understand fairly well, and you have good control on it.

But the problem is you want to have a process whose percentile difference behaves like a Brownian motion. And this gives you a way of describing it in terms of

Brownian motion, as an exponential function of it. Does that answer your question?

AUDIENCE: Right, distribution means that if we have a sample B_t , that would be the corresponding sample B_t [INAUDIBLE]?

PROFESSOR: That's a good question, actually. Think of it as a point related to valuation. That is not always correct, but for most of the things that we will cover, it's safe to think about it that way. But if you think about it path wise all the time, eventually it fails. But that's a very advanced topic.

So what this question is, basically B_t is a probability space, it's a probability distribution over passes. For this equation, if you just look at it, it looks right, but it doesn't really make sense, because B_t -- if it's a probability distribution, what is e to the B_t ?

Basically, what it's saying is B_t is a probability distribution over passes. If you take ω according to a pass according to the Brownian motion example probability distribution, and for this pass it's well defined, this function. So the probability density function of this pass is equal to the problem to density function of e to the whatever that is in this distribution. Maybe it confused you more. Just consider this as some pass, some well defined function, and you have a well defined function.

Integral definition. I will first give you a very, very stupid definition of integration. We say that we define F as the integration if d of F is equal to $f d B_t$ plus-- We define it as an inverse of differentiation.

Because differentiation is now all defined-- we just defined integration as the inverse of it, just as in classical calculus. So far, it doesn't have that good meaning, other than being an inverse of it, but at least it's well defined. The question is, does it exist? Given f and g , does it exist, does integration always exist, and so on. There's lots of questions to ask, but at least this is some definition. And the natural question is, does there exist a Riemannian sum type description?

That means-- if you remember how we defined integral in calculus, you have a

function f , integration of f from a to b . According to the Riemannian sum description was, you just chop the interval into very fine pieces-- $a_0, a_1, a_2, a_3, \dots$ and then sum the area of these boxes, and take the limit. And this is the limit of Riemannian sums. Slightly more, if you want, is it's the limit as n goes to infinity of the function $\frac{1}{n}$ times the sum of $[f(a_i) - f(a_{i-1})]$ over i from 1 to n . Does this ring a bell? Question?

AUDIENCE: [INAUDIBLE]

PROFESSOR: No, you're right. Good point, no we don't. Thanks. Does integral defined in this way have this Riemannian sum type description, is the question. So keep that in mind. I will come back to this point later.

In fact, it turns out to be a very deep question and very important question, this question, because if you remember like I hope you remember, in the Riemannian sum, it didn't matter which point you took in this interval. That was the whole point. You have the function. In the interval a_i to a_{i+1} , you take any point in the middle and make a rectangle according to that point. And then, no matter which point you take, when you go to the limit, you had exactly the same sum all the time.

That's how you define the limit. But what's really interesting here is that it's no longer true. If you take the left point all the time, and you take the right point all the time, the two limits are different. And again, that's used in quadratic variation, because that much of variance can accumulate over time.

That's the reason we didn't start with Riemannian sum type definition of integral. But I'll just make one remark. Ito integral is the limit of Riemannian sums when always take the leftmost point of each interval. So you chop down this curve at the time interval into pieces, and for each rectangle, pick the leftmost point, and use it as a rectangle.

And you take the limit. That will be your Ito integral defined. It will be exactly equal to this thing, the inverse of our Ito differentiation. I won't be able to go into detail.

What's more interesting is instead, what happens if you take the rightmost point all

the time, you get an equivalent theory of calculus. It's just like Ito's calculus. It looks really, really similar and it's coherent itself, so there is no logical flaw in it. It all makes sense, but the only difference is instead of a plus in the second order of term, you get minuses.

Let me just make this remark, because it's just a theoretical part, this thing, but I think it's really cool. Remark-- there's this and equivalent version. Maybe equivalent is not the right word, but a very similar version of Ito calculus such that basically, what it says is $d B_t^2$ is equal to minus dt .

Then that changed a lot of things. But this part, it's not that important. Just cool stuff.

