Stochastic Calculus

Backward Equation, duality

Jonathan Goodman, Fall, 2022

1 Introduction

2 Backward equation and generator

Suppose X_t is a diffusion process described by the SDE

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

Consider this to be a multi-component process with d components. This means that the SDE may be written in component form as

$$dX_{j,t} = a_j(X_t)\, dt + \sum_{k=1}^{m} b_{jk}(X_t)\, dW_{k,t} \ \ , \quad \text{for } j = 1, \cdots, d \ . \tag{2}$$

The $W_{k,t}$ are independent standard Brownian motions. The coefficient b_{jk} is the influence of Brownian motion W_k on component X_j of the process. These form the entries of the noise matrix b in the vector form (1). The number m is the number of *sources of noise*. We have seen that a non-diagonal b can be used to model correlations in the components $dX_{j,t}$. In fact, a model might start with an idea of single component infinitesimal variance and cross-component correlation:

$$\sigma_{j,t}^2\, dt = \text{var}(dX_{j,t}) \ , \ \ \rho_{jk,t} = \text{corr}(dX_{j,t}, dX_{k,t}) \ .$$

The noise coefficient matrix b then is constructed to match these modeling specifications.

A modeling example with correlations

It often happens that $m = d$, so b is a square matrix. If b is non-singular, we say the diffusion is *non-degenerate*. Degenerate diffusion models with $m < d$ also are common. For example, suppose r_t is the "short rate", which is the interest rate paid for loans that are repaid very soon (the next day, often) and have no risk of not being repaid. A simple model of the fluctuations in the short rate might be

$$dr_t = -\gamma(\overline{r} - r_t)\, dt + \sigma dW_t \ . \tag{3}$$

This is an *equilibrium* model (because there is a steady state probability distribution) built on the idea that there is a natural short rate $\overline{r}$ that the fluctuating rate returns to with rate γ, while being taken away from $\overline{r}$ by a constant infinitesimal variance term σdW_t. People used to criticize this model because it is possible to have $r_t < 0$. In the mean time, several European countries have at some time experienced periods of a negative short rate. A *money market* account is an amount of money whose value changes only because of interest at the short rate. This is modeled by

$$dM_t = r_t M_t \, dt \; . \tag{4}$$

Together, the components r_t and M_t form a two component degenerate diffusion process that may be written in matrix/vector form as

$$d \begin{pmatrix} r_t \\ M_t \end{pmatrix} = \begin{pmatrix} -\gamma(\overline{r} - r_t) \\ r_t \end{pmatrix} dt + \begin{pmatrix} \sigma \\ 0 \end{pmatrix} dW_t \; .$$

The matrix b is 2×1, with a single source of noise for a two component model.

This model could be enriched by adding a risky asset S_t that satisfies

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t \; . \tag{5}$$

Of course, the σ in the r_t equation (short rate volatility) is different from the σ in the risky asset equation (stock volatility). Moreover, we need to model correlations between interest rate and stock price changes. A natural to do this adapting the "one factor" market model that Markowitz used to model prices of different stocks. In that model, there is a *market factor*, which will be called $dZ_{0,t}$, and there are *idiosyncratic* factors for each stochastic process, which will be called $dZ_{r,t}$ and $dZ_{S,t}$. These three factors are taken to be independent. Correlations between market dS and dr are created by "weighting" factors differently.

A "market factor" is a source of noise that effects all the financial instruments in a market in some way. Here, there is just one market factor (a "one factor" model), $dZ_{0,t}$. The individual instrument prices have weightings β "to the market", so that the σdW_t in the r equation has a contribution $\beta_r dZ_{0,t}$. The σdW in the S equation has a contribution $\beta S dZ_{0,t}$. The β factors are different but the noise $dZ_{0,t}$ is the same. This term is responsible for correlations between dr and dS. The instruments also have independent idiosyncratic factors, σ_r and σ_S. These are sources of uncertainty that affect one instrument and not the other. Idiosyncratic factors are responsible for the less than 100% correlation between the noisy instrument values.

