Lesson 8, Multi-component diffusion

1   Theory (quick summary)

A multi-dimensional (or multi-component) diffusion is a collection $(X_{1,t}, \ldots, X_{d,t}) = X_t$. The stochastic dynamics of the components $X_{k,t}$ are related, but together the multi-component process $X_t \in \mathbb{R}^d$ is a Markov process. This means what it meant for single component diffusions: the distribution of X_{t_0+s} conditional on knowing X_t for all $t \leq t_0$ is the same as the distribution of X_{t_0+s} conditional on just knowing X_{t_0}. A multi-component diffusion is characterized by its infinitesimal mean and infinitesimal covariance.

The infinitesimal mean (also called drift) is called $a(x)$ and is defined by

$$\mathrm{E}_{x,t}[\Delta X] = \Delta t\, a(x) + O(\Delta t^2) \ . \tag{1}$$ `eq:im`

The notation is what we used before. We assume that $\Delta t > 0$ and take the limit as $\Delta t \to 0$. The subscript in $\mathrm{E}_{x,t}[\cdot]$ means that the expectation is taken assuming that $X_t = x \in \mathbb{R}^d$. The increment is $\Delta X = X_{t+\Delta t} - X_t$.

The infinitesimal covariance (related to quadratic variation) is a $d \times d$ positive semi-definite matrix $\mu(x)$ so that

$$\mathrm{E}_{x,t}\big[\Delta X \Delta X^t\big] = \Delta t\, \mu(x) + O(\Delta t^2) \ . \tag{2}$$ `eq:ic`

The (j, k) entry in $\mu(x)$ is the infinitesimal covariance between ΔX_j and ΔX_k:

$$\mathrm{cov}_{x,t}(\Delta X_j, \Delta X_k) = \mu_{jk}(x)\, \Delta t + O(\Delta t^2) \ .$$

The covariance is defined by subtracting the mean. This is irrelevant here (as it was for one dimensional diffusions) because the difference is $O(\Delta t^2)$. That is

$$\mathrm{E}_{x,t}[\Delta X_j \Delta X_k] = \mathrm{E}_{x,t}[(\Delta X_j - a_j(x)\Delta t)\,(\Delta X_k - a_k(x)\Delta t)] + O(\Delta t^2) \ .$$

The left side is the plain expectation as in the definition (2) `eq:ic` of infinitesimal covariance. The expectation on the right is almost the time Δt covariance. We subtracted out the approximation $a_j(x)\Delta t$ instead of the exact $\mathrm{E}_{x,t}[\Delta X_j]$.

There is a multi-dimensional Ito calculus and Ito's lemma for multii-component diffusions. The formula is a direct generalization of the one component Ito's lemma.

$$df(X_t, t) = \sum_{j=1}^{d} \partial_{x_j} f(X_t, t)dX_j + \partial_t f(X_t, t)dt + \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} \partial_{x_j}\partial_{x_k} f(X_t, t)\mu_{jk}(X_t)dt \ . \tag{3}$$ `eq:Il`

As for the one component formula, the derivation here has two steps. First you justify expanding f to first order in t and to second order in the x variables. Then you justify replacing $\Delta X_j \Delta X_k$ by its expectation value (to order Δt), which gives $\mu_{jk}(X_t)dt$. The simplest backward equation is

$$\partial_t f + \sum_{j=1}^{d} a_j(x)\partial_{x_j} f + \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}\mu_{jk}(x)\partial_{x_j}\partial_{x_k} f = 0 . \tag{4}$$ eq:be

This equation defines the generator of the diffusion process. The generator is a differential operator (now involving partial derivatives) in the "space variables" x_j.

$$Lg(x) = \sum_{j=1}^{d} a_j(x)\partial_{x_j} g(x) + \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d}\mu_{jk}(x)\partial_{x_j}\partial_{x_k} g(x) . \tag{5}$$ eq:g

The backward equation is

$$\partial_t f + Lf = 0 .$$

If f satisfies the backward equation, then $Y_t = f(X_t, t)$ is a martingale. This implies that if $T > t$, then

$$E_{x,t}[f(X_T, T)] = f(x, t) .$$

In particular, if we solve the backward equation (4) with final condition $f(x, T) = V(x)$, then f is the value function

$$f(x, t) = E_{x,t}[V(X_T)] . \tag{6}$$ eq:vf

You can find the value function by solving a partial differential equation, or you can express the solution of the partial differential equation as the value function for a diffusion process.

