[SQUEAKING]

[RUSTLING]

[CLICKING]

PETER KEMPTHORNE: And so from last time, we covered the Itô formula for the single case where we have the integral-- or sorry, the function of a Brownian motion, function of one variable. And importantly, with this function of a one variable, we get that the differential of the function is equal to the first derivative of the function evaluated at the Brownian motion plus 1/2 the second derivative of the function.

And so this is Itô's formula. And when we apply Itô's formula in one dimension, we can consider defining the antiderivative of the function F, and that's capital F. And so Itô's formula gives us this expression for the integral of f of Bs dBs. So this is a very useful way to compute Itô integrals.

Now, when we go to functions of two variables, so case 2, we're going to have a function of time. and Bt, say. When we want to evaluate this function of time and space as a function of Brownian motion, then our Itô's formula is expressed this way. We basically have the partials of little f with respect to t and with respect to x and then a second-order term, the second partial of f with respect to x. So this is how Itô's formula generalizes.

Now, what's really convenient to highlight is that, with Itô's formula, we get this top formula for the change in value of f of tx. So we have this theorem, which is for how f of Bt changes from an initial condition. And in looking at this expression, it's a sum of two terms.

The first term is an integral with respect to time, ds. And so that function will be non-random on paths. The second term is the integral of the partial of f with respect to x dBs And that will be random depending on DBs. So if we have this formula for the increment, then we basically have a constant term plus an integral of Brownian motion, which has mean 0 as a random variable. But if the term that's in the integrand over time, if that's equal to 0, then this function f will be a Martingale.

So importantly, in quantitative finance, we often are interested in solving for values of functions where the functions are the price or value of a derivative over time. And if we can express it in the right framework, where the price function follows a Martingale, then it will satisfy this PDE condition. Now, the notion of the value of some derivative contract or asset following a Martingale over time means that it has a constant expected value of time.

So we don't have the natural growth in value according to the risk-free rate. So to set this up, we'll need to normalize the prices discounted to the present value. So that comes into play in different applications.

Well, there are some examples of Itô's formula, applying it. And with the Brownian motion with drift, we went through that last time. And then, finally, in this section, I want to highlight how geometric Brownian motion is represented in terms of the stochastic differential equation as given here. So dPt is equal to Pt mu dt plus Pt sigma dPt dBt. And importantly, we basically have these factors Pt on both sides, which could be divided in both sides. So we might think that we're working with the derivative or differential of the log of the price. But it's a little bit different.

So with this geometric Brownian motion, if we apply Itô's lemma to the function log of dt-- so if we think of log of dt is equal to G of Pt, t, then the partial of G with respect to P is simply 1 over P. The partial of G with respect to t is 0, and the second partial of G with respect to P is minus 1 over P squared.

So Itô's Lemma gives us that the differential of log P is the sum of two terms, one that's times dt, the other that's times sigma dBt. And so log price is going to follow a generalized Wiener process with a drift rate given by mu minus sigma squared over 2 and variance rate equal to sigma squared. Now, what's important about this result is that the drift rate changes from mu to mu minus sigma squared over 2. So basically the drift rate on the log scale has a drag due to the quadratic variation.

And if we consider normalizing the log return-- well, actually not normalizing, but we look at an increment of the log return that's normally distributed. And then let's see, if we consider, I guess, looking at what the conditional distribution is of log Pt, it's given by this Brownian motion, which gives normally distributed log values with a particular mean mu star and variance sigma star. And so the expectation of our price at some time point capital T given the value at time little t, it'll be a lognormal random variable with expectation equal to the initial price Pt, P sub little t, times the exponential of mu times T minus t. There should be an e here, which is in that factor.

So despite there being a drag on the drift, the expectation of the exponential of the log doesn't have that drag due to drift. It actually has the mean log return in that factor. And with lognormal random variables, we have formulas for the mean and variance, which we covered, actually, in the first couple of lectures.

