Stochastic Calculus, Courant Institute, Fall 2021
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2021/index.html

Week 4

Jonathan Goodman, November, 2021

1 SDE models, diffusions

This is the most important of the seven classes. It describes how stochastic differential equations, SDEs, are used to create models of random processes in continuous time with continuous paths. The class starts with some terminology and the philosophy related to SDE models. This and the next section are all definitions and theory. The applications come in Sections 3 and 4, and in the exercises. There is less motivation here because the motivation is similar to Week 2 (for Ite's lemma) and Week 3 (for backward equations).

The fundamental theorem of ordinary calculus says that a function that can be written as an indefinite integral is differentiable. Written in the language of stochastic calculus, this ordinary fact is

$$Y_t = \int_0^t a_s \, ds \implies dY_t = a_t \; .$$

In stochastic calculus, a function that can be represented as an indefinite integral is called an *Ito process*:

$$X_t = \int_0^t a_s \, ds + \int_0^t b_s dW_s \; . \tag{1}$$

An Ito process is a sum of two indefinite integrals, an "ordinary" one and an Ito integral. The Ito integral makes X_t random, but more randomness is allowed: the integrands a_t and b_t may be random. As always, the integrand in the Ito integral must be non-anticipating (adapted). For example, we could have

$$X_t = \int_0^t s \, ds + \int_0^t W_{\frac{s}{2}} \, dW_s$$

The Ito integrand in this example is adapted in that it is a function of the path $W_{[0,t]}$, but it is not a function of W_t alone. An Ito process is a process that has an Ito differential, which is

$$dX_t = a_t dt + b_t dW_t \; . \tag{2}$$

The integrands a_s and b_s on the right side of the Ito process formula (1) determine the *infinitesimal mean* and *infinitesimal variance* of X_t. Suppose $dt > 0$ is an infinitesimal but non-zero increment of time (a more mathematical version is coming) and $dX = X_{t+dt} - X_t$ is the corresponding increment of X. The infinitesimal mean is a_t means

$$a_t \, dt = \mathrm{E}\big[\, dX_t \mid W_{[0,t]} \,\big] \ . \tag{3}$$

The conditional expectation on the right is the conditional expectation given that you know everything relevant that happened up to time t. This supposes that the Brownian motion path is the only source of randomness that is relevant to the process X_t. In more general situations, you might bet on one Brownian motion using input from an independent Brownian motion. We discuss this issue more below.

It might be more familiar to math-trained people to take a small but not infinitesimal increment $\Delta t > 0$. The corresponding increment of X is $\Delta X = X_{t+\Delta t} - X_t$. The infinitesimal mean formula is

$$a_t \, \Delta t = \mathrm{E}\big[\, \Delta X \mid W_{[0,t]} \,\big] + o(\Delta t) \ . \tag{4}$$

I think of the informal version (3) as a shorthand way to write this. I believe (personal belief, others disagree) that more formal statements like (4) do not make people who use them more likely to reason correctly. I see plenty of finance and economics papers written in terms of the fanciest mathematical formalism that make elementary reasoning mistakes that would be less likely using simpler and more intuitive reasoning such as (3).

The infinitesimal variance is

$$b_t^2 \, dt = \mathrm{var}\big(\, dX \mid W_{[0,t]} \,\big) \ . \tag{5}$$

This is the same as the expected square of the increment

$$b_t^2 \, dt = \mathrm{E}\Big[\, (dX)^2 \mid W_{[0,t]} \,\Big] \ . \tag{6}$$

The difference is whether or not the mean is subtracted. Subtracting the mean doesn't change the infinitesimal variance formula, because

$$\begin{aligned} \mathrm{var}(\, dX \mid \cdot\,) &= \mathrm{E}\Big[\, (dX)^2 \mid \cdot \,\Big] - \left(\mathrm{E}[\, dX \mid \cdot\,]\right)^2 \\ &= \mathrm{E}\Big[\, (dX)^2 \mid \cdot \,\Big] - a_t^2 \, dt^2 \ . \end{aligned}$$

In the language of Week 2, the dt^2 term on the right is *tiny* and can be ignored. In the Δt language, (5) would be

$$b_t^2 \Delta t = \mathrm{var}\big(\, \Delta X \mid W_{[0,t]} \,\big) + o(\Delta t) \ . \tag{7}$$

This is equivalent to an expected square formula

$$b_t^2 \Delta t = \mathrm{E}\Big[\, (\Delta X)^2 \mid W_{[0,t]} \,\Big] + o(\Delta t) \ . \tag{8}$$

The derivation is almost the same. If you're not used to "big Oh" and "little oh" reasoning, you can use the less formal version with differentials given above, or you can look it up in Wikipedia.

$$\begin{aligned}
\operatorname{var}(\,\Delta X \mid \cdot\,) &= \mathrm{E}\Big[\,(\Delta X)^2 \mid \cdot\,\Big] - (\mathrm{E}[\,dX \mid \cdot\,])^2 + o(\Delta t) \\
&= \mathrm{E}\Big[\,(\Delta X)^2 \mid \cdot\,\Big] - \big(a_t^2\,\Delta t + o(\Delta t)\big)^2 + o(\Delta t) \\
\operatorname{var}(\,\Delta X \mid \cdot\,) &= \mathrm{E}\Big[\,(\Delta X)^2 \mid \cdot\,\Big] + o(\Delta t)\;. \qquad (9)
\end{aligned}$$