The probability density for a multi-component diffusion will be called $u(x, t)$. It is defined by $X_t \sim u(\cdot, t)$. We can derive a PDE for u, the forward equation, using a duality argument like the one we used for one component diffusions. The tower property (the rules of conditional probability) the unconditional expectation of $V(X_T)$ as an integral, which is the expected value of the conditional expectations at time $t < T$. The first line below is an informal version of the second line:

$$E[V(X_T) = \int_{\mathbb{R}^d} E[V(X_T) \mid X_t = x] \Pr(X_t = x)$$
$$= \int_{\mathbb{R}^d} E_{x,t}[V(X_T)]\, u(x, t)\, dx$$
$$= \int_{\mathbb{R}^d} f(x, t)\, u(x, t)\, dx$$
$$= \langle f(\cdot, t), u(\cdot, t)\rangle .$$

The last line is the inner product "in space" (the x variables are the "space variables" and integration dx is integration "in space".) of $f(\cdot, t)$ with $u(\cdot, t)$. It

is defined by the integral above it. It is independent of t, so the derivative with respect to t is zero.

We can repeat the reasoning we used for one component diffusions to derive the forward equation:

$$
\begin{aligned}
0 &= \frac{d}{dt} \langle f(\cdot, t), u(\cdot, t) \rangle \\
&= \langle \partial_t f(\cdot, t), u(\cdot, t) \rangle + \langle f(\cdot, t), \partial_t u(\cdot, t) \rangle \\
&= -\langle Lf(\cdot, t), u(\cdot, t) \rangle + \langle f(\cdot, t), \partial_t u(\cdot, t) \rangle \\
&= -\langle f(\cdot, t), L^* u(\cdot, t) \rangle + \langle f(\cdot, t), \partial_t u(\cdot, t) \rangle \\
0 &= \langle f(\cdot, t), [\partial_t u(\cdot, t) - L^* u(\cdot, t)] \rangle \; .
\end{aligned}
\tag{7}
$$

eq:wfe

In this formula L^* is the adjoint of L, which is defined using integration by parts (as for one component diffusions). If $g(x)$ and $v(x)$ are functions of $x \in \mathbb{R}^d$ that behave OK for large x (this is a case-by case check), then L^* is defined by the duality relation

$$
\langle Lg, v \rangle = \langle g, L^* v \rangle \; .
$$

Some calculations show that the adjoint of the generator is

$$
L^* v(x) = -\sum_{j=1}^{d} \partial_{x_j} \left[a_j(x) v(x) \right] + \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} \partial_{x_j} \partial_{x_k} \left[\mu_{jk}(x) v(x) \right] \; .
$$

The line (7) holds for "every" function f, so the quantity in square brackets $[\cdot]$ on the right must be zero. This may be written out either abstractly in terms of the adjoint of the generator or concretely in terms of partial derivatives:

$$
\partial_t u(\cdot, t) = L^* u(\cdot, t)
\tag{8}
$$

eq:fea

$$
\partial_t u(x, t) + \sum_{j=1}^{d} \partial_{x_j} \left[a_j(x) u(x, t) \right] = \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{d} \partial_{x_j} \partial_{x_k} \left[\mu_{jk}(x) v(x) \right] \; .
\tag{9}
$$

eq:fec

If we have initial conditions such as $u(x, 0)$, then the forward equation determines $u(x, t)$ for $t > 0$.

