1 Diffusion processes, strong and weak

A *diffusion* is a random process that satisfies

$$dX_t \;=\; G_t dt \,+\, F_t dW_t \;. \tag{1}$$

To be more precise, the coefficients G_t and F_t are progressively measurable with respect to a filtration $\mathcal{F}_t$, as is X_t. In general the coefficients F_t and G_t are not required to be functions of X_t, but only to be known at time t. For example, we could take $F_t = \max_{0\leq s\leq t} X_s$. It also is possible to take F_t and G_t to depend on more than just the sample path $X_{[0,t]}$, such as another random process. In that case we talk about diffusions with random coefficients.

There are two ways to interpret the formula (1), which are called strong and weak respectively. The *strong* interpretation takes (1) literally and asks that $G_t dt + F_t dW_t$ be the Ito differential of X_t, in the sense that

$$X_t \;=\; X_0 \,+\, \int_0^t G_s\,ds \,+\, \int_0^t F_s\,dW_s \;. \tag{2}$$

This is most convenient for mathematical analysis. Clearly (2) implies (1).

More convenient for applications is the *weak* interpretation, which simply asks that if $\Delta X = X_{t+\Delta t} - X_t$ (and $\Delta t > 0$), then

$$E[\,\Delta X \mid \mathcal{F}_t\,] \;=\; G_t \Delta t \,+\, o(\Delta t)\;, \tag{3}$$

$$E\big[\,\Delta X^2 \mid \mathcal{F}_t\,\big] \;=\; F_t^2 \Delta t \,+\, o(\Delta t)\;, \tag{4}$$

and that X_t is a continuous function of t. The latter condition follows from, for example,

$$E\big[\,\Delta X^4 \mid \mathcal{F}_t\,\big] \;=\; O\big(\Delta t^2\big)\;. \tag{5}$$

Pay attention to the fact that these conditions do not mention the Brownian motion path W. Whether or not X_t satisfies (1) in the weak sense makes no reference to Brownian motion, but depends only on X_t. In applications, we want to study processes for which we can do the calculations (3) and (4) but have no interest in a hypothetical Brownian motion path that may be related to X_t but is not clear in the system we are modeling.

2 Quadratic variation and total variation

There are various ways to measure how much a function of t moves, or how "active" it is. This section concerns two of these, total variation and quadratic variation. Total variation is familiar to those who have taken enough mathematical analysis. It measures the total amount of movement of a function. Quadratic variation is more subtle. It is appropriate for a class of functions whose total variation is infinite but yet are not infinitely bad. Brownian motion and most other diffusions are in the category of functions with infinite total variation but finite quadratic variation. Quadratic variation is a useful practical quantity in some parts of finance. It measures the difficulty of carrying out hedging strategies for options. It even is possible to buy instruments whose value is determined by (an approximation to) the quadratic variation of a realized financial data series (variance swaps).

If $f(t)$ is a differentiable function of t, then the *total variation* of f up to time T is

$$V_f(T) \;=\; \int_0^T \left|\partial_t f(t)\right| \, dt \; . \tag{6}$$

This may be expressed as a limit like the limits used to define integrals. For small Δt define $t_k = k\Delta t$, then

$$V_f(T) \;=\; \lim_{\Delta t \to 0} \sum_{t_k < T} \left| f(t_{k+1}) - f(t_k) \right| \; . \tag{7}$$

Moreover, it is not even necessary that the points t_k be uniformly spaced. Let $\mathcal{S}_n(T)$ be the set of all increasing sequences of n times less than T. That is, a sequence $(t_1, t_2, \ldots, t_n)$ is in $\mathcal{S}_n(T)$ if it is increasing ($t_{k+1} \geq t_k$), non-negative ($t_1 \geq 0$), and bounded by T ($t_n \leq T$). We implicitly takt $t_0 = 0$ in formulas involving $\mathcal{S}_n(T)$. For differentiable functions, the definition (6) is equivalent to

$$V_f(T) \;=\; \lim_{n \to \infty} \; \sup_{\mathcal{S}_n(T)} \; \sum_{k=0}^{n-1} \left| f(t_{k+1}) - f(t_k) \right| \; . \tag{8}$$

(Recall that sup, for *supremum*, is appropriate if there is no actual largest value.) Even a discontinuous function can have well defined and finite total variation in the sense of (7) or (8). For example, if $f(t) = 2$ for $t < 1$ and $f(t) = 4$ for $t \geq 1$, then $V_f(T) = 0$ if $T < 1$ and $V_f(T) = 2$ for $T \geq 1$. In general, if f has jump discontinuities, the total variation includes the sizes of all the jumps up to time T.