The infinitesimal mean and infinitesimal variance formulas come from the Ito process representation and properties of integrals and continuous functions. For example,

$$\mathrm{E}\big[\,\Delta X \mid W_{[0,t]}\,\big] = \;\mathrm{E}\left[\int_t^{t+\Delta t} a_s\,ds \mid W_{[0,t]}\right]$$

The expectation of the Ito integral part is zero. If a_s is a continuous function of s, then

$$\int_t^{t+\Delta t} a_s\,ds = a_t\Delta t + o(\Delta t)\;.$$

For the infinitesimal variance, which is the same as the infinitesimal square, let Y_t be the Brownian motion integral

$$Y_t = \int_0^t b_s\,dW_s\;.$$

The increment of this is

$$\Delta Y = \int_t^{t+\Delta t} b_s\,dW_s$$

The Ito isometry formula from Week 2 gives

$$\mathrm{E}\left[\left(\int_t^{t+\Delta t} b_s\,dW_s\right)^2 \mid \cdot\,\right] = \int_t^{t+\Delta t} \mathrm{E}\big[\,b_s^2 \mid \cdot\,\big]\;ds\;.$$

The conditioning is the Brownian motion path up to time t. This means that in the conditional expectation b_t is known and $b_s \approx b_t$ if $s \approx t$ (because b_s is a continuous function of s). Therefore

$$\int_t^{t+\Delta t} \mathrm{E}\big[\,b_s^2 \mid \cdot\,\big]\;ds = b_t^2\Delta t + o(\Delta t)\;.$$

You can check, as we checked (9), that the ds integral in (1) changes this by a "tiny" amount, which means $o(\Delta t)$. This shows that the Ito process (1) has infinitesimal mean (4) and infinitesimal variance (7) if a_t and b_t are continuous functions of t. What I call "infinitesimal variance" is more commonly called *quadratic variation*. The infinitesimal mean is *drift*.

This reasoning is used to write integral expressions for stochastic processes that satisfy specified drift and quadratic variation conditions. Suppose you have a stochastic process X_t and some reasoning suggests that the drift is a_t and the quadratic variation is μ_t. You pick any square root $b_t^2 = \mu_t$. Then the integral (1) has the desired properties. From this we learn that an Ito process is completely determined by its infinitesimal mean and variance. You might of this as Gaussian-like. Gaussian random variables are determined by their mean and variance. But Ito processes do not have to be Gaussian.

Do not confuse the terms *Ito process* and *diffusion process*. An Ito process is any process with an Ito differential. A diffusion process is a stochastic model of some system. By "stochastic model", we mean that the statistics of dX_t (mean and variance) are determined by X_t and not by X_s for $s < t$. Technically, an Ito process is a diffusion process if it also is a Markov process. The *Markov property* is that the distribution of the future, conditional on the past, is the same as the distribution of the future conditional on the present. For an Ito process, this means that the Ito coefficients a_t and b_t are determined by X_t and t in a deterministic way: $a_t = a(X_t, t)$, and $b_t = b(X_t, t)$. This turns the Ito differential expression (2) into

$$dX_t = a(X_t, t)\, dt + b(X_t, t)\, dW_t \; . \tag{10}$$

The Ito differential (2) is a *stochastic differential equation* (*SDE*) if the differential coefficients are functions of X_t and t only, as in (10).

Week 1 defined *diffusion process* in a seemingly different way. Then, it just was a process with continuous "sample paths" whose infinitesimal mean and variance were specified functions of X_t. Now, we would add nuance by saying that these infinitesimal means and variances are conditional on the path $X_{[0,t]}$. At that time there was no mention of Brownian motion in the definition of diffusion process. Normally, a modeler would create a diffusion process stochastic model of a system by reasoning about infinitesimal mean and variance, rather than by a relation to Brownian motion.

The version here, with its connection to Brownian motion, is useful for technical manipulations. This version has the feature than the noise level is specified by the standard deviation, which is $b(x,t)$, not the variance, which is $\mu(x,t) = b(x,t)^2$. A function $X(t, W_{[0,t]})$ that satisfies (10) is called a *strong solution* of the SDE. A probability distribution on path space that has the right infinitesimal mean and variance is called a *weak solution*. This technical distinction is unimportant in moat practical modeling applications of stochastic calculus.

2 Ito's lemma for diffusion processes

Let X_t be an Ito process specified by either (1) or (2) For a general Ito process, as opposed to a diffusion process, the coefficients a_t and b_t are any random but non-anticipating functions. Suppose $f(x,t)$ is some function. The corresponding

"Ito's lemma" is

$$df(X_t,t) = \partial_x f(X_t,t)\,dX_t + \left[\partial_t f(X_t,t) + \frac{1}{2} b_t^2\, \partial_x^2 f(X_t,t) \right] dt \;. \tag{11}$$

You can derive this in two steps. First you expand df to second order in dX and first order in dt. This gives

$$df = \partial_x f(X_t,t)\,dX + \frac{1}{2}\partial_x^2 f(X_t,t)\,(dX)^2 + \partial_t f(X_t,t)\,dt \;.$$

Then you apply the "Ito rule", which is to replace $(dX)^2$ with its conditional expected value

$$\mathrm{E}\left[\, (dX)^2 \mid X_{[0,t]} \right] = b_t^2\, dt \;.$$

This gives the Ito's lemma formula (11).

