Stochastic Calculus

Stochastic Differential Equarions

Jonathan Goodman, Fall, 2022

1 General introduction to the course

These are the notes for a course called *Stochastic Calculus* offered at the Courant Institute in the Fall semester of 2022. It is a "half course", meaning that it consists of seven lectures of just under two hours. Most of the students are in the MS program *Mathematics in Finance* but the course was not intended for them only. I assumpe that the student have a tolerance for examples related to finance, and examples not related to finance.

The course is for a person interested in stochastic calculus as a practical modeling and analysis tool, not as a technical branch of probability theory. It describes creating and interpreting *SDE* (*stochastic differential equation*) models of random processes, and how to learn predictions of such models using simulation and the connections between SDE and *PDEs* (partial differential equations). There is little of the measure theory and no rigorous proofs of any of the main theorems. It is more like the "Freshman calculus" approach to stochastic calculus than the "mathematical analysis" approach.

The course is an experimental deviation from earlier versions that build up to diffusion processes after a discussion of Brownian motion. This course begins with and focused on SDE models, with Brownian motion mainly used as the simplest example. A more common approach is to describe Brownian motion in detail and then describe other diffusion processes as functions of Brownian motion. The "function of Brownian motion" approach has some mathematical technical advantages, but it can get in the way of modeling and computing in practical applications. If the class hates this approach, next year's class will revert to the old ways.

The course is at the level of beginning graduates or advanced undergraduates with the right technical background but not necessarily mathematics majors. Indispensable prerequisites are multi-variate calculus, linear algebra, and a probability background that includes multi-variate probability densities including the multi-variate normal and the central limit theorem. The assignments require students to use Python and the standard packages for numerical computing, Numpy, Scipy, and Matplotlib.

2 Introduction to the subject

Stochastic Calculus is a collection of methods for creating and analyzing models

of dynamical systems driven, in part, by noise. Stochastic calculus is an extension of ordinary calculus, which is used to create ordinary differential equations that model systems without noise. There are some major differences between stochastic modeling with *stochastic differential equations* (*SDE*) and modeling noise free dynamics with *ordinary* differential equations (*ODE*). One difference is that "the" solution of the SDE is random. Two simulations starting at the same place give different results. A theoretical description of the solution of an SDE involves probability density, expected values, mean and variance, and so on. By contrast, the solution for an ODE with a given starting point is a single trajectory. There are many more subtle differences that make SDE and stochastic calculus challenging to master.

Notation and terminology

A dynamical model involves a *state* at time t with n components. An ODE model might write this as $x(t) = (x_1(t), \ldots, x_n(t))$. SDE modelers typically write $X_t = (X_{1,t}, \ldots, X_{n,t})$. The time variable t is a subscript rather than a function argument. The state is a capital letter, indicating that its value is random. An ODE is written using derivatives, as

$$\frac{d}{dt}x(t) = f(x(t)) \; . \tag{1}$$

An SDE is written using differentials, as

$$dX_t = a(X_t)\, dt + b(X_t)\, dW_t \; . \tag{2}$$

The ODE could be written as $dx(t) = f(x(t))\, dt$, but many people prefer the derivative form (1) because the derivative has a mathematical definition while the differentials do not (in this context). The SDE expresses a model of stochastic dynamics in which the expected value of dX_t is $a(X_t)\, dt$ and the covariance matrix of dX_t is determined by $b(X_t)$. In notation to be clarified later, these have the form of conditional mean and covariance:

$$\mathrm{E}[\, dX_t \mid \cdots\,] = a(X_t)\, dt \; , \;\; \mathrm{cov}(\, dX_t \mid \cdots\,) = bb^t(X_t)\, dt \; . \tag{3}$$

The coefficient vector a is the *infinitesimal* mean, and the matrix $\mu = bb^t$ is the infinitesimal covariance. These are often called *drift* and *quadratic variation* respectively. The final factor dW_t is the *increment* of *Brownian motion*. This may be viewed literally as the differential of a random function of time called Brownian motion, or it may be viewed just as a way to say: independent of everything in the past, mean zero, covariance $I\, dt$.