For total variation, (8) is the "right" definition. Despite seeming more complicated than (7) it actually is easier to work with. Also, (8) can detect "variation" that (7) cannot. For example, suppose $f(t) = 0$ for all t except that $f(\sqrt{2}) = 1$. Then (8) gives $V_f(T) = 2$ as long as $T > \sqrt{2}$. But the quantity in the limit in (7) is equal to 0 unless $\sqrt{2}$ is a multiple of Δt. Be that as it may, the notion of quadratic variation is related to (7) rather than (8). The random functions we work with do not have one point discontinuities like that of f.

Although Brownian motion is continuous, its total variation is infinite, more precisely, infinite almost surely. To see this, take $T = 1$ and $\Delta t = 1/n$ and take the limit $n \to \infty$ in (7):

$$R_n \;=\; \sum_{k=0}^{n-1} \left| W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right| \; .$$

The increment of Brownian motion is a mean zero Gaussian, so the distribution of $\left| W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right|$ is the same as the distribution of $Z_k/\sqrt{n}$ where the Z_k are independent with $Z_k \sim \mathcal{N}(0,1)$. Of course $E[|Z_k|] = C > 0$ (the value is not important here, but for the curious, $C = \sqrt{2/\pi}$). Therefore,

$$E[\,R_n\,] \;=\; n\,\frac{C}{\sqrt{n}} \;=\; \sqrt{n}\,C \;\to\; \infty \quad \text{as } n \to \infty.$$

Moreover, the law of large numbers (the sample average of i.i.d. random variables converges almost surely to the actual mean) implies that

$$\frac{1}{\sqrt{n}} R_n \;=\; \frac{1}{n} \sum_{k=0}^{n-1} |Z_k| \;\to\; C \quad \text{as } n \to \infty, \text{ almost surely.}$$

If $\frac{1}{\sqrt{n}} R_n \to C$ then $R_n \to \infty$ as $n \to \infty$.

You can interpret this as saying that Brownian motion always moves infinitely fast. If f has finite total variation, you can define the average speed over an interval as

$$\text{average speed} \;=\; \frac{V_f(t_2) \;-\; V_f(t_1)}{t_2 - t_1} \;.$$

The numerator represents the total distance travelled by f in the time interval (t_1, t_2). The argument we just gave for Brownian motion applied to any time interval and shows that the total distance travelled is infinite in any interval. This means that the speed is always infinite. We already know that in time Δt the Brownian motion net change is $\Delta W = W_{t+\Delta t} - W_t$ is on the order of $\sqrt{\Delta t}$. Therefore, you might say that the average speed over an interval Δt is on the order of $\Delta t^{-1/2}$. But this is misleading. Even in a small interval, W reverses direction so often and so much that the total distance travelled is infinite. The net distance is finite only because of cancellation. The distance travelled in the up direction (which is infinite) is almost exactly balanced by the distanced travelled in the down direction.

The quadratic variation of a function is (notation as in (7))

$$Q_f(T) \;=\; \lim_{\Delta t \to 0} \sum_{t_k < T} |f(t_{k+1}) - f(t_k)|^2 \;. \tag{9}$$

This tolerates much larger $\Delta f = f(t_{k+1}) - f(t_k)$ than total variation does. If Δf is small, then $\Delta f^2 \ll \Delta f$, so it is possible for quadratic variation to be finite when total variation is infinite.

For Brownian motion, the quadratic variation is proportional to T:

$$Q_W(T) \;=\; T \quad \text{almost surely.} \tag{10}$$

To see this, again write $\Delta W_k = W_{t_{k+1}} - W_{t_k} = \sqrt{\Delta t} Z_k$. Then

$$\sum_{t_k < T} |\Delta W_k|^2 \;=\; \Delta t \sum_{t_k < T} |Z_k|^2 \;\to\; T \quad \text{as } \Delta t \to 0, \text{ almost surely.}$$

For small Δt, $|\Delta W|^2$ is smalser than $|\Delta W|$ roughly by a factor of $|\Delta W|$, which is of the order of $\sqrt{\Delta W}$. That's how the quadratic variation sums (9) stay bounded as the total variation sums (7) grow to infinity.