It may seem surprising that you can replace $(dX)^2$ with its expected value. You get the "right answer" doing this, but not because the error is small. On the contrary, $(dX)^2$ is on the order of dt, and the difference between this and its expected value is also on the order of dt:

$$(dX)^2 - b_t^2\, dt \;\text{ is on the order of } dt \;.$$

You can see this in the variance, which is the expected value of the square of the difference to the mean value. If the difference is $O(dt)$ then the variance of $(dX)^2$ should be on the order of $(dt)^2$. That variance formula is true already for Brownian motion. Here is the calculation, done carefully for finite Δt. The independent increments property of Brownian motion implies that you the answer is the same whether or not you condition on $W_{[0,t]}$. This calculation has been used a few times already in this course:

$$\mathrm{var}\big(\, (\Delta W)^2\big) = 2\Delta t^2 \;.$$

This implies that the difference between $(\Delta W)^2$ and it's expected value (which is Δt) is typically on the order of Δt.

You can replace $(dX)^2$ with its expected value because the error in this approximation not only has mean zero, but is nearly uncorrelated from one time interval to the next. These errors quickly cancel in the mean. That is (usual notation, $t_k = k\Delta t$, and $\Delta W_k = W_{t_{k+1}} - W_{t_k}$),

$$\sum_{t_k < T} \Big[(\Delta W_k)^2 - \Delta t \Big] \;\to\; 0 \;, \;\text{ as } \Delta t \to 0 \text{ (in probability or almost surely).}$$

Going further, the *quadratic variation* of an Ito process is

$$[X]_T = \lim_{\Delta t \to 0} \sum_{t_k < T} \left(X_{t_{k+1}} - X_{t_k} \right)^2 \;.$$

This limit is

$$[X]_T = \int_0^T b_s^2\, ds \;. \tag{12}$$

This course is too short for a full proof of this formula, but I hope the explanations here give you some confidence and understanding of it. The differential form of the quadratic variation formula (12) is

$$d\,[X]_t = b_t^2\,dt\ .$$

The informal "Ito rule" is to replace $(dX_t)^2$ with $d[X]_t$. That's how we get the Ito lemma formula (11).

There is a version of Ito's lemma (11) that uses the increment of Brownian motion instead of the increment dX. This is derived from (11) using the Ito differential (2). The effect is:

$$\partial_x f(X_t,t)\,dX_t = \partial_x f(X_t,t)\,(a_t dt + b_t dW_t)\ .$$

The dW form of Ito's lemma for the Ito process X_t is seen to be

$$df = b_t \partial_x f\,dW_t + \left[\partial_t f + a_t\partial_x f + \frac{1}{2}b_t^2\,\partial_x^2 f\right] dt\ . \tag{13}$$

I write f and $\partial_x f$ for $f(X_t,t)$ and $\partial_x f(X_t,t)$ (etc.) to make the formulas easier to write and to read. The dW form (13) shows that the process $Y_t = f(X_t,t)$ is a martingale if the coefficient of dt is zero. Therefore:

$$f(X_t,t) \text{ is a martingale } \iff \partial_t f + a_t\partial_x f + \frac{1}{2}b_t^2\,\partial_x^2 = 0\ . \tag{14}$$

This allows us to derive backward equations for general diffusions using the second derivation of the backward equation for Brownian motion from Week 3.

Suppose X_t is a diffusion process that satisfies the SDE (10). Consider a value function for a final time payout

$$f(x,t) = \mathrm{E}[\,V(X_T) \mid X_t = x]\ . \tag{15}$$

We might have conditioned on the whole path $X_{[0,t]}$. But X_t is a Markov process, so conditioning on the whole path up to time t is the same as conditioning on the location of the path at X_t. The backward equation for this value function is

$$\partial_t f + a(x,t)\partial_x f + \frac{1}{2}b^2(x,t)\partial_x^2 f = 0\ . \tag{16}$$

This is consistent with the backward equation from Week 3 for Brownian motion, because Brownian motion has $a=0$ and $b=1$.

The backward equation follows, as it did in Week 3, from the martingale property, and uniqueness of the solution of the backward equation with final conditions. If f satisfies the PDE (16) and has the final values $f(x,T) = V(x)$, then, because it's a martingale,

$$\mathrm{E}\big[\,f(X_T,T) \mid X_{[0,t]}\big] = f(X_t,t)\ .$$

The left side is the conditional expectation of $V(X_t)$ (because of the final condition). The right side, conditioned on $X_t = x$, is $f(x,t)$. This shows that "the" solution to the backward equation with the right final condition is the value function. But there is only one solution and one value function, so they must be the same.