The forward equation describes the movement of probability density over time. For diffusions, this movement may be described using a probability flux (or probability current) $\mathcal{F}(x, t)$. The probability flux has d components $\mathcal{F}_j(x, t)$. If $\mathcal{S}$ is a $d - 1$ dimensional surface in $\mathbb{R}^d$, then $\mathcal{F}$ describes the rate at which probability is crossing the surface. The rate is

$$
\int_{\mathcal{S}} \mathcal{F}(x, t) \cdot n(x, t) \, dA(x) \; .
$$

This is a $d - 1$ dimensional surface integral. At a point $x \in \mathcal{S}$, the unit normal vector to the surface is $n(x)$. The $d - 1$ dimensional unit of area on $\mathcal{S}$ is dA. If $\mathcal{V} \subset \mathbb{R}^d$ is a region with boundary $\mathcal{S}$, then

$$
\frac{d}{dt} \int_{\mathcal{V}} u(x, t) \, dx = \int_{\mathcal{S}} \mathcal{F}(x, t) \cdot n(x, t) \, dA(x) \; .
$$

3

The divergence theorem implies that this integral relation is equivalent to the differential relation

$$\partial_t u(x,t) + \text{div}\mathcal{F}(x,t) = 0 \ . \tag{10}$$

eq:uc

The divergence is

$$\text{div}\mathcal{F}(x,t) = \sum_{j=0}^{d} \partial_{x_j} \mathcal{F}_j(x,t) \ . \tag{11}$$

eq:dd

The probability flux is

$$\mathcal{F}_j(x,t) = -a_j(x)u(x,t) - \frac{1}{2}\sum_{k=1}^{d} \partial_{x_k}\left[\mu_{jk}(x)u(x,t)\right] \ . \tag{12}$$

eq:fu

The mathematical reasoning puts these formulas in the opposite order. First you define the flux using (12). This is just the definition of $\mathcal{F}(x,t)$. Then you calculate the divergence of this $\mathcal{F}$ using the definition of divergence (11). You see that the *conservation form* of the forward equation (10) is equivalent to the explicit concrete form (9).

The probability flux gives a physical picture of the forward equation. There is a probability current (flux, flow) $\mathcal{F}(x,t)$. At any point x at time t, the probability is flowing in the direction of $\mathcal{F}(x,t)$ with flow rate proportional to $|\mathcal{F}|$. If $\mathcal{F}$ is tangent to a surface $\mathcal{S}$ then probability flows along $\mathcal{S}$, but there is no net flow of probability from one side of $\mathcal{S}$ to the other. If you picture a collection of independent diffusions all at (x,t), then their random motions take them from one side of $\mathcal{S}$ to the other and back. But most of this random motion cancels out, as nearly the same number of particles cross from each side to the other. The flux represents average motion after this random motion is averaged out.

Probability flux is useful in modeling diffusions with boundaries. If it is an absorbing boundary, the flux tells you how fast particles are being absorbed. If it is a reflecting boundary, the boundary condition is $\mathcal{F}(x,t) \cdot n(x) = 0$ at a boundary point x. In one dimension, this was simply $\mathcal{F}(x,t) = 0$ because $n(x) = \pm 1$ in one dimension. Geometry of boundaries and surfaces is more complicated in multi-dimensional settings. It would be wrong to set $\mathcal{F}(x,t) = 0$ at boundary points in multi-dimensions because there are d components of the flux, this would be d boundary conditions, and you are supposed to give just one boundary condition. You can use the flux to model a situation where a particle may be absorbed at a boundary or may not be. In this case (as we saw in an earlier assignment), there is a relation between u and $\mathcal{F}$ at the boundary.

