Stochastic Calculus, Courant Institute, Fall 2019
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2019/StochCalc.html
Jonathan Goodman, September, 2019

Class 5, multi-component diffusions

1 Introduction

A stochastic dynamic model can have more than one component. Financial models used for planning investments, to take one example, model the simultaneous price changes of multiple assets. Engineering models need many variables to model the many moving parts of the device.

A multi-component stochastic process with d components (d is for dimension) is a random function of time, written $X_t \in \mathbb{R}^d$. We write $X_t = (X_{1,t}, \ldots, X_{d,t})$ when we need to talk about the individual components. If we have to do linear algebra, we may treat X_t as a column vector

$$X_t = \begin{pmatrix} X_{1,t} \\ \vdots \\ X_{j,t} \\ \vdots \\ X_{d,t} \end{pmatrix} .$$

The component $X_{j,t}$ might represent price of asset j at time t, or it might represent the distance between car j and car $j+1$ at time t.

Much of what this course has said about one component diffusions applies in a straightforward way to multi-component diffusions, but there are some new issues that do not arise in one component problems. We use $p(x,t)$ to denote the PDF of X_t. This looks the same as before, but now there are d variables $x = (x_1, \ldots, x_d)$. The infinitesimal mean is as before, except that in the expression both sides have d components. Choose $\Delta t > 0$ and define the change to be $\Delta X = X_{t+\Delta t} - X_t$. The infinitesimal mean is defined by

$$\mathrm{E}[\, \Delta X \mid X_t = x] = a(x)\Delta t + o(\Delta t) .$$

The "infinitesimal variance" is now a $d \times d$ matrix we call the infinitesimal covariance

$$\mathrm{E}\big[\, \Delta X \Delta X^t \mid X_t = x\big] = \mu(x)\Delta t + o(\Delta t) .$$

The expected value of a vector or matrix function in these formulas is done "componentwise". Note that if ΔX is a column vector then $\Delta X \Delta X^t$ is a $d \times d$ matrix, so its expected value, which is approximately $\mu(x)\Delta t$, is a $d \times d$ matrix.

A d component diffusion process satisfies an SDE of the form

$$dX_t = a(X_t)dt + b(X_t)dW_t \; . \tag{1}$$

The drift coefficient (the infinitesimal mean) $a(x)$ is a d component column vector. The Brownian motion is an m component *multi-dimensional* Brownian motion. The dimension of the driving Brownian motion is the number of *sources of noise*. To be stochastic, there needs to be at least one source, so $m \geq 1$. Many SDE models have $m = d$, which usually gives a *non-degenarate diffusion*. Other SDE models have $m < d$, which leads to a *degenerate diffusion*. A model with $m > d$ can always be replaced in a simple way by a model with $m = d$. There is no reason to study diffusions with more sources of noise than components. The *noise coefficient* $b(x)$ is a $d \times m$ matrix. It is related to the infinitesimal covariance (we will see) by

$$\mu(x) = b(x)b^t(x) \ . \tag{2}$$

For single component diffusions, this is the $\mu = b^2$ formula we had before. For d components, the right side is a $d \times m$ matrix multiplying an $m \times d$ matrix, which gives a $d \times d$ matrix. μ is singular (degenerate) if $m < d$.

The generator of the diffusion process is a partial differential operator. If $g(x) = g(x_1, \ldots, x_d)$, then

$$Lg(x) = \sum_{j=1}^{d} a_j(x)\partial_{x_j} g(x) + \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} \mu_{jk}(x)\partial_{x_j}\partial_{x_k} g(x) \ . \tag{3}$$

This turns into the generator formula we had before when $d = 1$. You will come to think of this as the natural version for $d \geq 1$. The value function is for simple expected values is

$$f(x,t) = \mathrm{E}_{x,t}[\,v(X_T)] \ . \tag{4}$$

The backward equation for this value function is

$$\partial_t f(x,t) + Lf(x,t) = 0 \ . \tag{5}$$

The backward equation has this form for every Markov process, unless the Markov process is so weird that it doesn't have a generator at all. Each Markov process has its own generator. The generator for a multi-component diffusion process has the form (3).

There are two general ways to get information about multi-component diffusions, numerical solution of backward equations and numerical simulation. The backward equation (or any PDE) requires a grid. Grids in high dimensions are impractical, so numerical solution of the backward equation is only practical for problems with a small number of components. More than four is likely to be impractical unless you have a very big computer. More than six wi probably impractical even with a big computer. Numerical simulation is practical for more complicated systems. As we have seen already, Monte Carlo results are noisy. It takes many sample paths to get an accurate average. Stochastic calculus can help design better simulation strategies. In finance, the simplest thing of a given type is often called "vanilla". [Vanilla is the simplest flavor of ice cream.] Fancy simulation methods can give results that are more accurate than those from vanilla methods. We will see in this course examples of two fancier simulation strategies, *control variates* and *importance sampling*. Importance sampling for SDE problems often uses the *Girsanov* re-weighting theorem, which comes later in the course.