We continue with the theme that quadratic variation is for rougher functions than total variation. If f is continuous and $V_f(T) < \infty$ then $Q_f(T) = 0$. Those who know enough mathematical analysis will be able to prove this using the fact that a continuous function on $[0,T]$ is uniformly continuous. But it is easier to see, informally, that if f is differentiable then $Q_f(T) = 0$. That is because $\Delta f_k = f(t_{k+1}) - f(t_k) \approx \Delta t f'(t_k)$ and

$$Qf(T) \;\approx\; \sum_{t_k<T} |\Delta f_k|^2 \;\approx\; \sum_{t_k<T} \left(\Delta t f'(t_k)\right)^2 \;=\; \Delta t \sum_{t_k<T} f'(t_k)^2 \Delta t \;\approx\; \Delta t \int_0^T f'(t_k)^2 \, dt \;.$$

The rightmost side goes to zero as $\Delta t \to 0$. Altogether, we see that quadratic variation is for continuous functions that move move infinitely fast despite being continuous.

3 Quadratic variation of a diffusion

The general formula for the quadratic variation of a diffusion process that satisfies (1) is

$$Q_X(T) \;=\; \int_0^T F_t^2 \, dt \;. \tag{11}$$

Note that the right side is random in that the values of F_t depend on the path. Unlike Brownian motion, the quadratic variation of a general diffusion is random and path dependent. Also note that the quadratic variation depends only on the noise coefficient F_t, not the drift part G_t. Technically, setting $G = 0$ makes X_t a martingale: $E[X_{t+\Delta t} \mid \mathcal{F}_t] = X_t$. This follows in the strong sense from the fact that the Ito integral with respect to Brownian motion is a martingale. In the weak sense it follows from (3).

We verify the quadratic variation formula (11) in the simpler case $G_t = 0$. Later it will turn out that adding G makes the derivation more complicated without changing the answer. So, suppose X_t satisfies (1) with $G = 0$ and consider the sum

$$Q_X^{\Delta t}(T) \;=\; \sum_{t_k \leq T} \left(X_{t_{k+1}} - X_{t_k}\right)^2 \;. \tag{12}$$

During the next manipulations we will write (12) simply as $Q^{\Delta t}$. As for the Ito integral, the main ideas reasons for things to be true are clear already in the proof that $Q^{\Delta t}$ has a limit as $\Delta t \to 0$. A similar argument evaluates the limit as (11). Also following our treatment of the Ito integral, we prove the limit exists not for every sequence $\Delta t \to 0$ but for the powers of 2 $\Delta t = 2^{-n}$. And the main point of that is to compare $Q^{\Delta t}$ to $Q^{\Delta t/2}$. The summand in $Q^{\Delta t}$, which is $\left(X_{t_{k+1}} - X_{t_k}\right)^2$ is broken into two terms in $Q^{\Delta t/2}$, which are $\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right)^2$, and $\left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2$. Recall that $t_{k+\frac{1}{2}} = (k + \frac{1}{2})\Delta t$ is the time point in the $\Delta t/2$ sum that is halfway between t_k and t_{k+1} in the Δt sum. The difference between $Q^{\Delta t}$ and $Q^{\Delta t/2}$ is (except for a possible term on the end that I ignore)

$$R \;=\; \sum_{t_k<T} R_k \;=\; \sum_{t_k<T} \left[\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right)^2 + \left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2 - \left(X_{t_{k+1}} - X_{t_k}\right)^2 \right] .$$

Note the algebraic "lemma"[1]

$$(c-b)^2 \,+\, (b-a)^2 \,-\, (c-a)^2 \;=\; 2(c-b)(b-a) \; .$$

Thus,

$$R_k \;=\; 2\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right)\left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right) \; .$$

This reveals the mechanism for R being small, which is $E[R_k] = 0$. We see this using the tower property with $\mathcal{F}_{t_{k+\frac{1}{2}}}$ and $\mathcal{F}_{t_k}$:

$$\begin{aligned}
E[R_k \mid \mathcal{F}_{t_k}] \;&=\; E\Big\{ E\Big[R_k \mid \mathcal{F}_{t_{k+\frac{1}{2}}}\Big] \mid \mathcal{F}_{t_k} \Big\} \\
&=\; 2E\Big\{ E\Big[\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right)\left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right) \mid \mathcal{F}_{t_{k+\frac{1}{2}}}\Big] \mid \mathcal{F}_{t_k} \Big\} \\
&=\; 2E\Big\{ \left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right) E\Big[\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right) \mid \mathcal{F}_{t_{k+\frac{1}{2}}}\Big] \mid \mathcal{F}_{t_k} \Big\} \; .
\end{aligned}$$

