6. Discovering stochastic integral

This section presents calculations that naturally lead to the discovery of the stochastic integral. The purpose of these is to motivate the definition of the stochastic integral while noting the important differences with the ordinary integration theory.

6.1 Quadratic variation.

We have seen that Brownian paths are roughly $1/2 + o(1)$-Hölder regular. This refers to local behavior but our future applications need a way to characterize the regularity of the path as a whole. For this we recall the following concepts:

Definition 6.1 *A* partition Π *of interval* $[0,t]$ *is an ordered collection of points* $0 =: t_0 < t_1 < \dots < t_n := t$. *The* mesh $\|\Pi\|$ *of the partition* Π *is then given by*

$$\|\Pi\| := \max_{i=1,\dots,n} |t_i - t_{i-1}| \tag{6.1}$$

Given $f\colon [0,\infty) \to \mathbb{R}$ *and* $p \geqslant 0$*, the* p*-variation of* f *associated with partition* Π *of interval* $[0,t]$ *is the quantity*

$$V_t^{(p)}(f,\Pi) := \sum_{i=1}^n \bigl|f(t_i) - f(t_{i-1})\bigr|^p \tag{6.2}$$

We are generally interested in the behavior of the p-variation $V_t^{(p)}(f,\Pi)$ as the mesh of Π tends to zero. For $p = 1$ (in fact $p \leqslant 1$), this is aided by the fact that the p-variation increases upon refinements of Π. It thus makes sense to take a supremum over Π which we then refer to (for $p = 1$) as *total variation*. The supremum can be taken for all $p \geqslant 1$, of course; if it turns out to be finite, we say that f is of *bounded p-variation*. (For $p = 1$, this is simply referred to as being of bounded variation.) We call $V^{(2)}(f,\Pi)$ *quadratic variation*.

As is readily checked, uniform Hölder continuity with exponent $\alpha \in (0,1]$ implies boundedness of p-variation for all p with $p\alpha \geqslant 1$. This suggests that $p = 2 + o(1)$ is the relevant case for standard Brownian motion. And, indeed, given a Brownian path $\{B_t\colon t \in [0,\infty)\}$, we have

$$\begin{aligned} EV_t^{(p)}(B,\Pi) &= \sum_{i=1}^n E\bigl(|B_{t_i} - B_{t_{i-1}}|^p\bigr) \\ &= E\bigl(|\mathcal{N}(0,1)|^p\bigr) \sum_{i=1}^n |t_i - t_{i-1}|^{p/2} \leqslant E\bigl(|\mathcal{N}(0,1)|^p\bigr) t \|\Pi\|^{\frac{p-2}{2}} \end{aligned} \tag{6.3}$$

which implies $V_t^{(p)}(B,\Pi_n) \to 0$ in probability and L^1 as $\|\Pi_n\| \to 0$ for all $p > 2$. For $p = 2$ we have

$$EV_t^{(2)}(B,\Pi) = t \tag{6.4}$$

and so $\{V_t^{(2)}(B,\Pi)\colon \Pi = \text{ partition of } [0,t]\}$ is a tight family of random variables. A natural question is whether $V_t^{(2)}(B,\Pi)$ converges weakly along sequences of partitions whose mesh tends to zero. This is addressed in:

Proposition 6.2 *Let $\{B_t \colon t \in [0,\infty)\}$ be a standard Brownian motion. For any sequence of partitions $\{\Pi_n\}_{n\geqslant 1}$ of $[0,t]$, we have:*

(1) $\|\Pi_n\| \to 0$ *implies* $V_t^{(2)}(B,\Pi_n) \to t$ *in probability and* L^2,

(2) $\sum_{n\geqslant 1} \|\Pi_n\| < \infty$ *implies* $V_t^{(2)}(B,\Pi_n) \to t$ *a.s.*

(3) $\{\Pi_n\}_{n\geqslant 1}$ *are nested and* $\|\Pi_n\| \to 0$ *imply* $V_t^{(2)}(B,\Pi_n) \to t$ *a.s.*

Here $\{\Pi_n\}_{n\geqslant 1}$ are said to be nested if Π_{n+1} contains all points of Π_n, for each $n \geqslant 1$.