Let $p(\cdot, t)$ be the PDF of X_t. This is the same as the one component case discussed in earlier classes, except that now there are d coordinates $x = (x_1, \ldots, x_d)$ as arguments to p. A *steady state* of the SDE (1) is a function $p(x)$ so that if $X_s \sim p(\cdot)$ then $X_{s_t} \sim p(\cdot)$ also. It is a theorem that unless the SDE is degenerate in some obvious way (more degenerate than just being a degenerate diffusion), and if there is a steady state, then

$$p(\cdot, t) \to p(\cdot) \ \text{ as } \ t \to \infty \ .$$

If there is a steady state, then the probability density converges to that steady state as $t \to \infty$. While the PDF has a limit, the path X_t continues to have random fluctuations. A statistical steady state is a steady state for the PDF, not for X_t. A steady state is sometimes called an *equilibrium* for the SDE. [Physicists use the term *equilibrium* for a steady state only if the SDE is "reversable", whatever that means.] A diffusion process in finance is called an *equilibrium model* if the SDE has a steady state. Stock price models are not equilibrium models, because the PDF for a stock keeps spreading as $t \to \infty$. But models of interest rates can be equilibrium models if there is a long term PDF of interest rates. The long term mean may be 3% or so with fluctuations around this value not decreasing nor growing as $t \to \infty$.

Linear multi-component diffusions can behave in many different ways. Importantly, it is possible to understand properties of linear diffusion processes in more detail than is possible for non-linear diffusions. In particular, the auto-covariance structure (covariance between X_{t_1} and X_{t_2}) may be calculated using the matrices describing the diffusion. If you observe a time series that you want to model, and if that time series is not Markov in itself, you might think of modeling it as one component of a linear multi-component diffusion. You might look for a diffusion that produces the auto-covariance structure of your time series.

2 Sources of noise

A "standard" $m-$component Brownian motion is

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

Each component process $W_{j,t}$ is a standard Brownian motion ($\mathrm{E}[\Delta W_j] = 0$, and $\mathrm{E}[\Delta W_j^2] = \Delta t$) as described in classes 1 - 4. The components are independent. This means that the time t covariance matrix is (with I being the $m \times m$ identity matrix)

$$\mathrm{cov}(\Delta W) = \mathrm{E}\big[\, \Delta W \Delta W^t \big] = \Delta t I \ .$$

You can define $m-$component Brownian motion directly. It's a random process with $X_t \in \mathbb{R}^m$ with the independent increments property and $\Delta W \sim \mathcal{N}(0, \Delta t I)$. This is the same thing. If one-component random variables are each one dimensional Gaussians, and if they are independent, then they form a multi-component Gaussian.

A multi-component diffusion, X_t, that is *driven* by an $m-$component standard Brownian motion is said to have *m sources of noise*. Each independent component of W_t is an independent source of noise. Being "driven by" means that X_t satisfies the SDE (1) with an *$n \times m$ noise coefficient matrix* $b(x)$. A "driving force", or "external force" (or "exogenous input", if you're an economist) in a differential equation is a term $F(t)$ so that $\dot{x} = f(x) + F(T)$. In differential form, it would be $dx = f(x)dt + F(t)dt$. For diffusions, this takes the form of $b(x)dW_t$.

Independent noise sources in (1) create correlated noise for the components dX_j. You can think of a two component diffusion as

$$\begin{aligned} dX_1 &= a_1(X)dt + noise_1 \\ dX_2 &= a_2(X)dt + noise_2 \ . \end{aligned}$$

As an example, think of $X_{1,t}$ and $X_{2,t}$ as the prices of stock 1 and stock 2. The noise that drives stock 1 is not independent of the noise that drives stock 2. If stock 1 goes up, there is an increased likelihood that stock 2 goes up. The noise coefficient matrix $b(x)$ in the SDE (1) creates these correlations. In the 2 component case, they are

$$\begin{aligned} dX_1 &= a_1(X)dt + b_{11}(X)dW_{1,t} + b_{12}(X)dW_{2,t} \\ dX_2 &= a_2(X)dt + b_{21}(X)dW_{1,t} + b_{22}(X)dW_{2,t} \ . \end{aligned}$$

Comparing these expressions, we see

$$\begin{aligned} noise_1 &= b_{11}(X)dW_{1,t} + b_{12}(X)dW_{2,t} \\ noise_1 &= b_{21}(X)dW_{1,t} + b_{22}(X)dW_{2,t} \ . \end{aligned}$$

If $b_{11} \neq 0$ and $B_{21} \neq 0$, then $W_{1,t}$ "drives" both components. This creates a correlation between $noise_1$ and $noise_2$. When you're building a diffusion process model for something, you may know the noise coefficient matrix and calculate the infinitesimal covariance, or you may know the infinitesimal covariance and invent a b that creates it. In the second case, modeling using the SDE (1) may seem a little artificial.

