STOCHASTIC INTEGRALS

Suppose S_t denotes the price of a stock at time t, for $0 \le t \le 1$. Let $0 = t_0 < t_1 < \ldots < t_n < t_{n+1} = 1$ be times at which you buy and sell stock: at time t_i you buy $H(t_i)$ stocks at a cost of $H(t_i)S(t_i)$ then you sell the same stocks at time t_{i+1} for $H(t_i)S(t_{i+1})$. Your total profit will be

$$\sum\nolimits_{i=0}^{n} H(t_i)\Delta_i S \qquad \text{where } \Delta_i S = S(t_{i+1}) - S(t_i).$$

This formula is also valid for purchases of random numbers of shares. In that case, $H(t_i)$ should depend only on information available at time t_i, otherwise you might be jailed for insider trading.

It is tempting to think that if the times between trades get smaller and smaller then we get closer and closer to some limit, an idealized continuous trading, with total profit being given by some sort of limit of the sums for trading in discrete time. To formalize this idea, we need to define a ***stochastic integral*** $\int_0^1 H_s dS_s$.

There is a large class of processes for which stochastic integrals can be defined. A complete treatment usually takes up a large fraction of the graduate course on Stochastic Calculus. However, with enough handwaving I can explain the main ideas.

TECHNICAL TERMS.

Throughout what follows, the information available up to time t will be denoted by $\mathcal{F}_t$. In rigorous developments of the theory, $\mathcal{F}_t$ is identified with a collection of subsets called a ***sigma-field***. A random vaiable tha depends only on $\mathcal{F}_t$-information is said to be $\mathcal{F}_t$***-measurable***. The flow of information $\{\mathcal{F}_t : 0 \le t \le 1\}$ is often called a ***filtration***.

A stochastic process $\{X_t(\omega) : 0 \le t \le 1\}$ is said to be ***adapted*** (to the filtration) if X_t depends only on $\mathcal{F}_t$-information, for each t.

I will denote the conditional expectation $\mathbb{E}(\ldots \mid \mathcal{F}_s)$ by $\mathbb{E}_s(\ldots)$. Remember that $\mathbb{E}_s Y$ is a random variable that depends only on $\mathcal{F}_s$ for which

$$\mathbb{E}\left(W(Y - \mathbb{E}_s Y)\right) = 0 \qquad \text{for all } W \text{ depending only on } \mathcal{F}_s.$$

In particular, if $\{Y_t : 0 \le t \le 1\}$ is a martingale then $\mathbb{E}_s Y_t = Y_s$ for all $s < t$, and hence

$$\mathbb{E}\left(W(Y_t - Y_s)\right) = 0 \qquad \text{for all } W \text{ depending only on } \mathcal{F}_s.$$

I often remind you of this equality by noting that the increment $\Delta Y := Y_t - Y_s$ is orthogonal to every W that depends only on $\mathcal{F}_s$-information.

I will construct stochastic integrals via approximation on a grid, $\mathbb{G} : 0 = t_0 < t_1 < \ldots < t_n < t_{n+1} = 1$. The quantity mesh($\mathbb{G}$) is defined to equal $\max_i(t_{i+1} - t_i)$. A grid $\mathbb{G}_1$ is said to be a refinement of a grid $\mathbb{G}_0$ if it is obtained by adding extra grid points. If $\mathbb{G}_1 : 0 = t_0 < t_1 < \ldots < t_n < t_{n+1} = 1$ and $\mathbb{G}_2 : 0 = s_0 < s_1 < \ldots < t_n < s_{m+1} = 1$ are grids then I will write $\mathbb{G}_1 \vee \mathbb{G}_2$ for their common refinement, the grid obtained by arranging all the s_j's and t_i's into one increasing sequence. Of course, we need retain only one copy of duplicate grid points.

> In what follows, I have been sloppy about stating regularity conditions. You should not take the assertions to be true precisely as stated. You would need to take the Stochastic Calculus course if you wanted to know the truth, almost the whole truth, and hardly anything but the truth.

It is easiest to start with a deterministic case.

1. Functions of bounded variation

Suppose f and g are continuous functions defined on the interval $[0, 1]$. Remember that the variation of f over a grid $\mathbb{G} : 0 = t_0 < t_1 < \ldots < t_n < t_{n+1} = 1$ is defined as

$$\mathcal{V}(f, \mathbb{G}) = \sum\nolimits_{i=0}^{n} |f(t_{i+1}) - f(t_i)|,$$

and f is said to be of bounded variation if $\mathcal{V}(f) = \sup_{\mathbb{G}} \mathcal{V}(f, \mathbb{G})$ is finite.