Here is another, better (in some cases) derivation of the backward equation. It is based on the *tower property*, which is that the expectation of a conditional expectation is the overall expectation. Consider a "running sum" functional

$$\int_0^T V(X_s)\,ds\ .$$

The value function for this is

$$f(x,t) = \mathrm{E}\left[\int_t^T V(X_s)\,ds \mid X_t = x\right]\ . \tag{17}$$

Consider an intermediate time t_1 with $t < t_1 < T$. The integral from t to T may separated into the parts before and after t_1:

$$\int_t^T V(X_s) = \int_t^{t_1} V(X_s)\,ds + \int_{t_1}^T V(X_s)\,ds\ .$$

The part after t_1 is included in the value function defined at t_1. Therefore

$$f(x,t) = \mathrm{E}\left[\int_t^{t_1} V(X_s)\,ds \mid X_t = x\right] + \mathrm{E}[\, f(X_{t_1},t_1) \mid X_t = x]\ .$$

We now take $t_1 = t + \Delta t$ and write $X_{t_1} = x + \Delta X$. For the first integral on the right, we approximate

$$\int_t^{t+\Delta t} V(X_s)\,ds = \Delta t V(x) + o(\Delta t)\ .$$

For the second expectation on the right, we approximate (standard notation, f means $f(x,t)$, and f_x is the partial derivative, etc.)

$$f(x+\Delta X, t+\Delta t) = f + f_x \Delta X + \frac{1}{2} f_{xx}\Delta X^2 + f_t \Delta t + \text{ smaller terms.}$$

Exercise 3 asks you to look at these "smaller terms". We put this approximation into the expectation:

$$\begin{aligned}
&\mathrm{E}[\, f(x+\Delta X, t+\Delta t) \mid X_t = x] \\
&= f + f_x \mathrm{E}[\,\Delta X \mid X_t = x] + \frac{1}{2} f_{xx}\mathrm{E}\big[\,\Delta X^2 \mid X_t = x\big] \\
&+ f_t\Delta t + \text{ smaller} \\
&= f + f_x a(x)\Delta t + \frac{1}{2} f_{xx} b(x)^2 \Delta t + f_t \Delta t + \text{ smaller}\ .
\end{aligned}$$

These results may be assembled to

$$f = \Delta t V(x) + f + \Delta t a(x) f_x + b^2(x)\frac{1}{2} f_{xx}\Delta t + f_t \Delta t + \text{ smaller} .$$

We cancel f from both sides, divide by Δt, let Δt go to zero, and the result is

$$0 = a(x) f_x + \frac{b^2(x)}{2} f_{xx} + f_t + V(x) . \tag{18}$$

3 Ornstein Uhlenbeck

The *Ornstein Uhlenbeck* process, or *OU* process, is a diffusion process that satisfies the SDE

$$dX_t = -aX_t\, dt + \sigma dW_t . \tag{19}$$

The infinitesimal mean is $-aX_t dt$. If $a > 0$ (the usual case), this moves X closer to $x = 0$. The coefficient a is sometimes called the *mean reversion* coefficient or mean reversion *rate*, with the understanding that reverting to the mean means reverting to zero. The infinitesimal variance is $\sigma^2 dt$. The noise level is uniform, independent of X. Without noise X_t would "revert" to zero exponentially fast. The noise takes X away from zero. The long time behavior of OU paths is a balance between noise and mean reversion.

The OU model was used by Einstein to model the random velocity of a tiny particle in a fluid. In this model, X_t is the velocity of the particle, a is a friction coefficient that represents the fact that the fluid slows the particle if it is moving. The noise represents the effect of individual fluid molecules hitting the particle and getting it to move. If the particle is much bigger than a fluid molecule, each individual collision has a small effect. The term $\sigma\Delta W_t$ represents the cumulative effect of many such collisions in a time Δt. The central limit theorem suggests that the effect of many collisions is approximately Gaussian, as is ΔW in the model.

The OU model is useful as a model for physical and economic/financial systems, and also because it is an example that may be solved in closed form. The solution suggests things that other diffusion processes may or may not do. One approach to the solution uses the method of *integrating factors*. This method may be familiar from a first course on differential equations, but it is given here in the notation of stochastic calculus. The method applies to differential equations in which the *endogenous* terms (the ones involving model variables) are linear in the model variable. In this case, $-aX_t dt$ would be called endogenous and σdW_t *exogenous* (given from the outside) We write the equation with the endogenous terms on the left and exogenous on the right:

$$dX_t + aX_t dt = \sigma dW_t .$$

The integrating factor allows us to combine the endogenous terms into a single "total differential".

We multiply both sides by the integrating factor e^{at}. On the left, you have

$$e^{at} dX_t + e^{at} a X_t dt \ .$$

In ordinary calculus, this might expressed in terms of derivatives (rather than differentials) as

$$e^{at}\frac{dX}{dt} + e^{at} aX = \frac{d}{dt}\left(e^{at}X\right) \ .$$

Ito's lemma is the mechanism for doing such calculations in stochastic calculus. We apply the differential version (2) to the function $f(x,t) = e^{at}x$. The following calculation uses the differential formula (2). We express partial derivatives of f using subscripts, so $\partial_x f = f_x$, $\partial_x^2 f = f_{xx}$, and $\partial_t f = f_t$. At the same time, X_t is just the value of X at time t.