Proof. Using that the Brownian increments are independent Gaussians, we observe

$$\begin{aligned}\operatorname{Var}\bigl(V_t^{(2)}(B,\Pi)\bigr) &= \sum_{i=1}^n \operatorname{Var}\bigl((B_{t_i}-B_{t_{i-1}})^2\bigr) \\ &= \operatorname{Var}\bigl(\mathcal{N}(0,1)^2\bigr)\sum_{i=1}^n |t_i - t_{i-1}|^2 \leqslant 3t\|\Pi\|\end{aligned} \tag{6.5}$$

In conjunction with (6.4) this gives L^2-convergence $V_t^{(2)}(B,\Pi) \to t$. Invoking the Markov inequality

$$P\bigl(|V_t^{(2)}(B,\Pi) - t| > \epsilon\bigr) \leqslant \frac{1}{\epsilon^2}\operatorname{Var}\bigl(V_t^{(2)}(B,\Pi)\bigr) \leqslant \frac{3t}{\epsilon^2}\|\Pi\| \tag{6.6}$$

we also get convergence in probability as $\|\Pi\| \to 0$, thus proving (1). For (2) we note that the summability of the mesh-sequence implies summability of the probabilities on the left. Then a.s. convergence $V_t^{(2)}(B,\Pi) \to t$ then follows via a standard argument based on the Borel-Cantelli lemma.

The convergence in (3) will be inferred from martingale convergence. We need a lemma whose statement is more scary that the proof:

Lemma 6.3 *Let $Z_1,\dots,Z_n \in L^2$ be independent with symmetric law, $\forall i \leqslant n \colon Z_i \overset{\text{law}}{=} -Z_i$, and let $I_1,\dots,I_m$ be non-empty disjoint sets such that $\{1,\dots,n\} = \bigcup_{j=1}^m I_j$. Then for any σ-algebra $\mathcal{G}$ satisfying*

$$\sigma\biggl(\Bigl(\sum_{i\in I_j} Z_i\Bigr)^2 \colon j = 1,\dots,m\biggr) \subseteq \mathcal{G} \subseteq \sigma\biggl(\Bigl(\sum_{i\in I} Z_i\Bigr)^2 \colon\ I \subseteq I_j \text{ for some } j \in \{1,\dots,m\}\biggr) \tag{6.7}$$

we have

$$E\biggl(\Bigl(\sum_{i=1}^n Z_i\Bigr)^2 \biggm| \mathcal{G}\biggr) = \sum_{j=1}^m \Bigl(\sum_{i\in I_j} Z_i\Bigr)^2 \tag{6.8}$$

Proof. Left to homework. □

We now pick a sequence of nested partitions $\{\Pi_n\}_{n\geqslant 1}$ where the points of Π_n will be denoted as $\{t_i^n\}_{i=0}^{k(n)}$. For $n \leqslant N$ set

$$\mathcal{G}_{n,N} := \sigma\Bigl((B_{t_i^r} - B_{t_{i-1}^r})^2 \colon i = 1,\dots,k(r),\ r = n,\dots,N\Bigr) \tag{6.9}$$

Lemma 6.3 then implies

$$V_t^{(2)}(f,\Pi_n) = E(B_t^2 \,|\, \mathcal{G}_{n,N}) \tag{6.10}$$

Taking $N \to \infty$ with the help of the Levy Forward Theorem gives

$$V_t^{(2)}(f,\Pi_n) = E(B_t^2 \,|\, \mathcal{G}_n) \tag{6.11}$$

where

$$\mathcal{G}_n := \sigma\Big(\bigcup_{N\geqslant n} \mathcal{G}_{n,N}\Big) = \sigma\Big((B_{t_i^r} - B_{t_{i-1}^r})^2 \colon i = 1,\dots,k(r),\, r \geqslant n\Big) \tag{6.12}$$

Lévy Backward Theorem then gives a.s. convergence of $V_t(f,\Pi_n)$ as $n \to \infty$. (Note that none of these require that the mesh of the partition tends to zero.) In light of part (1) of the claim, if $\|\Pi_n\| \to 0$ the a.s. limit must equal t, as desired. □