We might hope that $\int_0^1 g(t)df(t)$ could be obtained as a limit of approximating sums,

$$\mathcal{I}(g, \mathbb{G}) = \sum_{i=0}^{n} g(t_i)\Delta_i f \qquad \text{where } \Delta_i f = f(t_{i+1}) - f(t_i).$$

In fact, such a limit does exist, in the sense that there is number J such that $\mathcal{I}(f, \mathbb{G}) \to J$ as mesh($\mathbb{G}$) tends to zero. Of course, the limit J is then denoted by $\int_0^1 g df$.

<1> **Theorem.** *If g is continuous and f is both continuous and of bounded variation, then there is a number J for which $\mathcal{I}(g, \mathbb{G}) \to J$ as $\text{mesh}(\mathbb{G}) \to 0$.*

Proof. Continuity of g on a closed interval ensures that for each $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that

<2> $$|g(t) - g(s)| \le \epsilon \qquad \text{whenever } |t - s| \le \delta_\epsilon.$$

Let $\mathbb{G}_0$ be a grid with $\text{mesh}(\mathbb{G}_0) \le \delta_\epsilon$ and let $\mathbb{G}_1$ be a refinement of $\mathbb{G}_0$.

Consider the contributions to both $\mathcal{I}(g, \mathbb{G}_0)$ and $\mathcal{I}(g, \mathbb{G}_1)$ from the interval $[t_i, t_{i+1}]$. Suppose $\mathbb{G}_1$ puts grid points $s_0 = t_i < s_1 < \ldots < s_k < s_{k+1} = t_{i+1}$ in the interval. The contribution to $\mathcal{I}(g, \mathbb{G}_1)$ from the interval is

$$\sum_{j=0}^{k} g(s_j)\Delta_j f \qquad \text{where } \Delta_j f = f(s_{j+1}) - f(s_j)$$

The contribution to $\mathcal{I}(g, \mathbb{G}_0)$ is

$$g(t_i)\left(f(t_{i+1}) - f(t_i)\right) = g(t_i)\sum_{j=0}^{k} \Delta_j f.$$

The absolute value of the difference between the two contributions is bounded by

$$\sum_{j=0}^{k} |g(t_i) - g(s_j)|\,|\Delta_j f| \le \epsilon \sum_{j=0}^{k} |\Delta_j f| \qquad \text{because } |t_i - s_j| \le \delta.$$

Summing over all i, we conclude that

$$|\mathcal{I}(g, \mathbb{G}_0) - \mathcal{I}(g, \mathbb{G}_1)| \le \epsilon\mathcal{V}(f, \mathbb{G}_1) \le \epsilon\mathcal{V}(f)$$

Suppose $\{\mathbb{G}_n : n \in \mathbb{N}\}$ is a sequence of grids with each $\mathbb{G}_{n+1}$ a refinement of the preceding $\mathbb{G}_n$ and $\text{mesh}(\mathbb{G}_n) \to 0$. The argument in the previous two paragraphs implies that $J = \lim_n \mathcal{I}(g, \mathbb{G}_n)$ exists. (Formal reason: the real numbers $\mathcal{I}(g, \mathbb{G}_n)$ form a Cauchy sequence.) The limit does not depend on the choice of the $\mathbb{G}_n$. (Formal reason: if $\text{mesh}(\mathbb{G}) \le \delta_\epsilon$ then $\mathcal{I}(g, \mathbb{G} \vee \mathbb{G}_n)$ lies within ϵ of both $\mathcal{I}(g, \mathbb{G})$ and $\mathcal{I}(g, \mathbb{G}_n)$ for n large enough.) □

For the purposes of this handout, there are two important cases where a function f has bounded variation.

(i) If f is an increasing function on $[0, 1]$ then $\mathcal{V}(f) = f(1) - f(0)$, because

$$\sum_{i=0}^{n} |f(t_{i+1}) - f(t_i)| = \sum_{i=0}^{n} \left(f(t_{i+1}) - f(t_i)\right) = f(1) - f(0)$$

for every grid.