Let's think about this a little bit more, this fact. Taking the leftmost point all the time means if you want to make a decision for your time interval-- so at time t of i and time t of i plus 1, let's say it's the stock price. You want to say that you had so many stocks in this time interval. Let's say you had so many stocks in this time interval according to the values between this and this.

In real world, your only choice you have is you have to make the decision at time t of i . Your choice cannot depend on the future time. You can't suddenly say, OK, in this interval the stock price increased a lot, so I'll assume that I had a lot of stocks in this interval. In this interval, I knew it was going to drop, so I'll just take the rightmost interval. I'll assume that I only had this many stock.

You can't do that. Your decision has to be based on the leftmost point, because the time. You can't see the future.

And the reason Ito's calculus works well in our setting is because of this fact, because it has inside it the fact that you cannot see the future. Every decision is made based on the leftmost time. If you want to make a decision for your time interval, you have to do it in the beginning.

That intuition is hidden inside of the theory, and that's why it works so well. Let me reiterate this part a little bit more. It's the definition of these things where you're only

allowed to-- at time t , you're only allowed to use the information up to time t .

Definition Δt is an adapted process-- sorry-- adapted to another stochastic process X_t if for all values of time variables Δt depends only on X_0 up to X_t . There's a lot of vague statements inside here, but what I'm trying to say is just assume x is the Brownian motion underlying stock price.

Your stock is changing. You want to call it with a strategy, and you want to say that mathematically this strategy makes sense. And what it's saying is if your strategy makes your decision at time t is only based on the past values of your stock price, then that's an adapted process.

This defines the processes that are reasonable, that cannot see future. And these are all-- in terms of strategy, if Δt is a portfolio strategy, these are the only meaningful strategies that you can use. And because of what I said before, because we're always taking the leftmost point, adaptive processes just also fit very well with Ito's calculus. They'll come into play altogether.

Just a few examples. First, a very stupid example. X_t is adapted to X_t . Of course, because at time, X_t really depends on only X_t , nothing else.

Two, X_{t+1} is not adapted to X_t . This is maybe a little bit vague, so we'll call it Y_t equals X_{t+1} . Y_t is the value at $t+1$, and it's not based on the values up to time t . Just a very artificial example.

Another example, Δt equals \min is adapted. And I'll let you think about it. The fourth is quite interesting. Suppose T is fixed, some large integer, or some large real number. Then you let Δt to be the maximum where X of s . It's not adapted.

What is this? This means at time T , I'm going to take at it this value, the maximum of all value inside this part, the future. This refers to the future. It's not an adapted process. Any questions?

Now we're ready to talk about the properties of Ito's integral. Let's quickly review

what we have. First, I defined Ito's lemma-- that means differentiation in Ito calculus. Then I defined integration using differentiation-- integration was an inverse operation of the differentiation.

But this integration also had an alternative description in terms of Riemannian sums, where you're taking just the leftmost point as the reference point for each interval. And then, as you see, this naturally had this concept of using the leftmost point. And to abstract that concept, we've come up with this adapted process, very natural process, which is like the real life procedures, real life strategies we can think of.

Now let's see what happens when you take the integral of adapted processes. Ito integral has really cool properties.

The first thing is about normal distribution. B_t has normal distribution of 0 up to t . So your Brownian motion at time t has normal distribution with 0, t . That means if your stochastic process is some constant time B of t , of course, then you have 0 and c square t .

It's still a normal variable. That means if you integrate, that's the integration of some σ . That's the integration of σ of $d B_t$.

If σ is a fixed constant, when you take the Ito integral of σ times $d B_t$, this constant, at each time you get a normal distribution. And this is like saying the sum of normal distribution is also normal distribution. It has this hidden fact, because integral is like sum in the limit.

And this can be generalized. If σ is on a process depending only on the time variable-- so it does not depend on the Brownian motion-- then the process X of t equals the integration of σ $d B_t$ has normal distribution at all time. Just like this, we don't know the exact variance yet. The variance will depend on the sigmas, but still, it's like a sum of normal variables, so we'll have normal distribution.