Now, interestingly, if we think about just what kinds of results happen with values of mu and sigma that we might expect to see in financial markets, here's an example where we consider a drift rate mu of 0.3 per year, a time period for the lognormal return of one half year from time little t to capital T, and an annualized volatility of 0.4, or 40%-- so this is perhaps reasonable for some moderately risky stocks that have a modest return expectation of 30% per year and a volatility rate of 40% per year. And when we compute the expectation of the price, It equals the initial price Pt times 1.1618.

Now, that 1.1618 says that we expect the price to increase by about 16%. And if we think of a half year and a drift rate of 0.3, we might have thought it would be 15% If we didn't follow through the details. So it's a bit higher return. And similarly, for the volatility or standard deviation of this price, it turns out that the volatility or variance of the price at time capital T is 0.335 squared times the square of Pt. So the variance is also a bit higher than you might expect if you just naively interpreted the mu and sigma as appropriate annualized values.

Now, with this result, we can actually compute the continuously compounded percent return of a geometric Brownian motion. So we can look at the change in value from time little t to capital T-- the log of that ratio is the difference in logs-- and divide that by the length of the period, capital T minus t. And we get the annualized rate of return for the geometric Brownian motion. And what that rate of return as a random variable is, it is normal with mean mu minus a half sigma squared and variance sigma squared over T minus t.

So with this, what's significant or important is that the rate of return over long periods of time converges to a constant average rate. And the variance about that average rate is going to 0 as the time into the future continues. So this is an important property of geometric Brownian motion, that the growth of the process basically grows at a constant rate with some variability. But that variability goes to 0 as we consider very, very long periods.

So here, let's say I have a simulation of Brownian motion where we consider the same mu and sigma as the previous slides, and we consider a time period of capital T equal to four years. And we define increments on that interval. And look at, basically, Brownian motion with the appropriate mu and sigma, then geometric Brownian motion, which is the exponential of that Brownian motion, and then the log returns. So let's take a look at what we have.

So here's some simulated paths of Brownian motion with drift with those mu and sigma parameters. And importantly, we have the mean return as a function of time, or the mean value as a function of time of the Brownian motion being given by the blue line. And then plus or minus 2 standard deviations is given by the red curves.

And if we simulate not five but, say, 50 paths, this is what this Brownian motion looks like. And if we-- I guess, before doing this, I'd like everyone to just think about how is this set of paths going to transform when we take the exponential of them, so basically the exponential of Brownian motion with constant drift and variance?

Well, here's the geometric Brownian motion. And what I've done is I've drawn the same blue line and the same red lines for the Brownian motion and how that translates to the exponential scale. And what's rather striking here is that there are paths out of 50 that really have extraordinary performance, and so basically growing in value from by a factor of 10, or an order of 10, basically to an order of 20 on the exponential scale.

And so what I think is an interesting takeaway from this is that, if we have a simple Brownian motion process defining the log price processes, then with reasonably modest drift and volatility parameters, there can be some realizations that have phenomenal returns. And so when you look at the price returns of stocks, often our attention gravitates towards those that have extraordinary returns, like NVIDIA or Tesla, say, over some periods in the past, maybe not so much recently. But these extraordinary returns over a number of years is actually consistent with the possibility of geometric Brownian motion models underlying those.

Now, if we look at the average log return, we get this graph for these 50 paths just doing the time series of volatilities of the-- or sorry, of returns, average returns over across each of the paths. And if we look at zooming in on that path, this is the space-- or it's an illustration of the realizations of the average returns over time. So over short periods, we can expect very different return magnitudes, perhaps. But as the time period for measuring returns increases, it stabilizes to vary about this constant level. So let's go back here.

So with stochastic calculus, there are many follow-on topics. We're going to focus today on stochastic differential equations. But it's important to know that stochastic calculus and differential equations apply to other problems, such as population growth and filtering problems, optimal stopping problems. And so it's a very rich set of applications of the theory.

So let's go to stochastic differential equations. These slides are a bit dense. But I've color-coded some of the pieces so we can go through it more slowly. So let's go through the derivation of the Black-Scholes differential equation. Now, you've already seen this a few times in the class from a guest lecturer and from Vasily's lecture, but I think it bears repeating here.