$$\begin{aligned}
f(x,t) &= e^{at}x \\
f_x &= e^{at} \\
f_{xx} &= 0 \\
f_t &= ae^{at}x \\
df(X_t,t) &= f_x\, dX_t + \frac{1}{2} f_{xx}\, (dX_t)^2 + f_t\, dt \\
&= e^{at} dX_t + ae^{at} X_t dt \ .
\end{aligned}$$

The result is

$$e^{at}\, dX_t + ae^{at}X_t dt = d\left(e^{at}X_t\right) \ .$$

This shows that the SDE (19) is equivalent to

$$d\left(e^{at}X_t\right) = e^{at}\sigma dW_t \ .$$

The next step is to integrate both sides with respect to t. On the left side we have

$$\int_0^t d\left(e^{as}X_s\right) = e^{at}X_t - X_0 \ .$$

The right side is an Ito integral with integrand e^{as}. The result is

$$e^{at}X_t - X_0 = \int_0^t e^{as} dW_s \ .$$

Finally, you multiply by e^{-at} to arrive at a formula for X_t

$$X_t = e^{-at}X_0 + \sigma\int_0^t e^{-a(t-s)}\, dW_s \ . \tag{20}$$

This solution formula reveals important facts about the OU process. One is that the OU process "forgets" its initial state, X_0. As $t \to \infty$, the influence of the initial state, which is $e^{-at}X_0$ disappears exponentially. This is natural in Einstein's fluid friction model. The relevance of the initial velocity fades exponentially because of friction.

The solution formula (20) also shows that the distribution of X_t converges as $t \to \infty$, and it converges to a mean zero Gaussian. The "innovation" part of X_t is the Ito integral on the right of (20). This integral is "obviously" Gaussian (explanation below), as is any Ito integral with respect to Brownian motion where the integrand is not random.

Here is an explanation of the fact that the distribution of the Ito integral is Gaussian. Any Ito integral with a fixed deterministic integrand is Gaussian. Such an integral defines a random variable Z:

$$Z = \int_0^t c_t dW_s \ . \tag{21}$$

We choose a small $\Delta t > 0$ and approximate Z by

$$Z^{\Delta t} = \sum_{t_k < t} c_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \ .$$

This is Gaussian because every term on the right is Gaussian, the sum of independent Gaussians is Gaussian, and the coefficients c_{t_k} are just fixed numbers (not random). By contrast, the example $\int W_s \, dW_s$ has approximations with terms $W_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right)$. These are not Gaussian because the product of Gaussians typically is not Gaussian. The expected value of $Z^{\Delta t}$ is zero (all zero expectations on the right) and the variance is (because the terms on the right are independent)

$$\text{var}\left(Z^{\Delta t} \right) = \sum_{t_k < t} c_{t_k}^2 \Delta t \ .$$

Thus, Z is the limit of Gaussians $Z^{\Delta t}$ whose mean is zero and whose variances converge to

$$\text{var}(Z) = \int_0^t c_s^2 \, ds \ .$$

The variance of $Z^{\Delta t}$ is a Riemann sum that converges to this integral.

We just showed that Z is Gaussian and gave a formula for the variance. The variance formula is a special case of the Ito isometry formula

$$\text{var}\left(\int_0^t b_s \, dW_s \right) = \int_0^t \text{E}\left[b_s^2 \right] \, ds \ .$$

This variance formula applies whether or not the integral is Gaussian. But, when the integrand c_s is not random, there is no need for the expectation on the right. For our OU process, the variance is

$$\text{var}(X_t) = \sigma^2 \int_0^t \left(e^{-a(t-s)} \right)^2 ds = \sigma^2 \int_0^t e^{-2as} \, ds = \frac{\sigma^2}{2a} \left(1 - e^{-2t} \right) \ . \tag{22}$$

In the limit $t \to \infty$ this is just $\frac{\sigma^2}{2a}$. To check that this makes sense, note that it increases as σ increases (more noise means more variance). It decreases as

a increases, because stronger mean reversion (stronger friction) keeps X_t closer to zero, on average.

The OU model (19) is an example of an *equilibrium model*. This means that the distribution of X_t has a limit as $t \to \infty$. If $u(x,t)$ denotes the PDF of X_t (notation that was used earlier), then the following limit exists

$$u(x) = \lim_{t \to \infty} u(x,t) \; . \tag{23}$$

It is a fact that if sample paths are unlikely to "run off to infinity", then the limit (23) exists. The mean reversion term $-aX_t dt$ in the OU process is strong enough to keep paths from wandering to infinity. In this case there is a formula for the limiting PDF $u(\cdot)$. It is Gaussian with mean zero and variance $\frac{\sigma^2}{2a}$ (see (22) for the limiting variance), so

$$u(x) = \sqrt{\frac{a}{\pi \sigma^2}} e^{-\frac{a x^2}{\sigma^2}} \; .$$

The limiting distribution is called the *steady state* distribution. It is rare to have a formula for it.

The steady state is for the probability distribution, not the path X_t. An OU path does not settle down to a specific value as $t \to \infty$. In fact, it $X_t = x$, then the distribution at time $T > t$ quickly "forgets" x. The path is constantly changing. Only the probability density has a limit as $t \to \infty$. It is a *statistical* steady state, not a steady state for paths.