(ii) If $f(t) = \int_0^1 \lambda(s)\,ds$, with $\int_0^1 |\lambda(s)|\,ds < \infty$ then

$$\sum_{i=0}^{n} |f(t_{i+1}) - f(t_i)| = \sum_{i=0}^{n} \Big|\int_{t_i}^{t_i+1} \lambda(s)\,ds\Big| \le \sum_{i=0}^{n} \int_{t_i}^{t_i+1} |\lambda(s)|\,ds = \int_0^1 |\lambda(s)|\,ds.$$

In this case, it is not hard to show that $\int_0^1 g(s)\,df(s) = \int_0^1 g(s)\lambda(s)\,ds$.

A similar method of approximation could be used to define $\int_0^t g\,df$ for each t in $[0, 1]$. A better way is to build the dependence on t into the approximation, by defining

$$\mathcal{I}(g, \mathbb{G})_t = \sum_{i=0}^{n} g(t_i)\left(f(t_{i+1} \wedge t) - f(t_i \wedge t)\right).$$

If t equals t_i, we have $t_j \wedge t = t_i$ for all $j \ge i$, which ensures that the all summands for $j \ge i$ vanish. If $t_i < t < t_{i+1}$, the ith summand becomes $g(t_i)\left(f(t) - f(t_i)\right)$, which is continuous in t. Indeed, the insertion of the $\wedge t$ makes $\mathcal{I}(g, \mathbb{G})_t$ a continuous function of t. The argument from the proof of Theorem <1> still works, leading to the conclusion that $\int_0^t g\,df$ is a uniform limit of continuous functions, and hence is itself continuous as a function of t.

By various approximation arguments, the integral can also be extended to integrands g that are not continuous. I won't discuss this extension, because we will only need continuous integrands.

2. Stochastic integral for BV processes

Suppose $\{X_t(\omega) : 0 \le t \le 1\}$ is a stochastic process for which each sample path $X(\cdot,\omega)$ is continuous and of bounded variation. If $\{H_t(\omega) : 0 \le t \le 1\}$ is another stochastic process with continuous sample paths, then we can define the stochastic integral pathwise. That is, $\int_0^t H(s,\omega)\,dX(s,\omega)$ is defined using the method described above for each ω.

It often helps to think of the stochastic integral as defining a new stochastic process $H \bullet X$ with continuous sample paths:

$$(H \bullet X)(t,\omega) := \int_0^t H_s(\omega)\,dX_s(\omega)$$

The same notation will reappear in later sections, for stochastic integrals with respect to more complicated processes.

3. Stochastic integral with respect to Brownian motion

We know that almost all sample paths of a standard Brownian motion $\{B_t : 0 \le t \le 1\}$ have infinite total variation. We cannot expect the method from Section 2 to work to define a stochastic integral $\int_0^t H_s(\omega)\,dB_s(\omega)$.

For a smaller class of functions, the definition of the stochastic integral is easy. Suppose H is an ***elementary process***, that is, for some grid $0 = t_0 < t_1 < \ldots < t_n < t_{n+1} = 1$,

$$H(t,\omega) = \sum\nolimits_{i=0}^n h_i(\omega)\mathbb{I}\{t_i < t \le t_{i+1}\},$$

where h_i is a random variable that depends only on $\mathcal{F}_{t_i}$-information (that is, H is adapted). Then we define

$$\begin{aligned} H \bullet B_t := \int_0^t H_s(\omega)\,dB_s(\omega) &:= \sum\nolimits_{i=0}^n h_i(\omega)\left(B(t_{i+1}\wedge t,\omega) - B(t_i \wedge t,\omega)\right) \\ &= \sum\nolimits_{i=0}^{j-1} h_i(\omega)\left(B(t_{i+1}) - B(t_i)\right) + h_j(\omega)(B(t) - B(t_j)) \qquad \text{if } t_j \le t \le t_{j+1}. \end{aligned}$$

In the last expression I have omitted some ω argument to fit everything into one line.

Notice that, for a fixed elementary process H, nothing changes if we add extra grid points. For example, the addition of a new point $\bar{t}$ with $t_i < \bar{t} < t_{i+1}$ replaces the summand

$$h_i\left(B(t_{i+1}\wedge t) - B(t_i \wedge t)\right)$$

by two terms,

$$h_i(\omega)\left(B(\bar{t}\wedge t) - B(t_i\wedge t)\right) + h_i(\omega)\left(B(t_{i+1}\wedge t) - B(\bar{t}\wedge t)\right),$$

which leaves $H \bullet B_t$ is unchanged.