In fact, it just gets better and better. The second fact is called Ito isometry. That was cool. Can we compute the variance? Yes?

AUDIENCE: Can you put that board up?

PROFESSOR: Sure.

AUDIENCE: Does it go up?

PROFESSOR: This one doesn't go up. That's bad. I wish it did go up.

This has a name called Ito isometry. Can be used to compute the variance. B_t has a Brownian motion, Δt is adapted to a Brownian motion. Then the expectation of your Ito integral-- that's the Ito integral of your adapted process.

That's the variance-- we take the square of it-- is equal to something cool. The square just comes in. Quite nice, isn't it?

I won't prove it, but let me tell you why. We already saw this phenomenon before. This is basically quadratic variation. And the proof also uses it. If you take Δs equals to 1-- sorry, I was using Korean-- 1 at all time, then what we have is here you get a Brownian motion, B_t .

So on the left you get an expectation of B_t square, and on the right, what you get is t . Because when Δs is equal to 1 at all time, when you have to get from 0 to t you get t , and you have t on the right hand side. That's what it's saying. And that was the content of quadratic variation, if you remember. We're summing the squares-- maybe not exactly this, but you're summing the squares over small intervals.

So that's a really good fact that you can use to compute the variance. You have an Ito integral, you know the square, can be computed this simple way. That's really cool.

And one more property. This one will be really important. You'll see it a lot in future lectures. It's that when is Ito integral a martingale?

What's a martingale? Martingale meant if you have a stochastic process, at any time t , whatever happens after that, the expected value at time t is equal to 0. It

doesn't have any natural tendency to go up or go down. No matter which point you stop your process and you see your future, it doesn't have a natural tendency to go up or go down. In formal language, it can be defined as where F_t is the events X_0 gets t .

So if you take the conditional expectation based on whatever happened up to time t , that expectation will just be whatever value you have at that time. Intuitively, that just means you don't have any natural tendency to go up or go down. Question is, when is an Ito integral a martingale?

Adapted to B of t , then it is a martingale. As long as g is not some crazy function, as long as g is reasonable-- only can be reasonable if it's [INAUDIBLE]. If you don't know what it means, you can safely ignore it. Basically, if g doesn't-- it's not a crazy function if it doesn't grow too fast, then in most cases this integral is always a martingale.

If you flip it-- remember, integral was defined as the inverse of differentiation. So if dX_t is equal to some function μ that depends on both t and B_t times dt plus sigma of $d B_t$, what this means is X_t is a martingale if that is 0 at all time, always. And if it's not 0, you have a drift, so it's not a martingale. That gives you some classification.

Now, if you look at a differential equation of this stochastic-- this is called a stochastic differential equation-- if you know stochastic process, if you look at a stochastic differential equation, if it doesn't have a drift term, it's a martingale. If it has a drift term, it's not a martingale. That'll be really useful later, so try to remember it.

The whole point is when you write down a stochastic process in terms of something times dt , something times $d B_t$, really this term contributes towards the tendency, the slope of whatever is going to happen in the future. And this is like the variance term. It adds some variance to your stochastic process. But still, it doesn't add or subtract value over time, it fairly adds variation.

Remember that. That's very important fact. You're going to use it a lot.

For example, you're going to use it for pricing theory. In pricing theory, you come up with this stochastic process or some strategy. You look at its value. Let's say X_t is your value of your portfolio over time. If that portfolio has-- then you match it with your financial-- let me go over it slowly again.

First you have a financial derivative, like option of a stock. Then you have your portfolio strategy. Assume that you have some strategy that, at the expiration time, gives you the exact value of the option. Now you look at the difference between these two stochastic processes. Basically what the thing is, when your variance goes to 0, your drift also has to go to 0.

So when you look at the difference, if you can somehow get rid of this variance term, that means no matter what you do, that will govern the value of your portfolio. If it's positive, that means you can always make money, because there's no variance. Without variance, you make money. That's called arbitrage, and you cannot have that.