You can specify a multi-component diffusion by giving a stochastic differential equation. This would be a vector SDE

$$dX_t = a(X_t)\,dt + b(X_t)\,dW_t \ . \tag{13}$$

eq:sde

It could be written in components as (for $j = 1, \ldots, d$)

$$dX_{j,t} = a_j(X_t)dt + \sum_{k=1}^{m} b_{jk}(X_t)dW_{k,t} \ . \tag{14}$$

eq:sdec

4

We assume that the $W_{k,t}$ are independent standard Brownian motions. Any correlations in the noise are produced by the noise matrix b. The number m in the sum (14) is the number of sources of noise. It is common that $m < d$, which means that there are fewer sources of noise than components. This is a *degenerate diffusion*. It is also common to have $m = d$, which defines a *nondegenerate diffusion* if the $d \times d$ noise matrix is non-singular at every x. It is rare to have $m > d$, and it is always unnecessary (as we will see). The drift coefficient a in the SDE (13) is the same as the infinitesimal mean (1).

The noise matrix b is related to the infinitesimal covariance as follows (an informal derivation)

$$
\begin{aligned}
\mu(x)dt &= \mathrm{E}_{x,t}\big[dX dX^t \big] \\
&= \mathrm{E}_{x,t}\Big[(b(x)dW_t)\, (b(x)dW_t)^t \Big] + O(dt^2) \\
&= \mathrm{E}_{x,t}\big[(b(x)dW_t)\, \big(dW_t^t b(x)^t\big) \big] + O(dt^2) \\
&= b(x)\mathrm{E}_{x,t}\Big[(dW_t)\,(dW_t)^t \Big] b(x)^t + O(dt^2) \\
&= b(x)\, I dt\, b(x)^t + O(dt^2) \\
&= b(x)b(x)^t + O(dt^2) \; .
\end{aligned}
$$

In this derivation, the covariance of Brownian motion is $\mathrm{E}_{x,t}\Big[(dW_t)\,(dW_t)^t \Big] = I dt$, where I is the $m \times m$ identity matrix and dt is the time increment. We cancel the dt on both sides and arrive at

$$
\mu(x) = b(x)b^t(x) \; . \tag{15}
$$

2 Linear Gaussian processes

A system of linear equations takes the form

$$
\frac{d}{dt}x(t) = \dot{x}(t) = Ax \; .
$$

We use a dot to indicate the time derivative of a differentiable function. The state variable is $x(t) \in \mathbb{R}^d$ and the dynamics are specified by the $d \times d$ matrix A. You add noise to a linear system by adding a noise term

$$
dX_t = AX_t dt + B dW_t \; . \tag{16}
$$

The Ornstein Uhlenbeck process is the only one dimensional linear sytem. But in multi-dimensions there are many more possibilities and kinds of behavior. Nevertheless, for any matrices A and B, the process X_t is Gaussian. If X_0 is deterministic or if it has a Gaussian distribution, then X_t is Gaussian, as well as any linear function of the path.

The mean and covariance of X_t satisfy differential equations that lead to formulas, as they did for Ornstein Uhlenbeck. Let $m(t) = \mathrm{E}[X_t]$ be the mean.

The dynamics of m may be found as

$$
\begin{aligned}
m(t+dt) &= \mathrm{E}[X_{t+dt}] \\
&= \mathrm{E}[X_t + dX_t] \\
&= m(t) + \mathrm{E}[AX_t dt] + \mathrm{E}[BdW_t] \ .
\end{aligned}
$$

But $\mathrm{E}[dW_t] = 0$, so we get

$$
m(t+dt) = m(t) + Am(t)dt \ .
$$

This may be written in the usual way as a differential equation system without noise

$$
\dot{m} = Am \ .
$$

This is a characteristic property of linear stochastic systems – the mean evolves according to the same dynamics (16), but without noise.