The process $H \bullet B$ has continuous sample paths. It also inherits from B the martingale property. For suppose that $s < t$ and that W depends only on $\mathcal{F}_s$-information. With no loss of generality (as explained in the previous paragraph), we may assume that both s and t are grid points: $s = t_j$ and $t = t_k$. Then

$$\mathbb{E}W\left(H \bullet B_t - H \bullet B_s\right) = \sum\nolimits_{i=j}^{k-1} \mathbb{E}\left(W h_i \Delta_i B\right) \qquad \text{where } \Delta_i B = B(t_{i+1}) - B(t_i).$$

The ith summand vanishes because Wh_i depends only on $\mathcal{F}_{t_i}$-information and $\mathbb{E}_{t_i}\Delta_i B = 0$.

The martingale properties also lead to a simple expression for the second moment of $H \bullet B_1$.

$$\mathbb{E}\left(H \bullet B_1\right)^2 = \sum\nolimits_{i=0}^n \sum\nolimits_{j=0}^n \mathbb{E}\left(h_i h_j \Delta_i B \Delta_j B\right)^2 = \sum\nolimits_{i=0}^n \mathbb{E}\left(h_i \Delta_i B\right)^2 + 2\sum\nolimits_{i<j} \mathbb{E}\left((h_i h_j \Delta_i B)\mathbb{E}_{t_j}\Delta_j B\right).$$

For $i < j$, the product $h_i h_j \Delta_i B$ depends only on $\mathcal{F}_{t_j}$-information and $\mathbb{E}_{t_j} \Delta_j B = 0$. All the cross product contributions have zero expectation. Only the terms with $i = j$ survive, leaving

$$\begin{aligned}\sum\nolimits_{i=0}^{n} \mathbb{E}\left(h_i^2 (\Delta_i B)^2\right) &= \sum\nolimits_{i=0}^{n} \mathbb{E}\left((t_{i+1} - t_i) h_i^2\right) \qquad \text{because } \mathbb{E}_{t_i}(\Delta_i B)^2 = t_{i+1} - t_i \\ &= \mathbb{E} \int_0^1 H(s, \omega)^2 \, ds.\end{aligned}$$

I have included the ω argument in the final expression to emphasize the two averagings involved: one over s and the other over ω.

<3> **Definition.** *For a process $\{H(t, \omega) : 0 \le t \le 1\}$ define* $\llbracket H \rrbracket = \left(\mathbb{E} \int_0^1 H(s, \omega)^2 \, ds\right)^{1/2}$.

More generally, if G is also an elementary process, by taking a common refinement of the G and H grids we see that $G - H$ is also an elementary process, and hence

<4>
$$\mathbb{E}\left(H \bullet B_1 - G \bullet B_1\right)^2 = \llbracket H - G \rrbracket^2 = \mathbb{E} \int_0^1 \left(H(s, \omega) - G(s, \omega)\right)^2 \, ds.$$

<5> **Definition.** *Write $\mathcal{H}_2$ for the set of adapted processes H for which there exists at least one sequence of elementary processes $\{H_n\}$ with $\llbracket H_n - H \rrbracket \to 0$.*

Equality <4> justifies the definition the stochastic integral $H \bullet B$ for $H \in \mathcal{H}_2$ as a limit of stochastic integrals of elementary processes. The proof makes use of an inequality of Doob, which can be proved using the STL. For a martingale $\{M_t : 0 \le t \le 1\}$ with continuous sample paths (actually right continuity would suffice),

<6>
$$\mathbb{E} \sup\nolimits_{0 \le t \le 1} |M_t|^2 \le 4 \mathbb{E} M_1^2$$

<7> **Theorem.** *There is an extension of the stochastic integral for elementary processeses to a linear map $H \mapsto H \bullet B$ from $\mathcal{H}_2$ into the space of all martingales with continuous samples paths such that*

<8>
$$\mathbb{E}\left(G \bullet B_1 - H \bullet B_1\right)^2 = \llbracket G - H \rrbracket^2$$

for all $G, H \in \mathcal{H}_2$.