But I won't go into further detail because [INAUDIBLE] will cover it next time. But just remember that flavor. So when you write something down in a stochastic differential equation form, that term is a drift term, that term is a variance term. And if you don't have drift, it's a martingale. That is very important.

Any questions? That's kind of the basics of Ito calculus. I will give you some exercises on it, mostly just basic computation exercises, so that you'll get familiar with it. Try to practice it.

And let me cover one more thing called Girsanov theorem. It's related, but these are really basics of the Ito calculus, so if you have any questions on this, please ask me right now before I move on to the next topic. The last thing I want to talk about today.

Here is an underlying question. Suppose you have two Brownian motions. This is without drift. And you have another B tilde Brownian motion with drift. These are two probability distributions overpasses.

According to B_t , you're more likely to have some Brownian motion that has no drift. That's a sample path. According to \tilde{B} , you have some drift. Your Brownian motion will close it.

A typical path will follow this line and will follow that line. The question is this-- can we switch from this distribution to this distribution by a change of measure? Can we switch between the two measures to probability distributions by a change of measure?

Let me go a little bit more what it really means. Assume that you're just looking at a Brownian motion from time 0 up to time t , some fixed time interval. Then according to B_t , let's say this is a sample path ω . You have some probability of ω -- this is a p.d.f. given by this Brownian motion B . And then you have another p.d.f., \tilde{P} of ω , which is a p.d.f. given by \tilde{P} of t .

The question is, does there exist a Z depending on ω such that \tilde{P} of ω is equal to Z times \tilde{P} ? Do you understand the question? Clearly, if you just look at it, they're quite different. The paths that you get according to distributions are quite different. It's not clear why we should expect it at all. You'll see the answer soon. But let me discuss all this in a different context.

Just forget about all the Brownian motion and everything just for a moment. In this concept, changing from one probability distribution to another distribution, it's a very important concept in analysis and probability just in general, theoretically. And there's a name for this Z , for this changing measure. If Z exists, it's called the Radon-Nikodym derivative. Before doing that, let me talk a little bit more.

Suppose P is a probability distribution over ω . It's a probability distribution. So this is some set, and P describes the probability that you have each element in the set. And you have another probability distribution, \tilde{P} .

We define P and \tilde{P} to be equivalent if the probability that A is greater than zero if and only for all. These probability distributions describe the probability of the subsets. Think about a very simple case.

Σ is equal to 1, 2, and 3. P gives $1/3$ probability to 1, $1/3$ probability to 2, $1/3$ probability to 3. \tilde{P} gives $2/3$ probability to 3, $1/6$ probability to 2, $1/6$ probability to 1. We have two probability distributions over some space.

They are equivalent if, whenever you take a subset of your background set-- let's say $A = \{1, 2\}$. When A is equal to $\{1, 2\}$, according to probability distribution P , the probability you fall into this set A is equal to $2/3$. According to \tilde{P} , you have $5/6$.

They're not the same. The probability itself is not the same, but this condition is satisfied when it's 0. And when it's not 0, it's not 0. And you can just check that it's always true, because they're all positive probabilities. On the other hand, if you take instead, say, $1/3$ and 0, now you take your A to be 3. Then you have $1/3$ equal to 0.

This means, according to probability distribution P , there is some probability that you'll get 3. But according to probability distribution \tilde{P} , you don't have any probability of getting 3. So they're not equivalent in this case.

If you think about it, then it's really clear. The theorem says-- this is a very important theorem in analysis, actually. The theorem-- there exists a Z such that P of ω is equal to Z if and only if P and \tilde{P} are equivalent. You can change from one probability measure to another probability measure just in terms of multiplication, if and only if they're equivalent.

And you can see that it's not the case for this when they're not equivalent. You can't make a zero probability to $1/3$ probability by multiplication. So in the finite world this is very just intuitive theorem, but what this is saying is it's true for all probability spaces. And these are called the Radon-Nikodym derivative.