In the last step we used the fact that $X_{t_{k+\frac{1}{2}}} - X_{t_k}$ is known at time $t_{k+\frac{1}{2}}$, so it is a constant in $\mathcal{F}_{t_{k+\frac{1}{2}}}$ and comes out of the inner expectation. Finally, $E\Big[X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}} \mid \mathcal{F}_{t_{k+\frac{1}{2}}}\Big] = 0$ because X_t is a martingale. Altogether, we have shown that

$$E[R_k \mid \mathcal{F}_{t_k}] \;=\; 0 \; . \tag{13}$$

Still following the discussion of the Ito integral, we show that the sum R is small by computing the expected square and using (13) together with the tower property. The result is

$$E\big[R^2\big] \;=\; \sum_{t_j<T,\, t_k<T} E[R_j R_k] \; .$$

In the sum, either $j = k$ or $j < k$ or $j > k$. If $j < k$ then R_j is known in $\mathcal{F}_{t_k}$, so (with (13))

$$E[R_j R_k] \;=\; E\{E[R_j R_k \mid \mathcal{F}_{t_k}]\} \;=\; E\{R_j \, E[R_k \mid \mathcal{F}_{t_k}]\} \;=\; E\,\{R_j \cdot 0\} \;=\; 0 \; .$$

[1] You can derive this without doing the algebra by noting that the left side vanishes when $c = b$ or $b = a$. This means that there must be a factor of $(c-b)$ and a factor of $(b-a)$. Since the left side is quadratic, and $(c-b)(b-a)$ is quadratic, they are equal up to a constant. You can evaluate the constant by setting $a = c = 0$, in which case the left side is $2b^2$.

The terms $j > k$ vanish for the same reason. The $j = k$ terms do not vanish but are easy to estimate:

$$\begin{aligned}
E\left[R_k^2\right] &= 4E\left[\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right)^2 \left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2\right] \\
&= 4E\left\{E\left[\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right)^2 \left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2 \mid \mathcal{F}_{t_{k+\frac{1}{2}}}\right]\right\} \\
&= 4E\left\{\left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2 E\left[\left(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}\right)^2 \mid \mathcal{F}_{t_{k+\frac{1}{2}}}\right]\right\} \\
&\approx 4E\left\{\left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2 F^2_{t_{k+\frac{1}{2}}} \frac{\Delta t}{2}\right\} .
\end{aligned}$$

The last step uses (4) and the fact that $t_{k+1} = t_{k+\frac{1}{2}} + \frac{\Delta t}{2}$. Now suppose $F_{t_{k+\frac{1}{2}}}$ is bounded and get

$$E\left\{\left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2 F^2_{t_{k+\frac{1}{2}}}\right\} \le CE\left\{\left(X_{t_{k+\frac{1}{2}}} - X_{t_k}\right)^2\right\} \le C\Delta t .$$

Altogether, $E[R_k^2] \le C\Delta t^2$, so $E[R^2] \le C\sum_{t_k<T} \Delta t^2 = CT\Delta t$. Now look back at the Borel-Cantelli type lemma in the discussion of the Ito integral convergence and you will see that this implies that the limit of the $Q^{2^{-n}}$ exists almost surely as $n \to \infty$.

We can establish the value of the limit using similar methods. We want to show that

$$\sum_{t_k<T} \left(X_{t_{k+1}} - X_{t_k}\right)^2 \approx \sum_{t_k<T} F^2_{t_k} \Delta t . \qquad (14)$$

So we compute the expected square of the difference, which is

$$S = \sum_{t_k<T} S_k = \sum_{t_k<T} \left[\left(X_{t_{k+1}} - X_{t_k}\right)^2 - F_{t_k}\Delta t\right]$$

In the double sum for $E[S^2]$, the $j \neq k$ terms are small for the same reasons as before. This is the reason (4) refers to conditional variance with respect to $\mathcal{F}_t$ rather than simple unconditional variation. The terms $j = k$ have

$$E\left[\,S_k^2\right] = O(\Delta t^2) . \qquad (15)$$

Therefore, the difference between the right and left sides of (14) has expected square $E[S^2] = O(\Delta t) \to 0$ as $\Delta t \to 0$. The right side converges to the Riemann integral that is the right side of (11). The limit of the left is the definition of the left side of (11).