Sketch of a proof. Suppose $H \in \mathcal{H}_2$. By taking a subsequence if necessary, we may suppose that $\{H_n\}$ is a sequence of elementary processes for which

$$\llbracket H_n - H \rrbracket \le 2^{-n} \qquad \text{for each } n.$$

By inequality <6> applied to the martingale $H_n \bullet B - H_{n+1} \bullet B$,

<9>
$$\begin{aligned}\mathbb{E} \sup_{0 \le t \le 1} |H_n \bullet B_t - H_{n+1} \bullet B_t| &\le \left(\mathbb{E} \sup_{0 \le t \le 1} |H_n \bullet B_t - H_{n+1} \bullet B_t|^2\right)^{1/2} \\ &\le \left(4 \mathbb{E} |H_n \bullet B_1 - H_{n+1} \bullet B_1|^2\right)^{1/2} \\ &\le 2 \llbracket H_n - H_{n+1} \rrbracket \le 4/2^n.\end{aligned}$$

Thus

$$\mathbb{E} \sum\nolimits_{n \in \mathbb{N}} \sup\nolimits_{0 \le t \le 1} |H_n \bullet B_t - H_{n+1} \bullet B_t| < \infty,$$

which implies that there exists a process $\{J_t : 0 \le t \le 1\}$ for which

$$\sup_{0 \le t \le 1} |H_n \bullet B_t(\omega) - J_t(\omega)| \to 0 \qquad \text{for each } \omega \text{ in a set } \Omega_0 \text{ with } \mathbb{P}\Omega_0 = 1.$$

The limit process J inherits continuous sample paths and the martingale property from $H_n \bullet B$. (I am ignoring what happens on the set Ω_0^c, which has zero probability.)

By further subsequencing arguments, we could show that the limit process does not depend on the choice of the sequence of elementary processes. It therefore risks little ambiguity if we write $H \bullet B$ for J. We could also show, by an another argument starting from <9>, that

$$\mathbb{E} \sup\nolimits_{0 \le t \le 1} |H_n \bullet B_t - H \bullet B_t|^2 \to 0$$

Linearity and the isometry property <8> then follow from the analogous properties for stochastic integrals of elementary processes. □

<10> **Example.** In a course on stochastic integrals it is almost mandatory to show that $B \bullet B_t = \frac{1}{2}(B_t^2 - t)$. Notice the extra $-t$ on the right-hand side. Without it, we would not have a martingale because $\mathbb{E}_s B_t^2 = B_s^2 + (t-s)$ for $s < t$.

Define simple functions

$$H_n(t) = \sum\nolimits_{i=0}^{n} B(t_i)\mathbb{I}\{t_i < t \le t_{i+1}\} \qquad \text{where } t_i = i/(n+1).$$

Notice that

$$|H_n(t) - B(t)|^2 = \sum\nolimits_{i=0}^{n} \big(B(t) - B(t_i)\big)^2 \mathbb{I}\{t_i < t \le t_{i+1}\}$$

and

$$\int_0^1 \mathbb{E}|H_n(t) - B(t)|^2\,dt = \sum\nolimits_{i=0}^{n} \int_0^1 |t - t_i|\mathbb{I}\{t_i < t \le t_{i+1}\} = \tfrac{1}{2}\frac{n+1}{(n+1)^2} \to 0 \qquad \text{as } n \to \infty.$$

From Theorem <7>,

$$\mathbb{E}\big(H_n \bullet B_1 - B \bullet B_1\big)^2 = [\![H_n - B]\!]^2 \to 0.$$

We have only to calculate $H_n \bullet B_1$ then pass to the limit to find $B \bullet B_1$.

For a fixed n, write $\Delta_i B$ for $B(t_{i+1}) - B(t_i)$. Then

$$\begin{aligned}(B_1^2 - 0^2) - 2H_n \bullet B_1 &= \sum\nolimits_{i=0}^{n} \big(B(t_{i+1})^2 - B(t_i)^2\big) - \sum\nolimits_{i=0}^{n} 2B(t_i)\Delta_i B \\ &= \sum\nolimits_{i=0}^{n} (\Delta_i B)^2.\end{aligned}$$