A natural follow-up question is what happens with the corresponding quadratic variation of functions of the Brownian path. We start by looking at $V^{(2)}(B^2,\Pi)$. A calculation shows

$$\begin{aligned} V_t^{(2)}(B^2,\Pi) &= \sum_{i=1}^n (B_{t_i}^2 - B_{t_{i-1}}^2)^2 \\ &= \sum_{i=1}^n 4B_{t_{i-1}}^2 (B_{t_i} - B_{t_{i-1}})^2 + \sum_{i=1}^n 4B_{t_{i-1}} (B_{t_i} - B_{t_{i-1}})^3 + \sum_{i=1}^n (B_{t_i} - B_{t_{i-1}})^4 \end{aligned} \tag{6.13}$$

The reason for separating terms this way is that, by the defining properties of standard Brownian motion, $B_{t_{i-1}}$ is independent of $B_{t_i} - B_{t_{i-1}}$, and in fact of $B_{t_k} - B_{t_{k-1}}$ for all $k \geqslant i$. This allows us to estimate the L^1-norm of the second term on the right by

$$E|B_t| \sum_{i=1}^n E\big(|B_{t_i} - B_{t_{i-1}}|^3\big) \leqslant E|B_t|\, t\, E\big(|\mathcal{N}(0,1)|^3\big)\, \|\Pi\|^{1/2} \tag{6.14}$$

The L^1-norm of the last term on the right of (6.13) is in turn bounded by a constant times $\|\Pi\|$. These terms thus tend to zero in probability as $\|\Pi\| \to 0$. For the first term we in turn write

$$\sum_{i=1}^n 4B_{t_{i-1}}^2 (B_{t_i} - B_{t_{i-1}})^2 = \sum_{i=1}^n 4B_{t_{i-1}}^2 |t_i - t_{i-1}| + \sum_{i=1}^n 4B_{t_{i-1}}^2 \Big[(B_{t_i} - B_{t_{i-1}})^2 - |t_i - t_{i-1}|\Big] \tag{6.15}$$

The second term on the right has mean zero and variance equal to

$$\begin{aligned} &\sum_{i=1}^n 16 E\big(B_{t_{i-1}}^4\big) \operatorname{Var}\big((B_{t_i} - B_{t_{i-1}})^2\big) \\ &\qquad\qquad \leqslant 16 t^4\, E\big(|\mathcal{N}(0,1)^4\big)\, \operatorname{Var}\big(\mathcal{N}(0,1)^2\big) \sum_{i=1}^n |t_i - t_{i-1}|^2 \end{aligned} \tag{6.16}$$

where the bound on the right uses the scaling property of Brownian increments. The quantity on the right is order $\|\Pi\|$ and thus decays to zero as $\|\Pi\| \to 0$. Since the first term on the right of (6.15) is the Riemann sum, we have proved that, for any sequence $\{\Pi_n\}_{n\geqslant 1}$ of partitions of $[0,t]$,

$$\|\Pi_n\|\to 0 \quad\Rightarrow\quad V_t^{(2)}(B^2,\Pi_n) \xrightarrow[n\to\infty]{P} \int_0^t 4B_s^2\,\mathrm{d}s \tag{6.17}$$

More interesting that this conclusion is the mechanics of the underlying calculation. The key point is to arrange terms to isolate increments over non-overlapping intervals. The Brownian increments are handled according to the infinitesimal-calculus "rules"

$$(\mathrm{d}B_t)^2 = \mathrm{d}t \ \wedge\ \mathrm{d}B_t\,\mathrm{d}t = 0\ \wedge (\mathrm{d}t)^2 = 0 \tag{6.18}$$

Using these, we readily generalize the above calculations to the form:

Lemma 6.4 *Let $\{B_t\colon t\in[0,\infty)\}$ be standard Brownian motion. Then for any $f\in C^1(\mathbb{R})$, any $t\geqslant 0$ and any sequence $\{\Pi_n\}_{n\geqslant 1}$ of partitions of $[0,t]$,*

$$\|\Pi_n\|\to 0 \quad\Rightarrow\quad V_t^{(2)}(f\circ B,\Pi_n) \xrightarrow[n\to\infty]{P} \int_0^t f'(B_s)^2\,\mathrm{d}s \tag{6.19}$$

We leave the proof of this lemma to homework. Note that, unlike for the linear case treated in Proposition 6.2, the limiting *quadratic-variation process* is generally random.