The dynamics of the covariance matrix can be found in a similar way. The time t covariance matrix is

$$
C(t) = \mathrm{cov}(X_t) = \mathrm{E}\big[(X_t - m(t))(X_t - m(t))^t\big] \ .
$$

Here is an informal version of the calculation. A more formal version uses Ito's lemma (3). In a small increment of time dt we calculate dC. At first we expand the quantity in the expectation to second order in dX_t. Then we evaluate the quadratic terms (terms that involve two copies of dX_t by first ignoring the drift term $AX_t dt$ and then taking the expectation. The dynamics of $m(t)$ are known, and $dm(t) = \dot{m}(t)dt$. In the terms that are linear in dX, the expected value of BdW is zero, so we leave that out. Constant matrices may be taken out of the expectation, but you have to keep the order. For Brownian motion, $\mathrm{E}\big[(dW_t)(dW_t)^t\big] = I dt$.

$$
\begin{aligned}
dC_t &= \mathrm{E}\big[(dX_t - \dot{m}\,dt)(X_t - m(t))^t + (X_t - m(t))(dX_t - \dot{m}\,dt)^t + (dX_t)(dX_t)^t\big] \\
&= \mathrm{E}\big[(AX_t dt - Am\,dt)(X_t - m)^t\big] + \mathrm{E}\big[(X_t - m)(AX_t dt - Am\,dt)^t\big] \\
&\quad + \mathrm{E}\big[(BdW_t)(BdW_t)^t\big] \\
&= \mathrm{E}\big[A(X_t - m)(X_t - m)^t dt\big] + \mathrm{E}\big[(X_t - m)(X_t - m)^t A^t dt\big] \\
&\quad + \mathrm{E}\big[B(dW_t)(dW_t)^t B^t\big] \\
&= A\,\mathrm{E}\big[(X_t - m)(X_t - m)^t\big]\,dt + \mathrm{E}\big[(X_t - m)(X_t - m)^t\big]\,A^t dt \\
&\quad + B\,\mathrm{E}\big[(dW_t)(dW_t)^t\big]\,B^t \\
&= AC dt + C A^t dt + BB^t dt \ .
\end{aligned}
$$

The result is

$$
\dot{C} = AC + CA^t + BB^t \ . \tag{17}
$$

Some diffusion processes have limiting distributions $u(\cdot, t) \to u_*(x)$ as $t \to \infty$. The covariance dynamics equation (17) determines whether a linear Gaussian process is stable of unstable. If $C(t)$ has a limit as $t \to \infty$ then $C_* =$

6

$\lim_{t\to\infty} C(t)$ is the covariance matrix of $u_*(x)$. Furthermore, $u_*(x)$ is Gaussian, so it is determined by C_*.

If there is a steady state, the covariance satisfies the *Lyapunov equation*

$$AC_* + C_*A^t = -BB^t \ . \tag{18}$$

For a one component system, the Ornstein Uhlenbeck system $dX = -\gamma X_t dt + \sigma dW_t$, we have $A = -\gamma$ (a 1×1 matrix) and $B = \sigma$. The general equation (18) for the steady state variance c_* reduces to

$$-2\gamma c_* = -\sigma^2 \ . \tag{19}$$

There is a positive steady state variance, c_*, if and only if $\gamma > 0$ and $\sigma \neq 0$. If $\gamma < 0$ and $\sigma \neq 0$, then the equation steady variance equation (19) has a solution, but this c_* is negative and is not the variance of anything. The variance of X_t clearly goes to infinity as $t \to \infty$ if $\gamma > 0$ and there is no steady state.

The Lyapunov equation (18) is a collection of linear equations for the entries of the matrix C_*. There is one equation for each matrix index (i, j). Written in components, they become

$$\sum_{k=1}^{k} A_{ik}C_{*,kj} + \sum_{k=1}^{k} C_{*,ik}A_{kj} = -\sum_{k=1}^{d} B_{ik}B_{jk} \ .$$

The unknowns are the d^2 matrix elements $C_{*,ij}$ and there are d^2 equations, one for each (i, j). Well, this isn't exactly right because C_* is symmetric. But still the number of equations is the same as the number of unknowns, because the (i, j) equation is the same as the (j, i) equation and the (i, j) unknown is equal to the (j, i) unknown. As for Ornstein Uhlenbeck, the solution C_* may not be positive definite and therefore may not have physical significance.