This course uses $p(x,t)$ to denote the probability density of the random variable X_t. Figure 2 has estimates of $p(x,t)$ for two t values. This is for the *Ornstein Uhlenbeck* process, which described by the SDE

$$dX_t = -aX_t dt + \sigma dW_t \; . \tag{4}$$

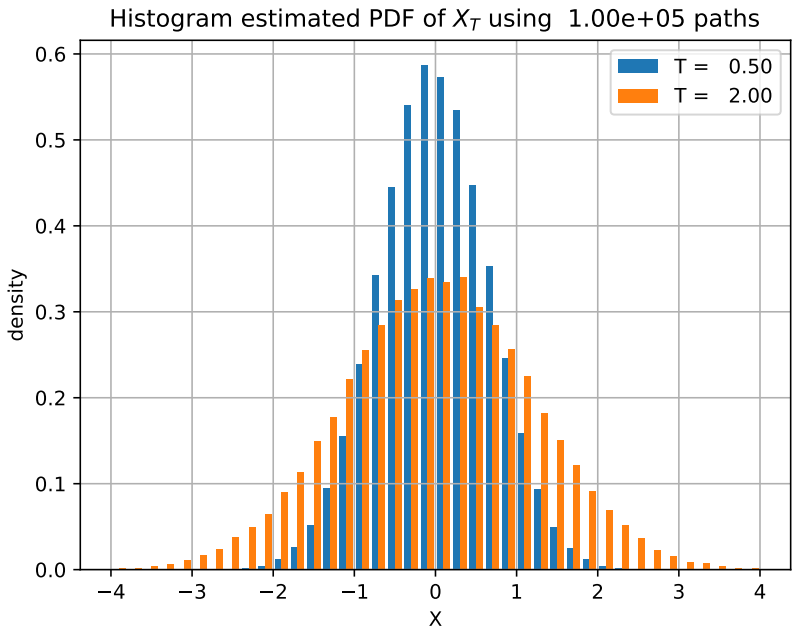

Figure 1: Probability density of X_t estimated for two t values using a histogram of many simulated paths.

All paths start with $X_0 = 0$. The PDF for $t = .5$ is more narrow than the PDF for $t = 2$ because paths have had less time to disperse. Figure 2 plots many paths and shows that they are generally closer to $x = 0$ at the earlier time. The narrower $t = .5$ density is also higher because the densities have the same integral.

Rough paths

In probability theory, a function of a time variable may be called a *path*. An SDE describes a random path X_t. If you simulate the SDE process many times, you get many different paths. The "solution" of an SDE is not a single path, but a probability distribution describing how likely different sets of paths can be. A typical path is not a "smooth", or even differentiable function of t. This explains the fact that some of the rules of ordinary calculus do not apply to SDE paths. Figure 2 plots a number of independently simulated paths to illustrate their behavior. Each path position X_t seems to be a continuous function of t, but they never seems to have a well defined slope.

To be more precise, let $\Delta X = X_{t+\Delta t} - X_t$ be the increment in the path corresponding to a small Δt. The quantitative analysis of Sections 4 and 5 will show that the size of a typical increment is $|\Delta X| \sim \sqrt{\Delta t}$. This $\sqrt{\Delta t}$ comes up

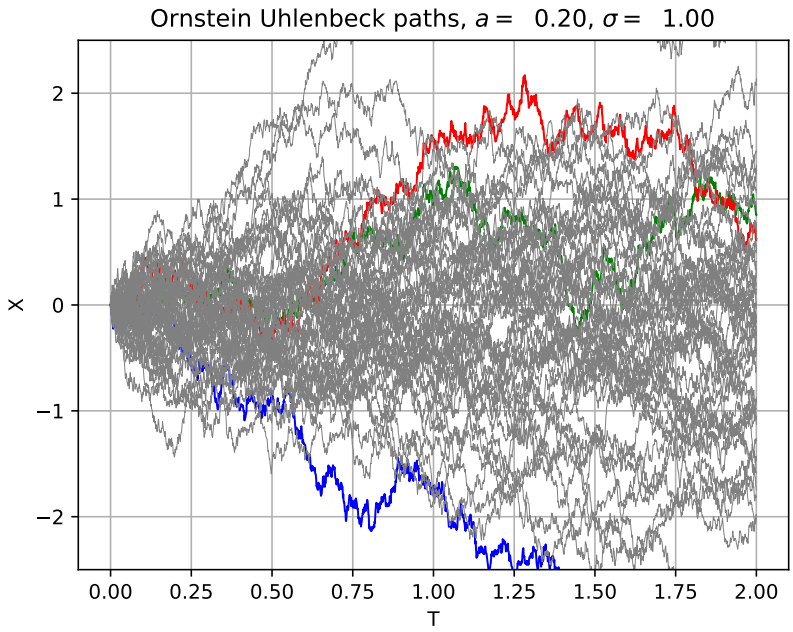

Figure 2: A collection of random paths generated by the Ornstein Uhlenbeck SDE (4).. Three paths are in bold color and the rest are thin gray. The paths take different random values but they share qualitative properties such as roughness.