You can find some of the OU formulas using the trick of putting a differential inside the expectation. If X_t is an Ito process, then

$$d\,\mathrm{E}[f(X_t,t)] = \mathrm{E}[\,df(X_t,t)] \; . \tag{24}$$

The point is that you can apply Ito's lemma to calculate df and then take the expectation. As a first example, let X_t by an OU process whose mean is $m(t) = \mathrm{E}[X_t]$. Then

$$\begin{aligned} d\,m(t) &= d\,\mathrm{E}[X_t] \\ &= \mathrm{E}[dX_t] \\ &= \mathrm{E}[-aX_t dt + \sigma dW_t] \\ &= -a\mathrm{E}[X_t]\,dt + 0 \\ &= -a\,m(t)\,dt \; . \end{aligned}$$

This may be written in terms of derivatives as

$$\frac{d}{dt} m(t) = -am(t) \; .$$

If X_0 is deterministic, then $m(0) = X_0$ and $m(t) = e^{-at} X_0$, which we already knew. This derivation might seem more straightforward. In a statistical steady state, $\frac{d}{dt} m(t) = 0$. This is part of the definition of "steady state", the probability distribution does not change with time. In particular, $dm = 0$, which implies that $m = 0$. The steady state has mean zero.

For a second example, the second moment is

$$s(t) = \mathrm{E}\left[\, X_t^2 \right] \ .$$

We calculate its dynamics using the same idea, but more of the Ito calculus, including the fact that dW_t is independent of X_t (because dW_t is in the future of t).

$$\begin{aligned} d\,s(t) &= d\,\mathrm{E}\left[\, X_t^2 \right] \\ &= \mathrm{E}\left[\, dX_t^2 \right] \\ &= \mathrm{E}\left[\, 2X_t dX_t + (dX_t)^2 \right] \\ &= \mathrm{E}\left[\, 2X_t(-aX_t\,dt + \sigma dW_t) + \sigma^2\,dt \right] \\ &= -2a\mathrm{E}\left[\, X_t^2 \right]\,dt + \sigma^2\,dt \\ d\,s(t) &= -2as(t)\,dt \ + \ \sigma^2\,dt \ . \end{aligned}$$

In statistical steady state, $ds(t) = 0$. This allows us to solve for $s(t)$ in steady state. The result is

$$\mathrm{E}\left[\, X^2 \text{ steady state} \right] = s(\text{steady state}) = \frac{\sigma^2}{2a} \ .$$

Since the mean in steady state is zero, the expected square is also the steady state variance. The formula is what we had before. You can, if you want, combine our formula for the expected first and second moments to verify that the previous formula (22) satisfies the differential equation derived in this way.

4 Geometric Brownian motion

A *geometric Brownian motion* is a diffusion process that satisfies the SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t \ . \tag{25}$$

The parameter μ is the *expected rate of return* (or just *expected return*). The parameter σ is the *volatility*. This is a simple model of the price of a traded asset (a stock). The change in S, in both the mean and the noise terms, is proportional to S_t. This means that the probability of $S = 100 \to S = 102$ is the same as the probability $300 \to 309$. Both are 2% increases.

As with the OU model, this may be solved using tricks from a course on differential equations, brought to the stochastic world using Ito's lemma. The differential equations method is *separation of variables*. This means putting S and dS on one side of the equation, dt on the other, then integrating both sides. In the stochastic world, we have to be careful to differentiate correctly (Ito's lemma) and to take care of the Brownian motion noise term. We start with

$$\frac{1}{S_t} dS_t = \mu dt + \sigma dW_t \ . \tag{26}$$

The indefinite integral of the right side is

$$\int_0^t \mu ds + \int_0^t \sigma dW_s = \mu t + \sigma W_t \ .$$

In ordinary calculus, you would look at the left side and recognize $\frac{1}{s}\frac{ds}{dt} = \frac{d}{dt}\log(s)$. The corresponding calculation for the diffusion process requires Ito's lemma:

$$\begin{aligned} f(s) &= \log(s) \\ f_s &= \frac{1}{s} \\ f_{ss} &= -\frac{1}{s^2} \\ f_t &= 0 \\ (dS_t)^2 &= \sigma^2 S_t^2 dt \ . \end{aligned}$$

Ito's lemma (11) then gives

$$d\log(S_t) = \frac{1}{S_t} dS_t - \frac{1}{2}\,\sigma^2 S_t^2\,\frac{1}{S_t^2}\,dt \ .$$

We re-write this for our calculation in the form

$$\frac{1}{S_t} dS_t = d\log(S_t) + \frac{\sigma^2}{2} dt \ .$$

This puts (26) into the form

$$d\log(S_t) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t \ .$$

We integrate both sides from $s = 0$ to $s = t$ to find

$$\log(S_t) - \log(S_0) = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t \ .$$

With a little more algebra, this is the solution formula

$$S_t = S_0 e^{\sigma W_t + \left(\mu - \frac{\sigma^2}{2}\right) t} \ . \tag{27}$$

We can check that this is correct using an Ito calculation for the right side of

(27),

$$
\begin{aligned}
S_t &= f(W_t, t) \\
f(w,t) &= S_0 e^{\sigma w + \left(\mu - \frac{\sigma^2}{2}\right)t} \\
f_w &= \sigma f \\
f_{ww} &= \sigma^2 f \\
f_t &= \left(\mu - \frac{\sigma^2}{2}\right) f \\
(dW_t)^2 &= dt \\
df(W_t,t) &= f_w\, dW_t + f_t\, dt + \frac{1}{2} f_{ww} (dW)^2 \\
&= \sigma f dW_t + \left(\mu - \frac{\sigma^2}{2}\right) f\, dt + \frac{\sigma^2}{2} f\, dt \\
&= \mu f dt + \sigma f dW_t \\
dS_t &= \mu S_t dt + \sigma S_t dW_t \; .
\end{aligned}
$$

Exercise 2 looks at this solution from different points of view.