From facts about the quadratic variation of Brownian motion, we know that the final sum converges (in probability) to 1 as $n \to \infty$. It follows that $2B \bullet B_1 = B_1^2 - 1$. We could carry out a similar approximation argument to find $B \bullet B_t$, but there is an easier way. The stochastic integral is a martingale, which implies

$$B \bullet B_t = \mathbb{E}_t(B \bullet B_1) = \tfrac{1}{2}\mathbb{E}_t(B_1^2 - 1) = \tfrac{1}{2}(B_t^2 - t),$$

as asserted. □

4. Stochastic integral with respect to a (square integrable) martingale

The method for defining stochastic integrals with respect to Brownian motion also works for a more general martingale $\{M_t : 0 \le t \le 1\}$ with continuous sample paths. Suppose that there exists an adapted process $\{A_t : 0 \le t \le 1\}$ with continuous, increasing sample paths for which

<11>
$$M_t^2 - A_t \qquad \text{is a martingale.}$$

For a fixed $s < t$ write ΔM for $M_t - M_s$ and ΔA for $A_t - A_s$. Then

<12>
$$0 = \mathbb{E}_s\big((M_s + \Delta M)^2 - A_s - \Delta A\big) - M_s^2 + A_s = \mathbb{E}_s\big((\Delta M)^2 - \Delta A\big).$$

For an elementary process $H(t,\omega) = \sum_{i=0}^{n} h_i(\omega)\mathbb{I}\{t_i < t \le t_{i+1}\}$, we define

<13>
$$(H \bullet M)_t := \int_0^t H_s(\omega)\,dM_s(\omega) := \sum\nolimits_{i=0}^{n} h_i(\omega)\big(M(t_{i+1} \wedge t, \omega) - M(t_i \wedge t, \omega)\big).$$

Almost the same argument as in Section 3 shows that $H \bullet M$ is a martingale with continuous sample paths. Moreover,

$$\begin{aligned}\mathbb{E}\big(H \bullet M_1\big)^2 &= \sum\nolimits_{i=0}^{n} \mathbb{E}\big(h_i^2(\Delta_i M)^2\big) \qquad \text{where } \Delta_i M = M(t_{i+1}) - M(t_i) \\ &= \sum\nolimits_{i=0}^{n} \mathbb{E}\big(h_i^2 \Delta_i A\big) \qquad \text{by <12>, where } \Delta_i A = A(t_{i+1}) - A(t_i) \\ &= \mathbb{E}(H^2 \bullet A_1) \qquad \text{with } H^2 \bullet A \text{ defined as in Section 2.}\end{aligned}$$

And so on.

The role of $[\![H]\!]^2$ is taken over by the quantity

$$[\![H]\!]_A^2 =: \mathbb{E}(H^2 \bullet A_1) = \mathbb{E}\int_0^1 H(t,\omega)^2\, dA(t,\omega).$$

It is possible to define $H \bullet M$ for H in the set $\mathcal{H}_2(A)$ of processes that can be approximated in the $[\![\cdot]\!]_A$ sense by elementary processes. The resulting stochastic process is again a martingale with continuous sample paths, for which

<14> $$\mathbb{E}\left(G \bullet M_1 - H \bullet M_1\right)^2 = [\![G - H]\!]_A^2 \qquad \text{for all } G, H \in \mathcal{H}_2(A).$$

<15> **Example.** Suppose $\{X_t : 0 \le t \le 1\}$ is an adapted stochastic process with continuous sample paths. Suppose also that there exist adapted processes μ and σ with continuous sample paths, such that

$$\mathbb{E}_t(X_{t+h} - X_t) = h\mu(t,\omega) + \text{ smaller order terms}$$
$$\mathbb{E}_t(X_{t+h} - X_t)^2 = h\sigma^2(t,\omega) + \text{ smaller order terms}$$

Interpret the first approximation to mean that

$$Z_t = X_t - \int_0^t \mu(s,\omega)\, ds \qquad \text{is a martingale.}$$

The second approximation then gives

$$\mathbb{E}_s\left(Z_{t+h} - Z_t\right)^2 = \mathbb{E}_s(X_{t+h} - X_t)^2 - \left(\mu(t,\omega)h + \ldots\right)^2 = h\sigma^2(t,\omega) + \text{ smaller order terms},$$

which we can interpret to mean that

$$Z_t^2 - \int_0^t \sigma^2(s,\omega)\, ds \qquad \text{is a martingale.}$$

If we write D_t for the drift $\int_0^t \mu(s,\omega)\, ds$ and A_t for the increasing process $\int_0^t \sigma^2(s,\omega)\, ds$, we can define

$$H \bullet X = H \bullet Z + H \bullet D,$$

with the martingale $H \bullet Z$ defined as above and the bounded variation process $H \bullet D$ defined as in Section 2. □