constantly in stochastic calculus. This implies that the average velocity over this interval of time satisfies

$$\left|\overline{V}\right| = \frac{|\Delta X|}{\Delta t} \sim \frac{1}{\sqrt{\Delta t}} \; .$$

This goes to infinity as $\Delta t \to 0$. The path X_t is "rough" (not differentiable or smooth) because its velocity is always (almost always) infinite, either positive or negative infinity.

Cancellation

Cancellation in a sum or an integral means that the positives nearly cancel the negatives, so the result is much smaller than if you just multiply the number of terms by the size of a typical term (for a sum) or multiply the typical size of the integrand by the length of the interval of integration (for an integral). An example is the indefinite integral

$$F_t = \int_0^t \sin(s)\, ds \; .$$

The typical size of the integrand is maybe $|\sin(s)| \sim \frac{1}{2}$. If you multiply this typical size by the length of the integration interval, you get $\frac{1}{2}t$. This becomes large when t is large. However, the actual value of the integral is $F_t = \cos(t)$, which is never bigger than one. When t is large, the positive and negative parts of the integrand roughly cancel.

Cancellation happens in probability when you take a sum or an integral of random variables with mean zero (or close to zero). The increment dX_t is such a quantity. We will see that $\mathrm{E}[dX_t]$ is on the order of dt but $\mathrm{E}[|dX_t|]$ is on the order of $\sqrt{dt}$.

To say this in a slightly different way, suppose $t_k = k\Delta t$ is a sequence of times with a small separation Δt. Let $\Delta X_k = X_{t_{k+1}} - X_{t_k}$ be the corresponding increments of X_t. Now take the limit $\Delta t \to 0$ with $t_n = n\Delta t = T$ fixed. It is always true, for any such Δt, that

$$X_T - X_0 = \sum_{k=1}^{n-1} \Delta X_k \; . \tag{5}$$

However, it is also true, almost surely (see below for the definition of "almost surely") that

$$\sum_{k=1}^{n-1} |\Delta X_k| \; \to \; \infty \; .$$

This resolves the paradox that the velocity is infinite but the random process goes only finitely far. Cancellation comes from the direction constantly changing.

Ito calculus and $(dX)^2$

The infinite velocity and large increments of sample paths lead to a breakdown of the ordinary rules of calculus. In studying solutions of and ODE (1), a typical chain rule calculation would be

$$\frac{d}{dt} U(x(t)) = \nabla U(x(t)) \, \dot{x}(t) = \nabla U(x(t)) \, f(x(t)) \; .$$

For example, this can be integrated to give

$$U(x(T)) - U(x(0)) = \int_0^T \nabla U(x(t)) \, f(x(t)) \, dt \; .$$

With a small Δt, the integral is approximated by a sum, so

$$U(x(T)) - U(x(0)) \approx \sum_{k=0}^{n-1} \nabla U(x(t_k)) \, f(x(t_k)) \, \Delta t \; . \tag{6}$$

In Freshman calculus you (should) learn that the sum on the left converges to the integral in the limit $\Delta t \to 0$.

The stochastic calculus version of this, which is the *Ito calculus*, relates to the old-fashioned "differential" version of the chain rule, which is

$$dU(x(t)) = \nabla U(x(t))\, dx \ .$$

The total change integral (6) is

$$U(x(T)) - U(x(0)) = \int_0^T \nabla U(x(t))\, dx(t) \ . \tag{7}$$

[A "modern" calculus course might tell you that "dx" doesn't "make sense", so that dx must be replaced by $\frac{dx}{dt}dt$.] The integral on the left has a finite Δt approximation, using the notation of (5) and $x_k = x(t_k)$.

$$U(x(T)) - U(x(0)) \approx \sum_{k=0}^{n-1} \nabla U(x(t_k))\, \Delta x_k \ . \tag{8}$$

The right side sum converges to the integral in (7) and therefore to the left side here, in the limit $\Delta t \to 0$.