The solution formula (27) allows you to find the PDF of S_t. The distribution is called *log-normal* because $\log(S_t)$ is normal. Exercise 5 asks you to calculate and plot this density. Other than plotting, I don't know much use for it. If I want to calculate expectations of functions of S_t, I use the Gaussian distribution of W_t instead. We will see this in Week 5, when we derive the *Black Scholes formula*. The formula (27) shows that $S_t > 0$. You will see in Exercise 5 that computations using the Euler Maruyama method (29) might fail to give positive results if the time step is too large.

The formula (27) has the striking feature that the typical growth rate of S_t, what you typically see for large t, is $\mu - \frac{\sigma^2}{2}$ rather than μ. This is because the term $\left(\mu - \frac{\sigma^2}{2}\right) t$ is much larger than ("dominates") σW_t. The typical size of W_t is the square root of the variance, which is $\sqrt{t}$ and much smaller than t for large t. For example, if $\mu = 0$ the S_t is a martingale (no dt term in dS_t). In that case, the expected value of S_t does not change, and is equal to S_0. However, the $-\frac{1}{2}\sigma^2 t$ in the exponent dominates σW_t in the exponent, which makes S_t go to zero almost surely. This GBM converges to zero in distribution. You will see this in the plots of Exercise 5. The expectation

$$
\mathrm{E}[S_T] = S_0
$$

is achieved by having rare paths much larger than S_0 while typical paths are much smaller. This is an example of a strongly *skewed* distribution with a PDF that is not symmetric around the mean.

Here is a simple "intuitive explanation" of the the fact that S_t goes to zero almost sure when it's a martingale. Suppose, in discrete time, there are equal probabilities of going up or down by 50%. If S goes up 50%, we multiply by

1.5. If it goes down 50%, we multiply by .5. If it goes up and then down, we multiply by both factors:

$$S \longrightarrow \frac{1}{2}\left(\frac{3}{2}S\right) = \frac{3}{4}S < S\ .$$

The expected value after two steps is still S (the martingale value) because the expected value is

$$\begin{aligned}
&\Pr(\text{down down})\cdot\frac{1}{4} + \Pr(\text{down up})\cdot\frac{3}{4} + \Pr(\text{up down})\cdot\frac{3}{4} + \Pr(\text{up up})\cdot\frac{9}{4} \\
&= \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{2}\frac{3}{4} + \frac{1}{4}\frac{9}{4} \\
&= \frac{1+6+6}{16} = 1\ .
\end{aligned}$$

However, S goes down by at least 25% in three out of four of the outcomes.

5 Computational methods

This section discusses two computational problems. One is generating sample paths for a diffusion process from the SDE. The other is finite difference methods for solving the backward equation. Most modeling projects involve first formulating a stochastic model such as an SDE and then doing computer work of some kind to explore the behavior of the model. The material here should seem natural, given similar methods for simpler problems we have already done. These methods do not produce the exact solution, neither for sample paths nor for backward equations. They have parameters Δt and/or Δx. As these parameters go the zero, the computed solution converges to the actual (model) solution. The trick in practice is to choose Δt or Δx small enough to get the accuracy you need. The computer time increases as Δt and Δx decrease, so you don't want to take these parameters smaller than necessary.

Consider the SDE (10). Choose a Δt and approximation times $t_k = k\Delta t$. Denote the values of the approximate sample path by

$$X_k^{\Delta t} \approx X_{t_k}\ .$$

An optimistic approximation to the SDE for non-zero Δt would be

$$X_{k+1}^{\Delta t} = a(X_k^{\Delta t})\,\Delta t + b(X_k^{\Delta t})\,\Delta W_k\ . \tag{28}$$

Everything here is known, except possibly ΔW_k. We take this to be the increment of Brownian motion over the time increment Δt. These increments (as we have seen since Week 1) are Gaussian with mean zero and variance Δt. You can ask the Gaussian random number generator to give you random variables with that distribution (see the code `StockSim.py`), or you can ask for $Z_k \sim \mathcal{N}(0,1)$

and take $\Delta W_k = \sqrt{\Delta t} Z_k$. In this case, the computer program would implement the formula

$$X^{\Delta t}_{k+1} = a(X^{\Delta t}_k)\, \Delta t + b(X^{\Delta t}_k)\, \sqrt{\Delta t}\, Z_k \ , \quad Z_k \sim \mathcal{N}(0,1) \ . \tag{29}$$

This is the *Euler Maruyama* method, which is how diffusion processes are usually simulated.