The SDEs that incorporate these modeling ideas are

$$dr_t = -\gamma(\overline{r} - r_t) \, dt + \beta_r \, dZ_{0,t} + \sigma_r \, dZ_{r,t} \; , \tag{6}$$

and

$$dS_t = \mu S_t \, dt + S_t \left(\beta_S \, dZ_{0,t} + \sigma_S \, dZ_{S,t} \right) \; . \tag{7}$$

Using the independence of the three noise terms $dZ_{0.t}$, $dZ_{r,t}$, and $dZ_{S,t}$, we find

$$\mathrm{var}(dr_t) = \left(\beta_r^2 + \sigma_r^2\right) dt \tag{8}$$

$$\mathrm{var}(dS_t) = S_t^2\left(\beta_S^2 + \sigma_S^2\right) dt \tag{9}$$

$$\mathrm{corr}(dr_t, dS_t) = \frac{\beta_r \beta_S}{\sqrt{\beta_r^2 + \sigma_r^2}\sqrt{\beta_S^2 + \sigma_S^2}} \tag{10}$$

$$d\begin{pmatrix} r_t \\ M_t \\ S_t \end{pmatrix} = \begin{pmatrix} -\gamma(\overline{r} - r_t) \\ r_t \\ \mu S_t \end{pmatrix} dt + \begin{pmatrix} \beta_r & \sigma_r & 0 \\ 0 & 0 & 0 \\ S_t\beta_S & 0 & S_t\sigma_S \end{pmatrix} \begin{pmatrix} dZ_{0,t} \\ dZ_{r,t} \\ dZ_{S,t} \end{pmatrix} . \tag{11}$$

The first row of this matrix equation is equivalent to the r equation (6). The second row is the noise-free M equation (4). The third row is the risky asset equation (7). This is in the form of the general SDE (1) with the identifications

$$X_t = \begin{pmatrix} r_t \\ M_t \\ S_t \end{pmatrix}$$

$$a(X_t) = a(r_t, M_t, S_t) = \begin{pmatrix} -\gamma(\overline{r} - r_t) \\ r_t \\ \mu S_t \end{pmatrix}$$

$$dW_t = \begin{pmatrix} dZ_{0,t} \\ dZ_{r,t} \\ dZ_{S,t} \end{pmatrix}$$

$$b(r_t, M_t, S_t) = \begin{pmatrix} \beta_r & \sigma_r & 0 \\ 0 & 0 & 0 \\ S_t\beta_S & 0 & S_t\sigma_S \end{pmatrix} . \tag{12}$$

The $d = 3$ component diffusion process (11) seems to have $m = 3$ sources of noise, Z_0, Z_r, and Z_S. But this is misleading, because the noise matrix (12) does not have full rank. The process is a degenerate diffusion. In fact, the same model can be re-written using just two independent noise sources.

Infinitesimal generator

The stochastic process X_t has the *infinitesimal generator* (usually just *generator*)

$$\mathcal{L}g(x) = \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} \mu_{jk}(x)\,\partial_{x_j}\partial_{x_k} g + \sum_{j=1}^{d} a_j(x)\,\partial_{x_j} g . \tag{13}$$

Let $f(x,t)$ be a value function of the form

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

Then f satisfies a backward equation

$$\partial_t f + \mathcal{L} f = 0 . \tag{14}$$

This has been stated before, but here is a derivation.

Before the derivation, some explanation of the notation (13) used to define the generator. The numbers μ_{jk} form a $d\times d$ matrix $\mu(x)$. This is related to the noise matrix b by

$$\mu(x) = b(x)\, b^T(x) \ . \tag{15}$$

If you want to work in component form, the entries of μ are given by

$$\mu_{jk} = \sum_{l=1}^{m} b_{jl}\, b_{kl} \ . \tag{16}$$

The formulas (15) or (16) are just covariance formulas.