There is a "uniqueness theorem" that says that if two diffusion processes have the same infinitesimal mean $a(x)$, and the same infinitesimal (co)variance $\mu(x)$, then they are the same process (more precise statement in a future class). [Something is "unique" if there is only one of that something. An equation has a unique solution if it has only one solution. The equation $x^2 = 2$ has two solutions. The equation $x^3 = 2$ has a unique solution (if you allow only real numbers x). An SDE has a unique solution (unless it's a bad SDE). The uniqueness theorem for diffusion processes is about probability distributions in path space, which we talk about in later classes.] But there are different $b(x)$ noise coefficient matrices that give the same infinitesimal covariance $\mu(x) = b(x)b^t(x)$. In fact if $b(x)$ and $\widetilde{b}(x) = b(x)q(x)$, with $qq^t = I$, then b and $\widetilde{b}$ lead to the same infinitesimal covariance matrix. For non-degenerate diffusions, if $bb^t = \widetilde{b}\widetilde{b}^t$, then there is a q so that $\widetilde{b} = bq$.

3 Linear systems

A linear system is an SDE of the form

$$dX_t = AX_t dt + BdW_t \ . \tag{6}$$

If the initial data X_0 is Gaussian, then X_t is Gaussian for later times (we will see this). Linear processes like this have many uses. For one thing, you may just be modeling a linear system or one that is approximately linear. Otherwise, systems like this are convenient ways to make noise (random processes) with interesting correlation structure. When you're trying to model random time series that you don't understand well (fluctuations in prices, wind patterns, etc.), you may just want to "build" a linear model that has a similar "correlation structure".

A linear diffusion may have a steady state or not. A *steady state* probability density is $u(x)$ so that if $X_t \sim u$ then $X_{t+s} \sim u$ for $s > 0$. The time dependent PDF for any diffusion satisfies a forward equation (a PDE described in a future class). If $X_t \sim u(\cdot, t)$, then $u(\cdot, t)$ satisfies that forward equation. It may happen that there is a $u(x)$ so that $u(x,t) \to u(x)$ as $t \to \infty$. This $u(x)$ is the steady state. Much of this section is about linear diffusions that have a steady state.

For any time series, the *lag t auto-covariance* function (sometimes called the *matrix* auto-covariance function) is the lag t covariance matrix

$$C(t) = \lim_{s\to\infty} \mathrm{cov}(X_s, X_{s+t}) \ . \tag{7}$$

The limit exists (we will see in a moment) if and only if there is a steady state. It may happen that you do not "observe" the whole state $X_t \in \mathbb{R}^n$, but only one component (say, $X_{1,t}$) or a "linear functional" $Y_t = v^t X_t$, where $v \in \mathbb{R}^n$ represents a linear combination of the components $X_{j,t}$. The time series you observe may be just Y_t instead of the whole X_t. This time series has "correlation structure" given by the *scalar* auto-covariance function

$$C_v(t) = \lim_{s\to\infty} \mathrm{cov}(Y_s, Y_{s+t}) \tag{8}$$

If you have a Y_t with a complicated auto-covariance structure, you can try to build matrices A and B so that the resulting scalar auto-covariance function matches what you see.

Much of what we need to know about solutions and auto-covariance functions for the SDE (6) is determined by the ODE

$$dx = Ax\,dt \ , \quad \text{equivalently} \ \ \dot{x} = Ax \ . \tag{9}$$

Much of the information we need about this is determined by the eigenvalues of A. Here is a quick review of this stuff. Many students will be familiar with this material, though possibly with different notation. The solution of the linear ODE system (9) may be expressed in terms of a matrix function $S(t)$ called the *fundamental solution*. This is a matrix that satisfies

$$\frac{d}{dt}S(t) = AS(t) \;, \quad S(0) = I \;.$$

The solution is

$$x_{t+s} = S(t)x_s \;.$$

The fundamental solution S_t advances the solution by a time t.