If we have an asset price that follows geometric Brownian motion, so dP_t is equal to mu-- P_t mu dt plus P_t sigma, dB_t, then let's consider the price of a derivative that's contingent on P_t. And as a motivating example-- say, a call option-- but we could consider arbitrary options or derivatives. And if we apply Itô's lemma to the price function G_t, then Itô's lemma gives us this sum of two terms, dt times the partial of G with respect to P mu P_t, partial of G with respect to dt, the second partial of G with respect to P times 1/2 times sigma squared P_t squared, and then, finally, this partial of G with respect to P times sigma P_t dB_t.

Now, if we think about discretizing the price dynamics of the stock or the asset, as well as the derivative, then we can say, OK, the change or increment in P_t is a mean growth rate times delta t plus sigma P_t delta B_t. So there's a random term and a deterministic term given time little t. And if we do the same for the increment of the derivative, we then basically substitute into dG, delta t's and delta B's. So we get the same constant factor times delta t plus dG dP times sigma P_t delta B_t.

And the brilliant argument for solving for the option price or derivative price is to consider constructing a portfolio V_t that eliminates dependence on the Brownian motion. So if we eliminate dependence on the Brownian motion, then we basically have a deterministic price function for the portfolio. So one way of doing this is to say, let us short one derivative, so minus G_t. If we sell the derivative at price G_t, we receive G_t dollars. And then we're going to add to that a number of shares of the asset given by dG dP.

So if we basically take minus 1 times this plus the dG dP times the asset, then the increment of value of this portfolio is simply delta t times this factor. And we've canceled out the terms in green when multiplied by minus 1 for the derivative or dG dP for the asset. So we end up getting a portfolio whose change in value depends just on this bracketed term here, with no Brownian motion increment involved.

And so having eliminated the risk in the portfolio, arbitrage arguments say that V_t must be risk-free. And as a risk-free asset equivalent, it should earn the riskless interest rate. So our change in value of this portfolio is simply the initial value times the rate of return of the riskless asset over the period delta t.

And so what we can do then is equate the two expressions for what the increment in V_t should be. If it's riskless, here's the right-hand side. And according to our Itô's formulas, the portfolio has an increment value given by this factor of delta t. So if we eliminate delta t from both sides, we get this formula, which is called the Black-Scholes differential equation.

Now, what's really significant is that at least when one studies Black-Scholes theory in finance classes, one thinks of call options and deriving their price or put options through put call parity. What's really important here is that this Black-Scholes differential equation applies for those vanilla options, but also for any contingent claim derivative where the derivative-- where the financial derivative G_t basically depends upon the value of the underlying asset.

So we have G_t, which is a contingent claim. And we assume that G_t is equal to some function f of P_t and t. So we assume that the derivative depends only on the current level of the asset and the current time. Yes?

AUDIENCE: So they won a Nobel Prize for it, right? What part of the paper was the Nobel Prize-winning part? Was is it the construction of the [INAUDIBLE] or was it the [INAUDIBLE]?

PETER KEMPTHORNE: It really was the extensions of the initial papers to handle modeling derivatives under no arbitrage assumptions. And so the basic theory extends amazingly to a broad range of applications. And so yeah. I mean, the initial papers were rather impressive in terms of solving these pricing problems using analysis, basically. And so the initial papers were really quite mathematical.

And since then, arguments and developments have been made that make it more straightforward. And what we're showing here is essentially the benefits from those later developments. Yes?

AUDIENCE: Is every [INAUDIBLE] changing what the portfolio is?

PETER KEMPTHORNE: I'm sorry, can you say--

AUDIENCE: Like, the amount of the asset changes over time. So then I'm not seeing the value of the portfolios, like Pt. Since the amount of Pt has changed, then they bought or sold some of it, [INAUDIBLE]

PETER KEMPTHORNE: Well, OK, I guess this equation should be interpreted in terms of just infinitesimal increments delta t. And as you point out, as time evolves, the value of the contract evolves. The partial derivatives of the asset with respect to time, those are not constant, as well. So as the process increments, those will also change.

So this is a simplification, I guess, of the argument. But if we're looking at what properties must hold in the infinitesimal time range, then this is exact. But OK.

AUDIENCE: Yes. Why can we always say that Vt equals minus Vt plus that term after t beta [INAUDIBLE]?