Our question is, are these two Brownian motions equivalent? The passes that this Brownian motion without drift takes and the Brownian motion with drift takes-- are they kind of the same but just skewed in distribution, or are they really fundamentally different? That's the question.

And what Girsanov's theorem says is that they are equivalent. To me, it came as a

little bit non intuitive. I would imagine that it's not equivalent, these two. These passes have a very natural tendency. As it goes to infinity, these passes and these passes will really look a lot different, because when you go really, really far, the passes which have drift will be just really close to your line μ of t , while the passes which don't have drift will be really close to the x axis.

But still, they are equivalent. You can change from one to another. I'll just state that theorem without proof. And this will also be used in pricing theory.

I'm not an expert enough to tell why, but basically what it's saying is, you switch some stochastic process into a stochastic process without drift, thus making it into a martingale. And martingale has a lot of meaning in pricing theory, as you'll see. This also application for it. That's why I'm trying to cover it, although it's quite a technical theorem. Try to remember, at least a statement and the spirit of what it means. It just means these two are equivalent, you can change from one to another by a multiplicative function. Let me just state it in a simple form.

GUEST SPEAKER: If I could just interject a comment.

PROFESSOR: Sure.

GUEST SPEAKER: With these changes of measure, it turns out that all of these theories with continuous time processes should have an interpretation if you've discretized time, and should consider sort of a finer and finer discretization of the process. And with this change of measure, if you consider problems in discrete stochastic processes like random walks, basically how-- say if you're gambling against a casino or against another player, and you look at how your winnings evolve as a random walk, depending on your odds, your odds could be that you will tend to lose. So there's basically a drift in your wealth as this random process evolves. You can transform that process, basically by taking out your expected losses, to a process which has zero change in expectation.

And so you can convert these gambling problems where there's drift to a version where the process, essentially, has no drift and is a martingale. And the martingale

theory in stochastic process courses is very, very powerful. There's martingale convergence theorems. So you know that the limit of the martingale is-- there's a convergence of the process, and that applies here as well.

PROFESSOR: You will see some surprising applications.

GUEST SPEAKER: Yeah.

PROFESSOR: And try to at least digest the statement. When the guest speaker comes and says by Girsanov's theorem, they actually know what it is. There's a spirit.

This is a very simple version. There's a lot of complicated versions, but let me just do it. So P is a probability distribution over passes from 0, T to infinity. What this means is just passes from that stochastic process defined from time 0 to time t . These are passes defined by a Brownian motion with drift μ .

And then \tilde{P} is a probability distribution defined by Brownian motion without drift. Then P and \tilde{P} are equivalent. Not only are they equivalent, we can actually compute their Radon-Nikodym derivative. And the Radon-Nikodym derivative Z which is defined as T of-- which we denote like this has this nice form.

That's a nice closed form. Let me just tell you a few implications of this. Now, assume you have some, let's say, value of your portfolio over time. That's the stochastic process. And you measure it according to this probability distribution.

Let's say it depends on some stock price as the stock price is modeled using a Brownian motion with drift. What this is saying is, now, instead of computing this expectation in your probability space-- so this is defined over the probability space P , our sigma \mathcal{F}_T defined by this probability distribution. You can instead compute it in-- you can compute as expectation in a different probability space.

You transform the problems about Brownian motion with drift into a problem about Brownian motion without a drift. And the reason I have \tilde{Z} instead of Z here is because I flipped. What you really should have is \tilde{Z} here as expectation of Z . If you want to use this Z .

I don't expect you to really be able to do computations and do that just by looking at this theorem once. Just really trying to digest what it means and understand the flavor of it, that you can transform problems in one probability space to another probability space. And you can actually do that when the two distributions are defined by Brownian motions when one has drift and one doesn't have a drift. How we're going to use it is we're going to transform a non martingale process into a martingale process. When you change into martingale it has very good physical meanings to it.

That's it for today. And you only have one more mass lecture remaining and maybe one or two homeworks but if you have two, the second one won't be that long. And you'll have a lot of guest lectures, exciting guest lectures, so try not to miss them.