6.2 Itô integral from Itô formula.

Building on the previous calculations, we will now explore the process $t\mapsto f(B_t)$ further. Fix a partition $\Pi=\{0=t_0<t_1<\dots<t_n=t\}$ of $[0,t]$. Assuming $f\in C^2(\mathbb{R})$, Taylor's Theorem gives

$$\begin{aligned} f(B_t)-f(B_0) &= \sum_{i=1}^n \bigl[f(B_{t_i})-f(B_{t_{i-1}})\bigr] \\ &= \sum_{i=1}^n f'(B_{t_{i-1}})(B_{t_i}-B_{t_{i-1}}) + \frac12\sum_{i=1}^n f''(B_{t_{i-1}})(B_{t_i}-B_{t_{i-1}})^2 \\ &\quad + \sum_{i=1}^n\biggl(\int_0^1\Bigl[f''(\theta B_{t_{i-1}}+(1-\theta)B_{t_i})-f''(B_{t_{i-1}})\Bigr](1-\theta)\mathrm{d}\theta\biggr)(B_{t_i}-B_{t_{i-1}})^2 \end{aligned} \tag{6.20}$$

Denoting by

$$\operatorname{osc}_f(A,\delta) := \sup\bigl\{|f(x)-f(x')|\colon x,x'\in A\ \wedge\ |x-x'|<\delta\bigr\} \tag{6.21}$$

the *oscillation* of f on A over distances less than δ, the term in the square brackets under the integral in (6.20) can be bounded by

$$\operatorname{osc}_{f''}\Bigl(B([0,t]),\operatorname{osc}_B\bigl([0,t],\|\Pi\|\bigr)\Bigr) \tag{6.22}$$

In light of continuity of B, the inner oscillation tends to zero as $\|\Pi\|\to 0$. The outer oscillation then does as well since f'' is continuous and $B([0,t])$ is compact. Since the last term in (6.20) is at most the product of (6.22) and $V_t^{(2)}(B,\Pi)$, which is tight, this term tends to zero in probability as $\|\Pi\|\to 0$.

The sum in the second term on the right of (6.20) is treated as in (6.15) thus replacing it with the Riemann sum

$$\sum_{i=1}^{n} f''(B_{t_{i-1}})|t_i - t_{i-1}| \tag{6.23}$$

In light of the continuity of $f'' \circ B$, the sum converges to the Riemann integral $\int_0^t f''(B_s)\mathrm{d}s$. Since the left-hand side of (6.20) does not depend on Π, also the first term on the right must converge in probability. This proves:

Lemma 6.5 *Given $f\colon \mathbb{R} \to \mathbb{R}$ and a sample $\{B_t\colon t \in [0,\infty)\}$ of standard Brownian motion, for each partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ of $[0,t]$ set*

$$I_t(f,\Pi) := \sum_{i=1}^{n} f(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) \tag{6.24}$$

Then for each $f \in C^2(\mathbb{R})$ and each $t \geqslant 0$ there exists a random variable $I_t(f')$ such that for any sequence of partitions $\{\Pi_n\}_{n\geqslant 1}$ as above,

$$\|\Pi_n\| \to 0 \quad \Rightarrow \quad I_t(f',\Pi_n) \xrightarrow[n\to\infty]{P} I_t(f') \tag{6.25}$$

Moreover, we then have

$$f(B_t) = f(B_0) + I_t(f') + \frac{1}{2}\int_0^t f''(B_s)\mathrm{d}s \tag{6.26}$$

where the integral on the right is a Riemann integral.

The limiting object is called the *Itô integral* after K. Itô, who developed these ideas in early 1940s in wartime Japan. To make a closer connection with (6.23), we write this in the integral notation as

$$\int_0^t f(B_s)\mathrm{d}B_s := I_t(f) \tag{6.27}$$

The identity (6.26) is then known as the *Itô formula*. Note that the Fundamental Theorem of Calculus does not hold with this integral due to the appearance of the second derivative term — referred to as the *Itô term*.