This is not true in stochastic calculus if X_t is the solution of an SDE model (2). The ordinary calculus approximation $dU \approx \nabla U\, dx$, which is the basis of (8) is not good enough. Instead, you need the *Ito's lemma* approximation, which is based on the second order Taylor approximation of dU:

$$dU \approx \nabla U\, dX + \frac{1}{2}(dX)^T \left(D^2U\right) dX \ . \tag{9}$$

The notation is that D^2U is the $n \times n$ *hessian* matrix of second partial derivatives

$$\left(D^2U\right)_{jk} = \frac{\partial^2 U}{\partial x_j \partial x_k} \ .$$

Because of cancellation (this subtle point is discussed later in the course), it suffices to use the expected values of the second terms on the right, but not the first term. Written out, the second term is

$$\frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \left(D^2U\right)_{jk} dX_{j,t} dX_{k,t} \ .$$

The expected value of one of the products on the right is

$$\mathrm{E}[\, dX_{j,t} dX_{k,t}\,] = \mathrm{cov}(\, dX_{j,t}\,,\, dX_{k,t}\,) + \mathrm{E}[\, dX_{j,t}\,]\, \mathrm{E}[\, dX_{k,t}\,] \ .$$

With all of this, the stochastic calculus version of (8) is

$$U(X_T) - U(X_0) \approx \sum_{i=1}^{N} \nabla U(X_{t_i}) dX_{t_i} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k=1}^{n} \left(D^2U\right)_{jk} \mathrm{cov}(\, dX_{j,t}\,,\, dX_{k,t}\,) \ . \tag{10}$$

By the end of the course, you should understand why we keep these terms and drop others. The two reasons are cancellation and smallness of neglected terms.

No PDF in path space

The probability distribution of the random path is not given by a probability density. The term *path space* is used for the collection of all functions of time that a random path might be. One specific path space is the set of continuous functions, saying that X_t is a continuous function of t, defined for t in the interval $0 \leq t \leq T$ and starting with a definite value such as $X_0 = 0$. This path space is a vector space in the sense of linear algebra, but it is *infinite dimensional* in the sense that it does not have a finite basis. By contrast, suppose Y is a random object defined by n components $Y = (Y_1, \cdots, Y_n)$. Then the distribution of Y might be defined by a probability density (PDF, for "probability density function") in the sense that for A set in $y-$space

$$\Pr(\, Y \in A) = \int_A p(y)\, dy_1 \cdots dy_n \ . \tag{11}$$

A general path in an infinite dimensional path space is not described by a finite list of parameters and general probabilities are not described by integration over a finite number of parameters as in (11).

Stochastic calculus looks for ways to find probability densities for some quantities, such as the n component random variable X_T for a fixed T. In finite dimensional probability, these *marginal* probability distributions would be found by "integrating out" the rest of the variables. This is impossible in path space because there are infinitely many values $t \neq T$. And, there is no PDF for the path. Simulation is an important tool in really challenging practical problems. The main theoretical tool is partial differential equations. These have names such as *forward* equation, *backward equation*, *Feynman Kac* equation (a misnomer used by many practitioners), etc.

Almost sure properties

A statement about a random object Y is true *almost surely* if the probability that it is true is one. In formulas, if A is a set of objects, then

$$\text{"}Y \in A \ \text{ almost surely"}, \quad \text{or "}Y \in A \ \text{ a.s."}, \quad \text{means } \Pr(\, Y \in A) = 1 \ .$$

Even if $Y \in A$ almost surely, there may be possible outcomes y with $y \notin A$. For example, if Y is a one component random variable with a PDF $p(y)$, then $y = 0$ is a possible outcome – an element of the probability space – but $\Pr(Y = 0) = 0$. This may be expressed by: $Y \neq 0$, a.s.

Interesting almost sure properties usually involve limits. A simple example is the *strong* law of large numbers, which concerns the average of N independent *samples* of a PDF $p(y)$. The hypothesis is that the absolute value has finite expectation:

$$\mathrm{E}[\, |Y| \,] = \int |y| \ dy < \infty \ .$$

The theorem is that if random variables Y_k are independent and have the same PDF, then, almost surely, the sample means converge to the actual mean:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} Y_k = \mathrm{E}[Y] \, .$$

Limits are the theoretical foundation of ordinary calculus. For example, if $f(x)$ is continuous then the Riemann sum approximation converges to the integral in the limit $\Delta x \to 0$. For the stochastic integrals of stochastic calculus, these limits are correct almost surely.