PETER KEMPTHORNE: Oh, why can we say that this equals this term?

AUDIENCE: The second to the [INAUDIBLE]

PETER KEMPTHORNE: OK, just repeat the question.

AUDIENCE: So the expression for Vt at the top, how do-- [INAUDIBLE] for all t's?

PETER KEMPTHORNE: OK. You're saying that the Vt equaling minus Gt plus-- OK, that's going to be the initial value of the portfolio at time little t. So at time little t, we are combining a short derivative claim.

AUDIENCE: So we just ignore what happened before.

PETER KEMPTHORNE: Exactly, we ignore what happens before. But this highlights in a good way-- it's a good question-- that, when t evolves-- or the time involves past time little t, then the partial of G with respect to P at that future time point will not necessarily equal this. And in order to have a portfolio that eliminates the Brownian motion term, one would need to adjust the number of shares in the asset.

So if we go back to this argument here, where we sell short one derivative and buy this many shares at time little t, this cancels out the Brownian motion. But then for increments of time t or time after t, we still are short the derivative. But we may need to adjust the number of shares of the asset to buy to hedge that derivative. And that's where delta hedging comes into play.

AUDIENCE: So that expression doesn't include-- if we're not assuming that we've done this through time 0 [INAUDIBLE]

PETER KEMPTHORNE: No, not at all, not at all. Yeah. And it's helpful just to see, OK, if, at any arbitrary time t, you want to consider buying the or selling the derivative, valuing the derivative, how should that be valued? Great. OK. So what's important with this Black-Scholes partial differential equation is that there are different boundary conditions that vary by the derivative that we are working with.

And so with call options, we have basically the final value at time capital T. The maturity of the option is given by the hockey stick function of Pt minus k if that's greater than 0 or 0. And similarly, for the put option, it has this terminal value. Now, what's, as I said, very powerful about this theory is that the partial differential equation, the Black-Scholes partial differential equation is satisfied by any derivative, so long as there is no arbitrage in the market. And so with different derivatives contracts, there will be different boundary conditions on those.

So what we now want to do is figure out how to solve these partial differential equations. And what we're going to do is go through the diffusion equation. And it turns out that the Black-Scholes partial differential equation can be transformed into this diffusion equation by changing the scale of time, reversing time, and doing some other normalizations. So this diffusion equation is very fundamental to solving the Black-Scholes partial differential equations. And it's the simplest form of this second-order partial differential equation that we can study and understand its solutions.

So the heat equation is an equation that is satisfied by a function u if the partial of u with respect to t is equal to lambda times the second partial of u with respect to x squared. So this is the heat equation. And we're going to give an example here which is very important and technical. But in terms of what kinds of functions satisfy this diffusion equation, we actually went through looking at the density of a Gaussian distribution, which is equal to 1 over root 2 pi root t e to the minus 1/2 x squared over t.

So if we think of a Gaussian density, which basically is a bell-shaped curve around 0 with some standard deviation here given by 1, this would be the density perhaps for t equals 1, the density function. And this is the density of a Brownian motion starting at time 0, ending at time t. And if we consider increasing time, then the change in P-- So P sub t is equal to lambda times P sub xx.

And so the change over time is proportional to the curvature of the density. And so if you look back in the course notes, course materials, we looked at how the density of Brownian motion over time, it drops where there's strong negative curvature, and it rises where there's positive curvature. And so we basically get flatter curves as time increases.

So the normal density function follows this heat equation. And what we're going to see is that functions that follow the heat equation can actually be represented as linear combinations of these Gaussian densities. So the special case of the Gaussian density is, in a sense, very, very general in terms of defining solutions to the heat equation.

So let's consider, I guess, how the heat equation is often set up. It says OK, let's consider the temperature in a long, thin bar of material that's perfectly insulated, so temperature only varies with distance along the bar at time t. So mu of xt is the value of the temperature over time. And if we consider an initial value for the heat in the bar given by u not of x, then what we want to do is solve for how that heat function changes.