Replacing f by its antiderivative, the above lemma asserts the convergence $I_t(f,\Pi) \to I_t(f)$ as $\|\Pi\| \to 0$ for all $f \in C^1(\mathbb{R})$. Requiring continuous differentiability is actually an overkill; just plain continuity and boundedness suffice. To prove this, we note:

Lemma 6.6 *For all $t \geqslant 0$, all bounded functions $f \in C(\mathbb{R})$ and all partitions $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ of $[0,t]$,*

$$E\bigl(I_t(f,\Pi)^2\bigr) = E\biggl(\sum_{i=1}^{n} f(B_{t_{i-1}})^2|t_i - t_{i-1}|\biggr). \tag{6.28}$$

Proof. Write the square of $I_t(f,\Pi)$ as a double sum and notice that, as before, all but the diagonal terms vanish under expectation. □

Hereby we get:

Corollary 6.7 *For all bounded $f \in C(\mathbb{R})$ and $t \geqslant 0$ there exists a random variable $I_t(f)$ such that for any sequence of partitions $\{\Pi_n\}_{n\geqslant 1}$ as above,*

$$\|\Pi_n\| \to 0 \quad \Rightarrow \quad I_t(f,\Pi_n) \xrightarrow[n\to\infty]{P} I_t(f) \tag{6.29}$$

Proof. Given $\epsilon > 0$, find $M > 0$ such that

$$P\Big(\sup_{0\leqslant u\leqslant t} |B_u| > M\Big) < \epsilon. \tag{6.30}$$

(This is possible because the Brownian motion is continuous and thus locally bounded.) For each $f \in C(\mathbb{R})$ we then find $f_\epsilon \in C^1(\mathbb{R})$ such that

$$\sup_{x\in[-M,M]} \big|f(x) - f_\epsilon(x)\big| < \sqrt{\epsilon}\,\|f\| \tag{6.31}$$

and

$$\|f_\epsilon\| \leqslant 2\|f\| \tag{6.32}$$

where $\|\cdot\|$ denotes the supremum norm over all of $\mathbb{R}$. (We are not justifying this step as that is standard in analysis.) From (6.28) we then get

$$\begin{aligned} E\Big(\big|I_t(f,\Pi) - I_t(f_\epsilon,\Pi)\big|^2\Big) &= E\big(I_t(f-f_\epsilon,\Pi)^2\big) \\ &= E\bigg(\sum_{i=1}^n (f-f_\epsilon)(B_{t_{i-1}})^2 |t_i - t_{i-1}|\bigg) \\ &\leqslant \|f-f_\epsilon\|^2 t P\Big(\sup_{0\leqslant u\leqslant t}|B_u| > M\Big) + t \sup_{x\in[-M,M]} \big|f(x) - f_\epsilon(x)\big|^2 \end{aligned} \tag{6.33}$$

By our choices above, there exists a constant $C > 0$ depending only on f and t such that the right-hand side is at most $C\epsilon$, regardless of Π.

The Chebyshev inequality converts the above to

$$P\Big(\big|I_t(f,\Pi) - I_t(f_\epsilon,\Pi)\big| > \delta\Big) \leqslant \frac{C}{\delta^2}\epsilon \tag{6.34}$$

We will now perform a variation on the 3ϵ argument. First note that, for any $0 < \tilde{\epsilon} < \epsilon$ the above shows

$$P\Big(\big|I_t(f_\epsilon,\Pi) - I_t(f_{\tilde{\epsilon}},\Pi)\big| > \delta\Big) \leqslant \frac{2C}{\delta^2}\epsilon \tag{6.35}$$

Given a sequence of partitions $\{\Pi_n\}_{n\geqslant 1}$ with $\|\Pi_n\| \to 0$, we know that $I_t(f_\epsilon,\Pi_n) \to I_t(f_\epsilon)$ in probability for each $\epsilon > 0$. With the help of Fatou's lemma (which works for convergence in probability by way of representing it as a.s. convergence along a subsequence), the above then gives

$$P\Big(\big|I_t(f_\epsilon) - I_t(f_{\tilde{\epsilon}})\big| > \delta\Big) \leqslant \frac{2C}{\delta^2}\epsilon \tag{6.36}$$

Setting $\epsilon_n := 8^{-n}$, $\tilde{\epsilon} := 8^{-n-1}$ and $\delta := 2^{-n}$ then shows

$$P\Big(\big|I_t(f_{\epsilon_n}) - I_t(f_{\epsilon_{n+1}})\big| > 2^{-n}\Big) \leqslant 2C2^{-n}. \tag{6.37}$$

This is summable on $n \geqslant 1$ and so, by Borel-Cantelli, the event occurs only for finitely many n a.s. It follows that $I_t(f) := \lim_{m\to\infty} I_t(f_{\epsilon_m})$ exists.