<16> **Lemma.** *If the martingale M has property <11> and if $Z = G \bullet M$ (a new martingale), then the increasing process $\Lambda = (G^2) \bullet A$ makes $Z_t^2 - \Lambda_t$ a martingale.*

Rough proof. Suppose G is an elementary process,

$$G_t = \sum\nolimits_{i=0}^n g_i(\omega)\mathbb{I}\{t_i < t \le t_{i+1}\}.$$

For fixed $s < t$, define $\Delta Z = Z_t - Z_s$ and $\Delta\Lambda = \Lambda_t - \Lambda_s$ We need to show that $\mathbb{E}\left(W((\Delta Z)^2 - \Delta\Lambda\right) = 0$ for W depending only on $\mathcal{F}_s$-information. With no loss of generality, we may assume that both s and t are grid points: $s = t_j$ and $t = t_k$. Then

$$\Delta Z = \sum\nolimits_{i=j}^{k-1} g_i \Delta_i M \qquad \text{where } \Delta_i M = M(t_{i+1}) - M(t_i)$$
$$\Delta\Lambda = \sum\nolimits_{i=j}^{k-1} g_i^2 \Delta_i A \qquad \text{where } \Delta_i A = A(t_{i+1}) - A(t_i).$$

Expand the quadratic then subtract.

$$\mathbb{E}\left(W((\Delta Z)^2 - \Delta\Lambda\right) = \sum\nolimits_{i=j}^{k-1} \mathbb{E} W g_i^2 \left((\Delta_i M)^2 - \Delta_i A\right) + 2\sum\nolimits_{i<\ell} \mathbb{E}\left(W g_i g_\ell \Delta_i M\right)\left(\Delta_\ell M\right)$$

Each term in the first sum vanishes, by virtue of the martingale property <12> and the fact that $W g_i^2$ depends only on $\mathcal{F}_{t_i}$-information. Each of the cross product terms in the double sum vanishes because $W g_i g_\ell \Delta_i M$ depends only on $\mathcal{F}_{t_\ell}$-information and $\mathbb{E}_{t_\ell}\Delta_\ell M = 0$.

Chant appropriate incantations as elementary functions converge to the general G, wave hands ignoring various hidden moment assumptions, then declare the same property to hold in the limit. □

Actually, you have most of the tools necessary to make the passage from elementary G to more general processes rigorous. Consider a sequence of elementary processes for which $[\![G_n - G]\!]_A \to 0$. Then what?

If the increasing process A in <12> is given as in Example <15>, then

$$(G^2) \bullet A_t = \int_0^t G_s^2 \sigma_s^2 \, ds.$$

If we choose $G_s = 1/\sigma_s$ then $\Lambda_t = t$. That is, the martingale

$$\tilde{B}_t = (1/\sigma) \bullet M_t = \int_0^t (1/\sigma_s) \, dM_s$$

has the property that $\tilde{B}_t^2 - t$ is also a martingale with continuous sample paths. The Lévy's characterization shows that $\tilde{B}$ is a Brownian motion. Moreover, another argument passing from elementary functions to the limit would show that $M = \sigma \bullet \tilde{B}$. That is, the martingale M can be constructed as a stochastic integral with respect to a Brownian motion.

5. Localization

If X is a process and τ is a stopping time, the stopped process $X_{\wedge\tau}$ is defined by $(X_{\wedge\tau})_t = X_{t\wedge\tau}$. For a given process X it is sometimes possible to find an increasing sequence of stopping times τ_k, with $\tau_k(\omega) \to \infty$ for each ω, such that each stopped process is better behaved than X itself. For example, if B is a standard Brownian motion and τ_k is defined by

$$\tau_k = \inf\{t : |B_t| \ge k\}$$

then the stopped process B_{τ_k} is bounded in absolute value by the constant k.

Typically, one can carry out stochastic calculus operations on suitably stopped processes, with boundedness taking care of various regularity problems, then pass to a limit as k tends to infinity. This technique is called ***localization***. In particular, a process M for which the stopped processes $M_{\wedge\tau_k}$ are all martingales is called a ***local martingale***.