And so some really neat properties to use are similar properties to what we have when we're trying to solve for ordinary differential equations. So first observation is, suppose we have two solutions, u1 and u2. Then the sum of those is going to also be a solution. So solutions are linear.

And if we have a collection of solutions indexed by little s, we can consider taking a weighted sum of those per an integral. And this will also be a solution. So the linearity of the solution spaces extends to this integral of solutions indexed by s also being a solution. Let's see.

So the simplest initial value problem is to consider this u function to be a Dirac delta function at x equals 0, so u of x0 equaling the Dirac delta function at t equals 0. And if we consider this function, which is essentially the PDF of a normal with mean 0 and variance 2 lambda t, then this density function turns out to solve the heat-- or satisfy the heat equation. But it also satisfies the initial condition, where, when we let t go to 0, then it equals the Dirac delta function.

So what's the Dirac delta function? It's equal to 0 for-- it equals 0 for x0 equal to 0. And it equals-- and the integral of u of x0 dx is equal to 1. So it's a function that's it's positive but has weight under the integral equal to 1. So this is a rather abstract kind of function to work with.

And actually, I don't know if-- is anyone here familiar with Paul Dirac? There's a great-- he was a physicist at Cambridge University. And if you're familiar with, I guess, Schrödinger and other physicists, he's like number three or four in the list of top physicists at the time.

And he came up with this Dirac delta function as something worth working with. So it's, in a sense, an imaginary function. It's basically the limit of this normal density function as lambda or as t goes to 0. So it becomes totally concentrated at 0. So with this function, u delta, we have this particular function satisfying the heat equation with that initial condition.

Now, if we want to generalize from this initial function or the Dirac delta function at 0, suppose we consider wanting to solve where the initial conditions are given by a function u0 of x. So in this discussion, we can think of u0 of x, the initial value, being some arbitrary known function. So it's like the boundary conditions of this function at time 0.

And if we integrate our u0 function with the Dirac delta function evaluated at x minus s, then we simply get back this u0 function at x. So the Dirac delta function allows us to pull out in an integral the u0 function itself. So there's an identity relationship between u0 and this transform of u0 with the Dirac delta function.

Now, if we consider different solutions. We can index solutions to this heat equation by delta-- or by s, rather. And then we can integrate this u delta function with respect to the factor u0 of s, and we get this formula for u of xt. So for any initial condition function, u0 of s, this integral will, in fact, give us the solution function u whose initial value u of x0 equals u0.

And so in order for this to work, we need to have the integrals existing. And if we have-- if the integrals exist, and we can take partial derivatives inside the integral sign, then we end up getting the partial of u with respect to t is the integral of the partial of u delta with respect to dt. That's the only part that depends on t. And similarly, the partial with respect to x, second partial, we can take that under the integral sign and get this expression here. So if this holds, then we have this function satisfying the heat equation.

Now, under what conditions can we differentiate integrals underneath the integral sign? I don't know if you remember from calculus, with real calculus, how that works. Yes?

AUDIENCE: [INAUDIBLE]?

PETER   Say that again.
KEMPTHORNE:

AUDIENCE: [INAUDIBLE] bit about, when does ux t converge?

PETER   When it converges? Well, let's see. OK, well, ux of t will exist if u0 is sufficiently smooth. And so if the initial
KEMPTHORNE: conditions are sufficiently smooth, then their first derivatives and second derivatives are bounded. And so these integrals are bounded, and they're absolutely integrable.

And so we're able to take derivatives underneath the integral sign. So when you have unbounded functions u0 that maybe are sufficiently smooth, then 1 can take limits as the bound on that function increases. And if those limits exist, then these solutions apply.

But what's important here is to see that we have this u0 of x is equal to this sum of the Dirac deltas multiplied by u0 of s value. So it's a weighted sum of Dirac delta. And the equation for the sum of these or integral of those weighted by u0 of s is given by this expression here.

Now, when we can differentiate underneath the integral sign, then basically the values of the-- or the u function can be represented as in a Taylor series with the terms in the Taylor series corresponding to these differentials. And they-- each of those satisfy the heat equation alone. So let's see how this works with where u0 is an indicator function. So we're thinking of-- let's see. Let me just write here. So let's see.