It is convenient in stochastic calculus to write differential expressions that are informal versions of more formal integral formulas. The differential expression related to (11) is

$$dQ_X(t) = F_t^2\, dt . \qquad (16)$$

Given the definition of quadratic variation, it makes some sense to write the limit of the left side of (12) as

$$Q_X(T) \;=\; \int_0^T \left(dX_t\right)^2 \;.$$

Although (16) is already informal, this leads to an even more informal expression:

$$\left(dX_t\right)^2 \;=\; F_t^2\, dt \;. \tag{17}$$

This is one of the formulas called Ito's lemma for general diffusions. It is true in the sense of our earlier expression for Brownian motion: $\left(dW_t\right)^2 = dt$. The sides are not equal "pointwise", because the left side is random and the right side is not. But they they have the same expected value and the statistical fluctuations become unimportant if you average over any interval of time. The original formula (11) is the expression of this fact that makes real mathematical sense, one actual well defined integral being equal to a different actual well defined integral. In the present case, it is $E\left[\left(dX_t\right)^2\right] \;=\; F_t^2 dt$, which is the infinitesimal version of (4), and $\mathrm{var}\left[\left(dX_t\right)^2\right] \;\propto\; dt^2$, which is the infinitesimal version of (15).

It is a straightforward if time consuming exercise to show that the quadratic variation exists and (11) holds (in any of its forms (16) or (17)) even when G_t is not zero.

If we take (1) seriously in the strong sense, then these facts follow immediately from corresponding facts in the ordinary Ito calculus. For example,

$$\left(dX_t\right)^2 \;=\; \left(G_t dt \,+\, F_t dW_t\right)^2 \;=\; G_t^2 (dt)^2 + 2G_tF_t(dt dW_t) + F_t^2\left(dW_t\right)^2 \;=\; F_t^2 dt \;,$$

as (17) (the first two terms are zero because they integrate to zero over any interval).

4 The stochastic integral with a diffusion

Suppose H_t is a non-anticipating function with respect to the filtration $\mathcal{F}_t$. Then we can define the stochastic integral with respect to X that generalizes the Ito integral with respect to Brownian motion. That is, we can define a process

$$Y_t \;=\; \int_0^T H_t dX_t \;. \tag{18}$$

By now you should be able to define this integral as

$$Y_T \;=\; \lim_{\Delta t \to 0} \sum_{t_k < T} H_{t_k}\left(X_{t_{k+1}} - X_{t_k}\right) \;.$$

This Y_t is a diffusion that satisfies

$$dY_t \;=\; H_t\, dX_t \;=\; H_tG_t\, dt \;+\; H_tF_t\, dW_t \;.$$

The weak interpretation of this is

$$E[\Delta Y_t \mid \mathcal{F}_t] \;=\; H_t G_t \, \Delta t \;+\; o(\Delta t) \;,$$

$$E\big[\Delta Y_t^2 \mid \mathcal{F}_t\big] \;=\; H_t^2 F_t^2 \, \Delta t \;+\; o(\Delta t) \;,$$

and the fact that Y_t is a continuous function of t.

5 Ito's lemma for a general diffusion

Suppose $u(x,t)$ is a smooth function of x and t. The analogue of Ito's lemma for a general diffusion clearly is

$$du(X_t,t) \;=\; \partial_x u(X_t,t) dX_t \;+\; \partial_t u(X_t,t) dt \;+\; \frac{1}{2} \partial_x^2 u(X_t,t) \, (dX_t))^2 \;.$$

Of course, we should rewrite this using (17):

$$du(X_t,t) \;=\; \partial_x u(X_t,t) dX_t \;+\; \partial_t u(X_t,t) dt \;+\; \frac{1}{2} \partial_x^2 u(X_t,t) F_t^2 dt \;. \tag{19}$$

We used (17) to go from $(dX_t)^2$ to $F_t^2 dt$. We could instead have used (16) to write the last term as $\frac{1}{2}\partial_x^2 u(X_t,t) dQ_X(t)$ as many mathematicians prefer. Of course, (19) really is an informal expression of the formal integral identity

$$u(X_T,T) - u(X_0,0) \;=\; \int_0^T \left(\partial_t u(X_t,t) \;+\; \frac{1}{2} \partial_x^2 u(X_t,t) F_t^2 \right) dt + \int_0^T \partial_x u(X_t,t) \, dX_t \;.$$

Of course, the first integral on the right is an ordinary Riemann integral. The second integral is the stochastic integral with respect to the diffusion process X_t.