The situation for the general system is similar to the situation for Ornstein Uhlenbeck. The ODE system $\dot{x} = Ax$ is stable if all the eigenvalues of A have negative real part (are in the left half of the complex plane). In that case, the equation system (18) has a solution C_* that is positive definite. Also, in that case $C(t) \to C_*$ as $t \to \infty$. The proof is not hard, but it takes some time to go through.

3   Correlated stocks

Suppose there are n *risky asset prices* S_k that satisfy

$$dS_k = \mu_k S_k dt + \sigma_k S_k dW_k \ . \tag{20}$$

If the driving noise processes W_k are independent, then the prices S_k are independent too. But asset prices are not independent. The dependence between risky assets can be modeled within the geometric Brownian motion framework

by assuming that the Brownian motions W_k are correlated. In fact, you can specify the correlations

$$\rho_{jk} = \text{corr}(dW_j, dW_k) \,. \tag{21}$$

The correlation matrix ρ with entries ρ_{jk} can be any positive definite matrix with ones on the diagonal, $\rho_{kk} = 1$. Giving the correlations instead of the covariances allows each S_k to be a geometric Brownian motion with expected return μ_k and volatility σ_k. The infinitesimal covariance of the asset prices is (to order dt^2, which is not written)

$$\text{cov}(dS_j, dS_k) = \text{E}[\, dS_j dS_k]$$
$$= \sigma_j \sigma_k S_j S_k \rho_{jk} dt \,.$$

The diagonal covariances are what they should be for geometric Brownian motion

$$\text{E}\big[\, dS_k^2\big] = \sigma_k^2 S_k^2 dt; \,.$$

Any positive definite matrix can be a covariance matrix. Any positive definite matrix with ones on the diagonal can be a correlation matrix. In fact, if Y has covariance matrix ρ then $\text{var}(Y_j) = \rho_{jj} = 1$ so ρ is also the correlation matrix of Y. It is often convenient (as here) to specify the correlations and the variances separately. If random variables have correlation matrix ρ_{jk} and if $X_j = c_j Y_j$, then the correlations are the same

$$X_j = c_j Y_j \implies \text{corr}(X_j, X_k) = \text{corr}(Y_j, Y_k) \,.$$

You can understand this as the motivation behind the definition of correlation – the relationship between variables that doesn't care about scalings.

If ρ is a positive definite matrix with $\rho_j = 1$ (a correlation matrix), then we can take the Cholesky factorization $LL^* = \rho$. If $Z_{k,t}$ are independent standard Brownian motions and if $W_t = LZ_t$, then W_t are Brownian motions with the desired correlations. Each $W_{k,t}$ is a standard Brownian motion, but the different components have the desired correlations.

A *basket* of stocks is a portfolio that has a_k "shares" (or *units*) of asset with price S_k. The total value of the basket is

$$R_t = \sum_{k=1}^n a_k S_{k,t} \,.$$

A straightforward but long calculation shows that R_t is not a geometric Brownian motion even when the S_k are geometric Brownian motions, correlated or not. This means that a stock index (the S&P 500 for example) is not a geometric Brownian motion if the individual stocks are.

The generator for a system of correlated geometric Brownian motions is

$$Lg = \frac{1}{2} \sum_k \sum_k \sigma_j \sigma_k s_j s_k \rho_{jk} \partial_{s_j} \partial_{s_k} g + \sum_j \mu_j s_j \partial_{s_j} g = 0 \,.$$

8

The backward equation for a payout that depends on correlated geometric Brownian motions is $\partial_t f + Lf = 0$. Written explicitly, this is

$$\partial_t f + \frac{1}{2} \sum_k \sum_k \sigma_j \sigma_k s_j s_k \rho_{jk} \partial_{s_j} \partial_{s_k} f + \sum_j \mu_j s_j \partial_{s_j} f = 0 \ .$$