The double sum in the generator formula (13) may be written in matrix form in several ways. These use notations for the matrix of second partial derivatives of g. This is the *hessian* matrix of g, which may be denoted H, with entries

$$H_{jk} = \partial_{x_j}\partial_{x_k} \ .$$

The hessian matrix is also written as $H = D^2 g$. The double sum in (13) is the trace of the matrix product μ and $D^2 g$

$$\sum_{j=1}^{d}\sum_{k=1}^{d} \mu_{jk}(x)\, \partial_{x_j}\partial_{x_k} g = \mathrm{Tr}(\mu D^2 g) \ . \tag{17}$$

If A is a square matrix, then the trace is the sum of the diagonal entries

$$\mathrm{Tr}(A) = \sum A_{jj} \ .$$

It may seem surprising that this seemingly arbitrary sum is so useful. One explanation is the fact that the trace "commutes" even if the matrices do not. If A is $d\times n$ and B is $n\times d$, then AB is $d\times d$ and BA is $n\times n$. If $n \neq d$, then $AB \neq BA$, just because they are different size. Even in that case, the trace of the $n\times n$ matrix is equal to the trace of the $d\times d$ matrix

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA) \ .$$

In particular, if A is diagonalizable with $RAR^{-1} = \Lambda$ (here, Λ is the diagonal matrix with eigenvalues on the diagonal), then

$$\begin{aligned}
\mathrm{Tr}(\Lambda) &= \mathrm{Tr}(RAR^{-1}) \\
&= \mathrm{Tr}\big[\, R\left(AR^{-1}\right)\big] \\
&= \mathrm{Tr}\big[\left(AR^{-1}\right) R\,\big] \\
&= \mathrm{Tr}\big[\, A\left(R^{-1}R\right)\big] \\
\mathrm{Tr}(\Lambda) &= \mathrm{Tr}(A) \ .
\end{aligned}$$

This shows that the trace of A is the trace of Λ, which is the sum of the eigenvalues of A. If A and B are compatible for multiplication, then

$$(AB)_{jl} = \sum_l A_{jk} B_{kl} \ .$$

Thus,

$$\mathrm{Tr}(AB) = \sum_j \left(\sum_k A_{jk} B_{kj} \right) \ .$$

If B is symmetric, we can replace B_{kj} with B_{jk}. This justifies the trace formula (17).

Another notation for the double sum in (17) is $\mu{::}D^2 g$. You get the inner product of two vectors by multiplying corresponding components and adding the results. You can think of doing the same thing with two matrices. This is written

$$A{::}B = \sum_{jk} A_{jk} B_{jk} \ .$$

These are related by

$$A{::}B = \mathrm{Tr}(AB^T) \ .$$

This is the same as AB if B is symmetric. Thus, the generator may be written in matrix/vector form as

$$\mathcal{L}g(x) = \frac{1}{2}\mu(x){::}D^2 g + a(x) \cdot \nabla g(x) \ .$$

Here is a derivation of the generator formula (13) from the general definition of generator

$$\mathcal{L}g(x) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{E}[\, g(X_{\Delta t}) \mid X_0 = x \,] - g(x)}{\Delta t} \ .$$

For this, we use the Euler approximation

$$X_{j,\Delta t} = x_j + a_j(x)\,\Delta t + \sqrt{\Delta t} \sum_{k=1}^m b_{jk}(x)\,\xi_k \ , \quad [\, \xi_k \sim \mathcal{N}(0,1) \ , \quad \text{i.i.d.}\,] \ .$$

For the upcoming calculations, we write this in the form

$$X_{j,\Delta t} = x_j + \Delta X_j \ , \quad \Delta X_j = a_j(x)\,\Delta t + \sqrt{\Delta t} \sum_{k=1}^m b_{jk}(x)\,\xi_k \ , \qquad (18)$$

We expand $g(X_{\Delta t})$ in a Taylor series around x. We simplify by leaving out the argument x, writing g for $g(x)$, $\partial_j g$ for $\partial_j g(x)$, etc. The result is

$$g(x + \Delta X) = g + \sum_{j=1}^d \partial_j g \,\Delta X_j + \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \partial_{x_j} \partial_{x_k} g \,\Delta X_j \Delta X_k + O(\|\Delta X\|^3) \ .$$

The calculation is done to second order because ΔX is on the order of $\sqrt{\Delta t}$ so ΔX^2 (the quadratic terms in the sum) are on the order of Δt. We do not calculate the ΔX^3 terms because they are of order $\Delta t^{\frac{3}{2}}$ and vanish in the limit $\Delta t \to 0$ even when divided by Δt.