This may seem odd, particularly if you have experience with numerical methods for ordinary or partial differential equations. For those problems there are families of sophisticated and extremely accurate methods, including Runge Kutta methods, finite element methods, and so on. There are whole graduate courses devoted to such methods. The simplest method, which is Euler's method, is explained in the first class. The rest of the course explains better methods. Yet, for SDE, there do not seem to be methods that are much better than the simple Euler Maruyama method (28).

6 Exercises

1. Let X_t be an OU process with a deterministic starting point $X_0 = x_0$. Let $u(x,t)$ be the PDF for X_t.

   (a) Use the solution formula (20) to show that $u = \mathcal{N}(\mu_t, v_t)$ and find formulas for the mean μ_t and variance v_t. This is basically done in the text, so summarize and write a formula for $u(x,t)$

   (b) Suppose X_t satisfies the SDE (10) and $u(x,t)$ is the PDF of X_t. The *forward equation* is

   $$\partial_t u = -\partial_x \left(a(x) u(x,t) \right) + \frac{1}{2} \partial_x^2 \left(b(x)^2 u(x,t) \right) \ .$$

   Check by explicit calculation that the solution formula from part (a) satisfies the forward equation for (19).

   (c) Show that if $X_t \sim \mathcal{N}(0, \frac{\sigma^2}{2a})$, then X_T has the same distribution for any $T > t$. *Hint*, show that this satisfies the forward equation. This is a slightly different way of saying that $\mathcal{N}(0, \frac{\sigma^2}{2a})$ is the statistical steady state distribution.

   (d) Use your formulas for the mean and variance of X_t to find the value function for an OU process (19) and final time payout $V(x) = x^2$ at time T. Show that your formula satisfies the backward equation for OU and appropriate final condition. Interpret the fact that the value function becomes increasingly flat as $T - t$ becomes large. Find a formula for this "flat" value using the steady state probability distribution and check that it agrees with the rest of this exercise.

2. The geometric Brownian motion SDE 25 may be solved using the *log variable transformation.* There are several equivalent ways to derive the transformation.

(a) Set $X_t = \log(S_t)$. Use Ito's lemma to find the SDE that X_t satisfies. Show that $X_t = a + bt + cW_t$ is a solution. The process X_t is *Brownian motion with drift.* This is the derivation given above, explained slightly differently.

(b) Write the backward equation for S_t. Consider the log change of variables $f(s,t) = g(\log(s),t)$. Show that this g satisfies

$$\partial_t g + \alpha \partial_x g + \beta \partial_x^2 g = 0 \ .$$

Find the relation between α and β here to a, b, and c from part (a). What diffusion process has this PDF as its backward equation?

3. Assume that the higher moments of the diffusion process are of the size they would be for Brownian motion, which is

$$\begin{aligned} \mathrm{E}\Big[\,|\Delta X|^3\Big] &\le C\Delta t^{\frac{3}{2}} \\ \mathrm{E}\Big[\,|\Delta X|^4\Big] &\le C\Delta t^2 \end{aligned}$$

Suppose $f(x,t)$ has partial derivatives up to order 4 in both variables. Define $\Delta f = f(X_{t+\Delta t}, t+\Delta t) - f(x,t)$ Show that

$$\mathrm{E}[\,\Delta f \mid X_t = x] = \left[\partial_t f(x,t) + a(x)\partial_x f(x,t) + \frac{1}{2}b(x)^2\partial_x^2 f(x,t)\right]\Delta t + o(\Delta t) \ .$$

Show that

$$\mathrm{E}\big[\,(\Delta f)^2 \mid X_t = x\big] = \left[b(x)\partial_x f(x,t)\right]^2 \Delta t + o(\Delta t) \ .$$

This is a *weak* version of Ito's lemma.

4. Consider the Ornstein Uhlenbeck process (19) and running payout

$$\int_t^T X_s^2 \, ds \ .$$

Evaluate the value function explicitly using variance formulas for the OU process. Verify that this function satisfies the backward equation (18).

5. The code `StockSim.py` does the Euler Maruyama method to compute a geometric Brownian motion governed by the SDE (26).

(a) Run with a larger T and Δt and see that it is common to produce negative approximate prices. For this, you do not need so many paths and the plots will not look good. This is OK because the results are not good either.

(b) Find a formula for $u(s,t)$, which is the PDF of S_T. You can use the solution formula (27) for this. Add the exact PDF to the plot and see what Δt you need to get a good match.

(c) The problem gets harder for larger T. Do a calculation with larger T to see that the PDF of S_T is not at all symmetric.

(d) A *volatility surface* model makes σ a function of s. A volatility *skew* adds a slope, which is $\sigma(s) = \sigma_0 + \sigma_1(s - S_0)$. A volatility *smile* adds a positive quadratic term. Experiment with volatility skew, both positive and negative and see how this impacts the PDF. Add two curves to the plot, one with positive and one with negative skew. Choose the slopes s_1 so that the PDF is noticeably different but not completely different. At this point, your plot will have four curves. Make sure the vol surface curve legend labels have the corresponding skew values. Volatility skew and smile are used to explain the observed fact that market option prices for put options that are unlikely to be "in the money" are much higher than the Black Scholes theory says they should be.