For this description, I want to actually think about, ultimately, a Brownian motion process xt that starts at 0. And so we have different paths, possible paths, for this Brownian motion process. And we're going to be interested in, basically, an indicator for the path lying between the intervals a and b. So let's consider a u0 function for the initial condition to be the indicator of whether a process lies between a and b. So there will be some paths that start with the indicator being 1 and others with the indicator being equal to 0.

And it turns out that the expectation of u0, the indicator function with respect to this Gaussian density is simply the probability of lying between a and b. So u of xt, as defined in the top, is simply the expectation of the indicator function, which is the probability of lying between a and b. And we can solve for this probability in terms of the cumulative normal CDF, standard normal, evaluated at the upper limit minus the mean divided by the standard deviation minus the CDF at the lower limit minus the mean divided by the standard deviation. So this u of xt, which is the difference of two functions of Gaussian densities, integrals of them, is a solution to this problem.

So now, in the next slide-- whoops OK, so that solution. All right, so the next set of concepts we want to introduce you to are solving these equations using properties of similarity and invariance. And so we're going to be considering changes in the time and space variables. So here, we have time variable, and here we have the space variable with Brownian motions.

And so what if we were to rescale time to a tau which is equal to alpha t, and we were to rescale x to a variable y, which is equal to some beta x. So we're just changing the units of time and the units of space in a constant way. Well, if we consider a v function which equals the u function, then we basically can map u into v by redefining u in terms of y and-- sorry, not u in terms of y, u in terms of tau.

And so if we apply the chain rule to ut, we get alpha vt. And if we apply the chain rule to the partial of u with respect to x, we get beta vy. And the second partial of u with respect to y is beta squared vyy. And so we end up getting our initial heat equation in x and t to be this equation in terms of tau and y.

And if alpha equals beta squared-- so we could arbitrarily, if we knew the-- if we're free to choose alpha and beta arbitrarily, we could choose alpha equal to beta squared. And we then get v tau is equal to lambda vy So this is a heat equation in terms of variables tau and y. And this must solve the original problem if we substitute back.

So we can think of looking for solutions that are functions that are invariant under rescaling time and rescaling the space. So we must have that u of xt is equal to u of root alpha x and alpha t if we consider this particular transformation. Now, with the choice of alpha and beta, those are arbitrary. We can choose alpha equal to 1 over t.

And then that means our u of xt is equal to u of x over root t and 1, which is a function just of x over root t. So we have a transformation of our original scales to a problem where our solution function v basically-- or our solution function u has this form. And so it's a function just of one variable.

And so when this is a function of one variable, if we substitute that into the heat equation, we get the derivative of h with respect to t on the left-hand side is equal to lambda times the partial of h with respect to x, second partial, on the right-hand side. So those are easily computed to give that top equation. And then we can say, well, let's stop carrying around x over root t everywhere. Let's just use y divided by root t as our variable of interest. And we get h prime of y times y over 2 equals lambda h double prime, which is given by this last equation here.

Now, this second derivative of h divided by the first derivative of h is equal to the derivative of the log of the first derivative. So it's 1 over the first derivative times the derivative of the first derivative. So we have that the derivative of the log of the derivative of h is equal to minus lambda over 2y. And we just integrate both sides to get the solution for h prime is just a constant times the exponential of the integral of this, which is y squared over 4 lambda-- minus y squared over 4, and lambda in the exponent.

And then we integrate this again to get a solution for h of y. And it turns out that we're basically integrating a Gaussian density function. This is proportional to a Gaussian density. So its antiderivative is-- or one antiderivative is actually the cumulative normal, standard normal evaluated at y over 2 root 2 lambda. So with this, we can substitute in our value for y back in terms of x. And we get this formula here for what the u function is.

Now, with all of this development, we have a formula for what u is over the entire time domain. But at time 0, it's the initial condition. And so the initial condition is the limit of this phi function as t goes to 0. And so when x is negative, this goes to minus infinity. And so we get 0 for the cumulative distribution function when x is less than 0.