We now find the expectation values needed to evaluate the generator. We start with the simpler terms:

$$\mathrm{E}[\Delta X_j] = a_j(x)\,\Delta t + O(\Delta t^{\frac{3}{2}})\ .$$

This is because the terms ξ_k have mean value zero. Thus

$$\mathrm{E}[\partial_{x_j} g\, \Delta X_j] = \partial_{x_j} g\, a_j(x)\,\Delta t + O(\Delta t^{\frac{3}{2}})\ .$$

Next, we look at the quadratic terms. For these, use the fact that $\Delta X_j \Delta X_k$ has a contribution of order Δt that involves products $\xi_i \xi_l$, but that all other terms in $\Delta X_j \Delta X_k$ have powers $\Delta t^{\frac{3}{2}}$ or Δt^2, which do not contribute as $\Delta t \to$ after dividing by Δt. Therefore, we calculate keeping only terms of order Δt:

$$\mathrm{E}[\,\Delta X_j \Delta X_k] = \Delta t \sum_{i=1}^{m} \sum_{l=1}^{m} b_{ji} b_{kl} \mathrm{E}[\xi_i \xi_l] + O(\Delta t^{\frac{3}{2}})\ .$$

Now, use the fact that ξ_i and ξ_l are independent and mean zero if $i \neq l$, while $\mathrm{E}[\xi_l \xi_l] = 1$. This leads to (using the relation $\mu = bb^T$)

$$\mathrm{E}[\,\Delta X_j \Delta X_k] = \Delta t \sum_{l=1}^{m} b_{ji} b_{kl} + O(\Delta t^{\frac{3}{2}}) = \Delta t\, \mu_{jk} + O(\Delta t^{\frac{3}{2}})\ .$$

This evaluates the hard part of the expectation

$$\mathrm{E}\left[\frac{1}{2} \sum_{jk} \partial_{x_j} \partial_{x_k} g\, \Delta X_j \Delta X_k\right] = \frac{1}{2} \sum_{jk} \partial_{x_j} \partial_{x_k} g\, \mu_{jk} \Delta t + O(\Delta t^{\frac{3}{2}})\ .$$

This is the derivation of the generator formula (13) from the definition of the generator. It is one of the fundamental calculations of the subject of Stochastic Calculus.

Backward equation, tower property

The backward equation can be derived in generator form but here is a slightly different presentation that emphasizes the tower property. The *tower property* is that the conditional expectation of the conditional expectation is the conditional expectation. Specifically, suppose (U, V, W) is a three component random variable. Let $F(u, v, w)$ be some function of these variables and consider the the conditional expectations

$$\begin{aligned} G(u, v) &= \mathrm{E}[\,F(u, v, W) \mid U = u, V = v\,] \\ H(u) &= \mathrm{E}[\,F(u, V, W) \mid U = u\,]\ . \end{aligned}$$

Then $H(u)$ is the conditional expectation of F, conditioned on $U = u$. It is also the conditional expectation of G, which is the conditional expectation of F, conditioned on $U = u$ and $V = v$. In formulas,

$$H(u) = \mathrm{E}[\, G(u,V) \mid U = u\,] \ .$$

We apply this generic tower property to the three variables X_t, X_{t+s}, and X_T, and to the function $F = V(X_T)$. We assume $t < t+s < T$. We are interested in the conditional expectation of a function that depends on one of the three variables

$$f9x) = \mathrm{E}\,[V(X_T) \mid X_t = x\,] \ .$$