3 Introduction to the Section

This section is about creating and simulating SDE models of the form (2).

3.1 Multi-variate normal

The formulas and linear algebra of multi-variate normals and covariances is used constantly in stochastic calculus, particularly for multi-variate processes. Here is a quick summary. Let $p(y)$ be the PDF of a multi-component random variable $Y = (Y_1, \cdots, Y_n)$. Then Y is a multi-variate normal if there is a vector $\mu = (\mu_1, \cdots, \mu_n)$ and a symmetric positive definite $n \times n$ matrix Σ so that

$$p(y) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \, e^{-\frac{1}{2}(y-\mu)^t \Sigma^{-1} (y-\mu)} \, . \tag{12}$$

This formula has the properties that

$$\begin{aligned}
\int_{\mathbb{R}^n} p(y)\, dy &= 1 \\
\mathrm{E}[Y] &= \int_{\mathbb{R}^n} y p(y)\, dy \\
&= \mu \\
\mathrm{cov}(Y) &= \mathrm{E}\big[\, (Y-\mu)(Y-\mu)^t \big] \\
&= \int_{\mathbb{R}^n} (y-\mu)(y-\mu)^t p(y)\, dy \\
&= \Sigma \, .
\end{aligned}$$

That is to say that (12) defines the probability density of a random variable with mean μ and covariance matrix Σ.

A linear mapping applied to a multi-variate normal gives another multi-variate normal. If Y is multi-variate normal and $X = AY + b$, then X is multi-variate normal.

If A is an $m \times n$ matrix and $b \in \mathbb{R}^m$ is an $m-$component "offset", and if $X = AY + b$, then X is multivariate normal. The mean and covariance of X are given by (in hopefully clear notation)

$$\mu_X = A\mu_Y + b \tag{13}$$
$$\Sigma_X = A\Sigma_Y A^t \ . \tag{14}$$

The second formula follows from the calculation, which uses facts of matrix multiplication,

$$\begin{aligned}
\Sigma_X &= \mathrm{E}\big[\,(X-\mu_X)(X-\mu_X)^t\big] \\
&= \mathrm{E}\Big[\,[A(Y-\mu_Y)]\,[A(Y-\mu_Y)]^t\Big] \\
&= \mathrm{E}\big[\,A(Y-\mu_Y)(Y-\mu_Y)^t A^t\,\big] \\
&= A\,\mathrm{E}[\,(Y-\mu_Y)(Y-\mu_Y)]^t\,A^t \\
&= A\Sigma_Y A^t \ .
\end{aligned}$$

The probability density formula for the PDF of X involves Σ_X^{-1}, so it makes strict sense only if Σ_X is non-singular. This means that $m \leq n$, so X cannot have more components than Y. If $m = n$, then A must be non-singular. If $m < n$, then A must have maximal rank, which is m.

4 SDE models, infinitesimal mean and variance

We describe SDE models by analogy to ODE models. We give similarities and differences in some technical notation.

An ODE model involves a state "variable", which is a vector with d components

$$x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_d(t) \end{pmatrix} \ .$$

We usually treat vectors as column vectors when we do matrix/vector algebra. But we might write $x = (x_1, \cdots, x_d)$ instead of the more correct $x = (x_1, \cdots, x_d)^T$ if we are just listing components of x.

ODE models assume that $x(t)$ is a differentiable function of t. This means that if Δt is small, then $\Delta x = x(t + \Delta t) - x(t)$ is approximately proportional to Δt. The constant of proportionality may be written as

$$\dot{x} = \frac{d}{dt}x = \begin{pmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_d(t) \end{pmatrix} \ .$$

"Approximately proportional" means that[1]

$$\Delta x = \dot{x}\,\Delta t + O(\Delta t^2)\ .$$

ODE models give $\dot{x}(t)$ in terms of the state $x(t)$. For component k, this is $\dot{x}_k(t) = f_k(x(t))$. The model specifies that if Δt is small, then the change in component x_k over that time interval is

$$\Delta x_k \approx f_k(x_1(t), \cdots, x_d(t))\,\Delta t\ . \tag{15}$$

These functions f_k are the ODE model. The hardest part of modeling a system is deciding what these f_k should be.