When x is greater than 0, this goes to plus infinity, so we get 1. And if x equals 0, then it's always equal to 1/2. So we end up having this function which corresponds to the initial conditions. And this is the solution of the heat equation when the initial condition is given by this function. So let's see. Let's see if we want-- Well, let me just write, as a function of x, I guess, the function at 0. And then it's 1/2, and then it's 1 here. So this is the function u of x0 for this case.

Now, if we consider different linear combinations of this solution, we can actually consider taking-- instead of u of x0 equaling this, we can consider u of x corresponding to phi of x minus a over root 2 lambda t. And so this basic-- these first two bullets correspond to shifting this u of x0 to a values a and b. And then, with other one, b. Let's see. Sorry, the-- so at a, we have the same. And then at b we have the same of 0 up to b and then a half at b and then 1 beyond.

So we basically have two step functions centered where the steps are at x equals and at x equals b. And so both of these u functions will satisfy the heat equation. And if we consider the difference of the two functions, that will also satisfy the heat equation.

And so if we look at the difference in the initial conditions for these two solutions, we basically have the indicator of being between a and b. So one function for b is 1 forever. And then the function for a is 1 up to-- or sorry. Both functions are equal to 1 when x is greater than b, in which case we get 0. And otherwise, we get 1 when a is between x and b and 1/2 otherwise.

So what we end up having is, for any initial conditions that are representable as a step function or a sum of step functions, we can get a solution in this way. So for our initial conditions, u0, we can basically use limits of these step functions to approximate any initial condition.

Well, with stochastic differential equations that we want to solve, we consider generalizations of the drift and volatility parameters of the process. And so at the top here is a generalization of simple geometric Brownian motion, where the drift and the volatilities depend upon time as well as the level of the process x. And so what we want to be able to do is understand when solutions to these differential equations exist.

And there are conditions that guarantee solutions. One of them is a Lipschitz condition which says that the change in the mu function at a fixed time t for different values of x and y plus the change in the volatility for the same t with different x's and y's, that those are bounded by some constant times the difference between x and y. And if we have the-- so this is a space time condition.

And with the spatial growth condition, we also want the magnitude of the drift term plus the magnitude of the squared volatility term to be bounded by 1 plus the square of the x values. So what we have here are conditions for the drift not getting too wild and the volatility not getting too wild in order for there to be solutions to this.

A really important theorem that we won't get into but which are dealt with in advanced books on stochastic calculus deal with, the existence and uniqueness of solutions to these stochastic differential equations. And so if we consider two processes, xt and yt, that are both solutions to the stochastic differential equation, then subject to the conditions that we indicated, the two solutions are actually equal to each other almost everywhere. So we don't get different limits or different solutions to the equations. So they're essentially unique, and "essentially" means that, with probability 1, the paths are the same.

Now, in solving these stochastic differential equations, a method for solving these that turns out to be very useful is a method called coefficient matching. And so let's just illustrate coefficient matching for geometric Brownian motion. So here's the differential equation for geometric Brownian motion.

And suppose we have a solution that's some function of time and the underlying Brownian motion. Let's apply the Itô's formula to x of t. And we get dx of t is equal to this expression dt plus this expression dBt. And so the-- basically this dx of t is applying the chain rule to f of t and Bt. And we can then say, well, let's match mu x of t with this red term here and match sigma x of t, the coefficient of dBt with df dx.

And if we do that, well, the first or the second equation there, sigma f equals df dx, gives a solution of f equaling the exponential sigma x plus g of t. So this first equation, mu f equals-- or sorry, the second equation is sigma f df dx is equal to-- let me just write that out. So we have-- we have sigma f is equal to df by dx. And so we get sigma is equal to df by dx over f. And so this is equal to d by dx of log of f of x.

So f of x is going to be the exponential of the integral of sigma with respect to x And the g of t function will just be an arbitrary constant of that equation. And then, if we substitute this form formula for f into the first equation, we get this differential equation for f. And with this differential equation for f, mu f equals g prime f plus sigma squared over 2f, notice that f is a constant factor in each of those terms. So we can eliminate that constant factor or term, the same term, and we get g prime is equal to mu minus sigma squared over 2.