To conclude the desired convergence from this, observe that

$$\begin{aligned}P\Bigl(\bigl|I_t(f,\Pi_n) - I_t(f)\bigr| > 3\delta\Bigr) &\leqslant P\Bigl(\bigl|I_t(f,\Pi_n) - I_t(f_{\epsilon_m},\Pi_n)\bigr| > \delta\Bigr)\\ &\quad + P\Bigl(\bigl|I_t(f_{\epsilon_m},\Pi_n) - I_t(f_{\epsilon_m})\bigr| > \delta\Bigr) + P\Bigl(\bigl|I_t(f_{\epsilon_m}) - I_t(f)\bigr| > \delta\Bigr)\end{aligned} \tag{6.38}$$

Taking $n \to \infty$ followed by $m \to \infty$ then makes all three terms on the right vanish. Since δ was arbitrary, the claim follows. □

As we shall see later, the statement of Corollary 6.7 is not the end of the story as the stated limit exists in considerably larger generality. But the mechanics of the proof is based on the same idea; namely, an extension via the Itô isometry. This will in fact be the principal tool in our treatment of the general stochastic integral later.

We finish this section by noting that the stochastic integral $f \mapsto I_t(f)$, although akin to the Riemann-Stieltjes integral, does not exist in the ordinary sense. The easiest way to see this is to replace the "left endpoint rule" by a different rule (which does not make a difference when the integral exists in classical sense). The following lemma will be relegated to a homework exercise:

Lemma 6.8 *Fix $\theta \in [0,1]$ and let $\{B_t\colon t \in [0,\infty)\}$ be a standard Brownian motion. For any $f\colon \mathbb{R} \to \mathbb{R}$, $t \geqslant 0$ and a partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ of $[0,t]$ set*

$$I_t^{(\theta)}(f,\Pi) := \sum_{i=1}^n f\bigl(B_{(1-\theta)t_{i-1}+\theta t_i}\bigr)(B_{t_i} - B_{t_{i-1}}) \tag{6.39}$$

Assuming $f \in C^1(\mathbb{R})$ with f, f' bounded, for any sequence of partitions $\{\Pi_n\}_{n\geqslant 1}$, we then have

$$\|\Pi_n\| \to 0 \quad \Rightarrow \quad I_t^{(\theta)}(f,\Pi_n) \underset{n\to\infty}{\overset{P}{\longrightarrow}} \int_0^t f(B_s)\mathrm{d}B_s + \theta\int_0^t f'(B_s)\mathrm{d}s \tag{6.40}$$

Here the first integral on the right is as in (6.27)*.*

For the "mid point rule", $\theta := 1/2$, the limit object on the right is called the *Stratonovich integral*. This integral can be defined directly from the Itô integral as

$$\int_0^t f(B_s)\circ \mathrm{d}B_s := \int_0^t f(B_s)\mathrm{d}B_s + \frac{1}{2}\int_0^t f'(B_s)\mathrm{d}s \tag{6.41}$$

The choice $\theta = 1/2$ ensures that the Fundamental Theorem of Calculus is restored,

$$f(B_t) - f(B_0) = \int_0^t f(B_s)\circ\mathrm{d}B_s, \tag{6.42}$$

thus making the Stratonovich integral attractive in certain situations (such as analysis on manifolds) where the Itô term wrecks havoc in calculations. The Stratonovich integral also arises naturally in solutions of stochastic differential equations obtained by approximating standard Brownian motion by smoothed-out versions thereof. (Indeed, the FTC holds in these and thus survives the limit as all approximations are taken away.)

Still, the Stratonovich integral is not good for theory development and so we will work mostly with the Itô integral in the sequel.

Further reading: Chapter 3 of Øksendal