An ODE model depends on the hypothesis that Δx depends on $x(t)$ and Δt, but not on $x(s)$ for $s < t$. The functions $f_k(x)$ give this dependence. The update formula (15) has only components of $x(t)$ on the right side, not $x(s)$ for $s \neq t$.

Here is an example to emphasize this point. Suppose $C(t)$ is the concentration of a chemical in air. There is a chemical reaction that consumes the chemical at a rate proportional to the concentration and the temperature, $T(t)$

$$\frac{dC}{dt} = -k\,T(t)\,C(t)\ .$$

The chemical reaction increases the temperature in a way that is proportional to a "heat release" coefficient r and the reaction rate $kC(t)$

$$T(t) = T(0) + k\,r\int_0^t C(s)\,ds\ .$$

This gives a model

$$\frac{dC}{dt} = -k\left(T_0 + kr\int_0^t C(s)\,ds\right)C(t)\ . \tag{16}$$

This is an equation involving derivatives, but it is not an ODE model because the right side depends on values $C(s)$ for $s < t$. To put this into the form of an ODE model, define a two component state variable $x(t) = (C(t), T(t))$. The dynamics are given by

$$\frac{dC}{dt} = -kT(t)C(t) \tag{17}$$

$$\frac{dT}{dt} = krC(t)\ . \tag{18}$$

It takes a two component *state space* to make this model into an ODE model. We can put this model into generic notation using all the notation

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} C \\ T \end{pmatrix}$$

$$\dot{x} = f(x) = \begin{pmatrix} f_1(C,T) \\ f_2(C,T) \end{pmatrix} = \begin{pmatrix} -kTC \\ krT \end{pmatrix} = \begin{pmatrix} -k\,x_2\,x_1 \\ k\,r\,x_1 \end{pmatrix}\ .$$

[1]There is a summary of "big Oh" notation, as in $O(\Delta t^2)$ at the end. A good calculus book should describe it.

An SDE model specifies the *infinitesimal mean* and the *infinitesimal covariance* at a time t, as a function of X_t, which is the state at time t. The change in time interval $[t, t + \Delta t]$ is

$$\Delta X = X_{t+\Delta t} - X_t \ .$$

The expected value of ΔX is approximately proportional to Δt:

$$\mathrm{E}[\,\Delta X] = a(X_t)\Delta t + O(\Delta t^2) \ .$$

It is traditional to use $a(x)$ for the infinitesimal mean, also called *drift*, instead of the $f(x)$ that is used in ODE models. The infinitesimal mean/drift has d components.

The infinitesimal mean formulas just above need to be made more precise. In an SDE model, you are modeling the future evolution of a path given that you already know the whole past from the "beginning of time" (time $t = 0$ by tradition) up to time t. This path may be called $X_{[0,t]}$. What we model is the the statistics of the future $X_{t+\Delta t}$, with $\Delta t > 0$, conditioned on knowing the path up to time t. This is analogous to the difference between the dynamics of $C(t)$ in (16), which is not independent of $C([0, t])$, and the ODE model of C and T together in (17) and (18). The latter depend on the path $C([0, t])$ and $T([0, t])$ only through the present values $C(t)$ and $T(t)$. In the stochastic model, the probability distribution of the future $X_{t+\Delta t}$ depends on the past $X_{[0,t]}$ only through the "present" value X_t. A more correct version of the infinitesimal mean formula uses conditional probability notation, as

$$\mathrm{E}\big[\,\Delta X \mid X_{[0,t]}\big] = a(X_t)\Delta t + O(\Delta t^2) \ . \tag{19}$$

It is common to use the notation of differentials to signal that the $O(\Delta t^2)$ terms will be ignored. In that way of thinking, we write dt for a small positive increment of time. We write $dX_t = X_{t+dt} - X_t$ for the corresponding increment of X. The infinitesimal mean formula, in this notation, is just

$$\mathrm{E}\big[\,dX_t \mid X_{[0,t]}\big] = a(X_t)dt \ . \tag{20}$$

SDE modeling differs from ODE modeling in that an SDE model determines the probability distribution of $X_{t+\Delta t}$, conditioned on the known X_t, but the SDE model does not determine the value of $X_{t+\Delta t}$ as an ODE model would. Therefore, an SDE model also must include a measure of the conditional uncertainty of $X_{t+\Delta t}$. This is the *infinitesimal variance* or *infinitesimal covariance*.

$$\mathrm{cov}\big[\,\Delta X_t \mid X_{0,t}\big] = C(X_t)\Delta t + O(\Delta t^2) \ . \tag{21}$$

The infinitesimal covariance matrix C should be called the "covariance rate" because the actual covariance is this matrix multiplied by Δt. Mathematicians often call C, or the integral of C, the *quadratic variation*. It turns out that the probability distribution of paths $X_{[0,T]}$ is determined by the infinitesimal mean/drift and infinitesimal (co)variance/quadratic variation. To be clear, a

multivariate Gaussian is determined by its mean and covariance matrix. But the distribution of SDE paths does not have to be Gaussian. The distribution of ΔX is approximately Gaussian if Δt is small, which is enough.