So with that, we end up getting a solution for f, which is the initial value x0 times the exponential sigma x plus this factor times little t. And that factor is the adjusted drift of the geometric Brownian motion. So we get that-- we basically get our solution for geometric Brownian motion coming out from that differential equation.

Now, an interesting paradox is that the drift term is this factor, and the variance is sigma t. Our standard deviation is sigma t for the value of a geometric Brownian motion process. But if our drift parameter is smaller than sigma squared over 2, then this mu star is going to be negative.

And so if we have a Brownian motion process, a geometric Brownian motion process where on the log scale the drift is negative, then this is going on the log scale to go to minus infinity as t grows large. And so the probability that the geometric Brownian motion is smaller than any constant is going to go to 1 as time grows large.

So when the mu terms, the drift term, is small relative to sigma squared over 2 less than, then the likelihood of this process going to 0 occurs with probability approaching 1. But complementing that is that the mean value of x of t is growing exponentially at the rate mu t, exponential rate mu t. So with geometric Brownian motion, we have this process that has a mean level that's growing arbitrarily large, but the process is, with probability approaching 1, approaching 0 for all paths.

So that's an interesting paradox. But it appears consistent with what actually happens often in equity markets, where you'll have some stocks that do really, really well and others that haven't done so well and go close to 0. On average, maybe there's a mean growth rate, but many of them have ultimate values that approach 0.

Well, in the remaining time, there's an introduction here to the Ornstein-Uhlenbeck process, which basically is a stochastic differential equation where instead of a constant drift term, we have a drift rate that depends on how far away the process is from the mean level. And so this model is one that arose with physicists studying this, where you look at the average velocity of molecules in a container with a certain heat or temperature. And so particles should be moving around at some average velocity. And this model is one way of trying to characterize that.

Outside the range of physics, people have used this model in finance. Vasicek used it as a model for interest rate dynamics, where there's a mean level that is expected in the long run, and there's variation about that. When the process is above the average level, it tends to drift down. And when it's below, it tends to drift up.

So with this Ornstein-Uhlenbeck process, we can consider a solution which is a product process. Now, you may recall with some differential equation problems where you want to solve a differential equation with two variables. And maybe the solution is a product of one variable times a product of another-- times the a function of the other variable. So in this case, we can think of a possible solution being given by an a of t function times the integral of the Brownian motion process with respect to a little b of s differentiable function.

So if we just tentatively assume this possible solution, we can apply the chain rule to get this first equation. dXt is the derivative of a times the bracketed term, then plus a of t times the derivative of the right-hand term with respect to t. And we can then use coefficient matching as per these expressions.

And at the end of the day, we end up getting our Ornstein-Uhlenbeck process. If we start at x0, it-- well, let's see. If we consider the process with mu equal to 0 and initial value of x0, then it decreases or increases exponentially to 0. And then the extra term is this integral of Brownian motion with the exponential term.

And what's interesting about this process is that it has a limiting distribution that's not like a non-stationary random walk of Brownian motion. But it's centered at the mean level, which was assumed in this development to be 0, and with some constant variance. So there's an ergodic property to this process. The long-term distribution is simply this fixed distribution with constant mean and variance.

Well, just to finish up, we can extend stochastic differential equations in one space variable and time to multi dimensions. And so we can think of having m dimensional processes evolving for the functions. So Xt could be m-dimensional. And then we would have a drift rate corresponding to different components of the dimensions mu 1 to mu m, different Brownian motion terms corresponding to m independent Brownian motions and then a covariance matrix reflecting how similar increments of x are related to each other in terms of the covariance terms here in this expression.

So we can have multiple multi-dimensional stochastic differential equations. And so here's just an illustration of thinking about mean reverting velocity and position and how that would be set up. So the theory of stochastic differential equations is very powerful. And it's been exploited a lot in physics and engineering.

So with that, let's see, I just want to highlight some useful references for stochastic differential equations. There's this really excellent monograph on by Oksendal which is very mathematical, not necessarily motivated by finance but motivated by physics problems, as well. And then the other references here are really good ones in quantitative finance. So we have our guest lecturer, John Hull, to thank for giving us this book on *Options, Futures, and Other Derivatives.* All right.