The fundamental solution may be constructed in terms of the eigenvalues and eigenvectors of A. The precise formula will be next week, but it involves a lot of numbers related to eigenvectors, and the functions $e^{\lambda_j t}$, where the λ_j are the eigenvalues of S. The eigenvalues generally are complex, which we write $\lambda_j = \mu_j + i\omega_j$, with μ_j and ω_j being real. These are the *real part* and *imaginary part* of λ_j, and written $\mu_j = \text{Re}(\lambda_j)$ and $\omega_j = \text{Im}(\lambda_j)$. The exponential of a complex number may be written as

$$e^{\lambda_j t} = e^{\mu_j t}e^{i\omega_j t} = e^{\mu_j t}\left(\cos(\omega_j t) + i\sin(\omega_j t)\right) \;.$$

The exponential $e^{\lambda_j t}$ is an exponentially growing if $\mu_j = \text{Re}(\lambda_j) > 0$ and exponentially decaying if $\text{Re}(\lambda_j) < 0$. You can plot the point λ_j in 2d by giving it coordinates (μ_j, ω_j). The points with $\mu = 0$ correspond to the "y axis", which is the *imaginary axis*. A point with $\mu < 0$ is in the *left half plane*. A point with $\mu > 0$ is in the *right half plane*. An eigenvalue λ_j in the left half plane corresponds to *stable* dynamics in that $e^{\lambda + _j t}$ converges to zero exponentially. An eigenvalue λ_j in the right half plane corresponds to *strongly unstable* dynamics in that $e^{\lambda + _j t}$ grows exponentially. The matrix A is *stable* if all its eigenvalues are in the left half plane. The matrix A is unstable if it has any eigenvalues in the right half plane. An eigenvalue on the imaginary axis is called *neutrally stable* because the exponential neither grown nor decays. For the linear SDE (6), there is good reason (explained next week) to classify the system as unstable. If A is a stable matrix, then $S(t) \to 0$ exponentially as $t \to \infty$.

The ODE

$$\dot{x}_t = Ax_t + F_t \tag{10}$$

is called *homogeneous*. The English word "homogeneous" means "all the same". A group of rocks is homogeneous if they all have the same color, size, and shape. The rocks are inhomogeneous if they have different sizes or shapes, etc. If the "forcing function" $F(t)$ depends on t, in that it is different for different values of t, then the equation (10) is "inhomogeneous". As often happens, the mathematical meaning is a bit different from the meaning in ordinary language. In economics, the function $F(t)$ might be called *exogenous*, which means "specified from the outside". In an economic model, a function is exogenous if it is determined by something outside the model and *endogenous* if it is determined by the model itself. You might say that $x(t)$ is determined endogenously by (10) from the exogenous F.

The solution of the inhomogeneous ODE system (10) satisfies

$$x_t = S(t)x_0 + \int_0^t S(t-t')F_{t'}\, dt' \ . \tag{11}$$

This formula is called *Duhamel's principle*. Suppose A is stable. Then x_t "forgets" the initial data x_0 because the contribution $S(t)x_0$ converges to zero exponentially. In the integral, the forcing with small t' (time close to zero) also carries exponentially small weight. The forcing carries more weight when $t-t'$ is not large, which is for t' close to t. That is, x_t "knows most" about forcing that happened "recently" (t' close to t) and has little memory of forcing that happened long ago.

You can verify the Duhamel formula (11) by differentiating the right side with respect to t. The integral depends on t in two ways, through the upper limit of integration and through $S(t-t')$. With some algebra, you see that the x given by the Duhamel formula (11) satisfies the inhomogeneous ODE (10). But the formula is natural without this mathematical verification. The ODE is linear, so the sum of solutions is a solution. In the integral part of (11), the contribution $S(t-t')F_{t'}dt'$ is the influence of the force between times t' and $t'+dt'$ on x_t. The delay between t' and t is $t-t'$, which is how long the influence at time t' has been acting. That explains the $S(t-t')$. The ODE is linear, so the total solution is given as a sum (an integral) of all the individual influences at times $t' < t$.

The linear SDE (6) has exogenous forcing W_t. If we put this into Dummel's formula, the result is

$$X_t = S(t)X_0 + \int_0^t S(t-t')dW_{t'} \ . \tag{12}$$

Now the right side is an Ito integral involving the Brownian motion W_t. We can verify this formula using Ito's lemma, but its intuition should make it believable for now.

The point is: If A is a stable matrix if and only if the linear SDE 6 is an equilibrium model. An equilibrium model must "forget" the initial distribution (because the limit distribution is the same no matter what the initial distribution was). If $X_0 = x_0$ is deterministic, then because $S(t) \to 0$ exponentially, the influence of x_0 on X_t disappears as $t \to \infty$.