A *stochastic differential equation* description of the dynamics has the form (2). The first term on the right is the infinitesimal mean (19). The term $b(X_t)dW_t$ is the $d\times d$ *noise coefficient* matrix and dW_t is the *increment of Brownian motion*. It is a model of uniform yet totally unpredictable noise. "Totally unpredictable" means that dW_t is independent of $W_{[0,t]}$, and $\mathrm{E}[\, dW_t] = 0$ always. We don't have to write $\mathrm{E}\big[\, dW_t \mid W_{[0,t]}\big] = 0$ because dW_t is assumed to be independent of $W_{[0,t]}$. "Uniform" means that the statistics of dW_t are the same at every time. This means that the infinitesimal (co)variance of Brownian motion C_B is a constant independent of t and (as just stated) $W_{[0,t]}$. By convention, we take $C_B = I$, so

$$\mathrm{cov}(\, dW_t) = I\, dt\ . \tag{22}$$

In components, this is

$$\mathrm{var}(\, dW_{k,t}) = dt\ , \tag{23}$$

and

$$\mathrm{cov}(\, dW_{j,t}, dW_{k,t}) = 0\ ,\ \text{if } j \neq k\ . \tag{24}$$

A d component Brownian motion process is written

$$W_t = \begin{pmatrix} W_{1,t} \\ \vdots \\ W_{d,t} \end{pmatrix}\ .$$

The components of Brownian motion are independent and have the same distribution.

5 Ito's lemma, the stochastic chain rule

6 Simulating diffusion processes

Simulating a diffusion process governed by an SDE is a lot like simulating an ODE trajectory. You replace dt by a positive but small Δt and *march* forward in time in a sequence of *time steps*. By convention, the starting time will be $t = 0$ and the time step times are $t_n = n\Delta t$. The time between them is $t_{n+1} - t_n = \Delta t$. A superscript Δt will indicate that we are making the finite Δt approximation. The notation may be confusing, but an approximation to the ODE dynamics (1) is

$$\Delta x^{\Delta t}(t_n) = f(x^{\Delta t}(t_n)\Delta t\ .$$

This may be written as

$$x^{\Delta t}(t_{n+1}) = x^{\Delta t}(t_n) + f(x^{\Delta t}(t_n))\Delta t\ . \tag{25}$$

This is called the *forward Euler* method, or just *Euler's method*.

We simplify notation for SDE paths by writing $X_n^{\Delta t}$ instead of $X_{t_n}^{\Delta t}$. The analogous time step for the SDE (2) is

$$\Delta X_n^{\Delta t} = a(X_n^{\Delta t}\Delta t + b(X_n^{\Delta t})\,\Delta W_n \ .$$

The $a(\cdot)\Delta t$ is the same as for the ODE. The ΔW_n are random and independent noise variables that have

$$\mathrm{E}[\,\Delta W_n] = 0 \ , \ \ \mathrm{cov}(\Delta W_n) = I\Delta t \ . \tag{26}$$

It is convenient (but not necessary!) to satisfy the statistical conditions (26) by taking

$$\Delta W_n = \sqrt{\Delta t} Z_n \ , \ \ Z_n \sim \mathcal{N}(0, I) \ .$$

If you do this, the algorithm to simulate an SDE path is first to take X_0 either to be the known starting value or to have the specified starting probability distribution. The rest of the path is found using the above, in the form

$$X_{n+1}^{\Delta t} = X_n^{\Delta t} + a(X_n^{\Delta t})\Delta t + \sqrt{\Delta t}\, b(X_n^{\Delta t})\, Z_n \ , \ \ Z_n \sim \mathcal{N}(0, I) \ , \ \text{independent} \ . \tag{27}$$

This is *Euler's method* for the SDE, also called the *Euler Maruyama* method.