It it possible to develop a satisfyingly complete stochastic calculus for integration with respect to processes (the ***semimartingales***) expressible as the sum of a local martingale and a process whose sample paths are locally of bounded variation, with integrands that are locally bounded,

Stopping times can also be used in place of deterministic grids to control spatial increments of a process. For example, if X is a process with continuous paths, then by defining $\tau_0 = 0$ and

$$\tau_{k+1} = h \wedge \inf\{t \ge \tau_k : |X(t) - X(\tau_k)| > \delta\}$$

we get increments with both $|\tau_{k+1} - \tau_k| \le h$ and $|X(\tau_{k+1}) - X(\tau_k)| \le \delta$. Such properties are very useful if we wish to carry out Taylor expansions with respect to both time and spatial variables.

6. Quadratic variation

For Brownian motion on [0, 1], you know that $\sum_{i=0}^n (\Delta_i B)^2$ converges (in probability) to 1 as the mesh of the underlying grid tends to zero. A similar property holds for all semimartingales. For processes X with continuous sample paths of bounded variation, the quadratic variation

$$\mathbb{Q}(X, \mathbb{G}) = \sum\nolimits_{i=0}^n (\Delta_i X)^2$$

tends (in probability) to zero as $\text{mesh}(\mathbb{G}) \to 0$.

For a martingale M with the property that $M_t^2 - A_t$ is also martingale, for some adapted, increasing process A with continuous sample paths, the quadratic variation over a grid does not vanish in the limit.

We can establish this fact by mimicking the argument from Example <10>. For simplicity, suppose $M_0 = A_0 = 0$. Define an elementary process

$$H_n(t,\omega) = \sum\nolimits_{i=0}^{n} M(t_i,\omega)\mathbb{I}\{t_i < t \le t_{i+1}\}$$

Then

$$\begin{aligned}(M_1^2 - 0^2) - 2H_n \bullet M_1 &= \sum\nolimits_{i=0}^{n} \left(M(t_{i+1})^2 - M(t_i)^2\right) - \sum\nolimits_{i=0}^{n} 2M(t_i)\Delta_i M \\ &= \sum\nolimits_{i=0}^{n} (\Delta_i M)^2.\end{aligned}$$

As the mesh of the grid tends to zero, we force $\sum_{i=0}^{n}(\Delta_i M)^2$ to converge (in what sense?) to $M_1^2 - 2M \bullet M_1$. More generally, the quadratic variation up to time t over the grid, $\sum_{i=0}^{n}\left(M(t \wedge t_{i+1}) - M(t \wedge t_{i+1})\right)^2$, converges to an increasing process

$$[M]_t := M_t^2 - 2M \bullet M_t$$

This process is called the quadratic variation of the martingale.

In fact, for martingales with continuous sample paths, $[M]$ is the same as the increasing process A. To see why $[M]_1 = A_1$, note that the difference $\xi_{i+1} := (\Delta_i M)^2 - \Delta_i A$ depends only on $\mathcal{F}_{t_{i+1}}$-information and $\mathbb{E}_{t_i}(\xi_{i+1}) = 0$. The random variable

$$D_n := \sum\nolimits_{i=0}^{n} (\Delta_i M)^2 - A_1 = \sum\nolimits_{i=1}^{n+1} \xi_i$$

has $\mathbb{E}D_n = 0$ and

$$\mathbb{E}D_n^2 = \sum\nolimits_i \mathbb{E}\xi_i^2 + 2\sum\nolimits_{i<j} \mathbb{E}(\xi_i \xi_j) \le \sum\nolimits_i \mathbb{E}(\Delta_i M)^4 + 0.$$

If the grid is fine enough to make $\max_i |\Delta_i M| \le \epsilon$ (an effect that could be achieved by using stopping times instead of deterministic grid points t_i) the last sum is smaller than

$$\epsilon^2 \mathbb{E} \sum\nolimits_i (\Delta_i M)^2 = \epsilon^2 \mathbb{E} A_1.$$

It follows that D_n converges (in probability) to zero as $\text{mesh}(\mathbb{G}) \to 0$. In the limit we have $[M]_1 - A_1 = 0$.

Similar arguments—starting from elementary H then passing to a limit—gives a result that will be most useful when we develop Ito's lemma:

$$\sum\nolimits_{i=0}^{n} H(t_i)\left(M(t \wedge t_{i+1}) - M(t \wedge t_{i+1})\right)^2 \to H \bullet A_t \qquad \text{in probability}$$

as $\text{mesh}(\mathbb{G}) \to 0$. In fact the convergence is uniform over all t in any bounded interval.