Chapter 1

Brownian Motion

In these notes we build the objects the rest of the book computes with. Filtrations and martingales first, then the Itô integral — which is defined by a limit that is not obviously there, so we show why it exists rather than assuming it, and find that the Itô isometry is what supplies the limit rather than merely describing it. Along the way the Stratonovich integral appears as the alternative that obeys the ordinary chain rule and is not adapted, which is the trade between geometry and predictability. The chapter ends with Itô's lemma and the diffusions it applies to.

1.1 Stochastic Processes

Definition 1.1 (Brownian Motion)**.** A brownian motion is a stochastic process $W_{t\geq 0}$ such that the following properties hold:

- $W_0 = 0$ with probability 1.
- $W_a - W_b \sim N(0, |a-b|)$ where $N(m, v)$ is a Normal distribution with mean m and variance v.
- $W_a - W_b$ and $W_c - W_d$ are independent random variables for $a > b \geq c > d$.

So the process starts at zero, moves by independent normal steps, and a step's variance is exactly its length — which is why the path accumulates randomness at a constant, known rate.

Definition 1.2 (σ-Algebra)**.** A σ-algebra on a set X is a collection Σ of subsets of X such that the following properties hold:

- $X \in \Sigma$.
- $A \in \Sigma \Rightarrow A^c \in \Sigma$ (closed under complement)
- $A_i \in \Sigma \Rightarrow \bigcup_i A_i \in \Sigma$ (closed under countable union)

A probability is a map $\Sigma \to \mathbb{R}$ assigning a number to each event in Σ — so the σ-algebra is exactly the collection of events whose probability is known, and the closure axioms say that knowing an event's probability means knowing its complement's and any countable combination's as well.

Definition 1.3 (Filtration)**.** A filtration is an indexed σ-algebra $\mathcal{F}_t$ of subsets of the probability sample space such that $F_a \supseteq F_b$ for $a \geq b$.

So $\mathcal{F}_t$ is the information available at time t — a finer σ-algebra distinguishes more events — and the nesting condition says that information only grows: nothing is forgotten as t increases.

Definition 1.4 (Measurable Function)**.** Given a σ-algebra $\mathcal{F}$ on the sample space Ω, a real-valued random variable X is $\mathcal{F}$-measurable if the X-inverse of any open set in $\mathbb{R}$ is contained in $\mathcal{F}$ i.e.

$$\{A \in \Omega | X(A) \in (a, b)\} \in \mathcal{F} \qquad \forall a, b \in \mathbb{R} \quad a > b$$

So observing the value of a measurable function identifies an event already in $\mathcal{F}$ — the set of outcomes consistent with that observation — which is what lets $\mathcal{F}$ "know" the function's value.

Definition 1.5 (Adapted Process)**.** A stochastic process X_t is adapted to filtration $\mathcal{F}_t$ if for each t, X_t is F_t measurable.

So a process adapted to $\mathcal{F}_t$ has a value at time t already implied by the information available then — nothing about X_t is still hidden from $\mathcal{F}_t$.

Definition 1.6 (Generated Filtration)**.** The filtration generated by a stochastic process is the smallest filtration to which the process is adapted.

Example 1.1. Let's take an example of 3 unbiased coin tosses with process X_t counting the number of heads until toss t. The probability sample space Ω is

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

The filtration generated by the process is

$$
\begin{aligned}
\mathcal{F}_0 =\{ & \\
& \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}, \quad (X_0 = 0) \\
& \emptyset \\
\} & \\
\mathcal{F}_1 =\{ & \\
& \{HHH, HHT, HTH, HTT\}, \quad (X_0 = 0, X_1 = 1) \\
& \{THH, THT, TTH, TTT\}, \quad (X_0 = 0, X_1 = 0) \\
& \emptyset \\
& \text{and all unions of the above} \\
\} & \\
\mathcal{F}_2 =\{ & \\
& \{HHH, HHT\}, \quad (X_0 = 0, X_1 = 1, X_2 = 2) \\
& \{HTH, HTT\}, \quad (X_0 = 0, X_1 = 1, X_2 = 1) \\
& \{THH, THT\}, \quad (X_0 = 0, X_1 = 0, X_2 = 1) \\
& \{TTH, TTT\}, \quad (X_0 = 0, X_1 = 0, X_2 = 0) \\
& \emptyset \\
& \text{and all unions of the above} \\
\} & \\
\mathcal{F}_3 =\{ & \\
& \{HHH\}, \quad (X_0 = 0, X_1 = 1, X_2 = 2, X_3 = 3) \\
& \{HHT\}, \quad (X_0 = 0, X_1 = 1, X_2 = 2, X_3 = 2) \\
& \{HTH\}, \quad (X_0 = 0, X_1 = 1, X_2 = 1, X_3 = 2) \\
& \{HTT\}, \quad (X_0 = 0, X_1 = 1, X_2 = 1, X_3 = 1) \\
& \{THH\}, \quad (X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 2) \\
& \{THT\}, \quad (X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 1) \\
& \{TTH\}. \quad (X_0 = 0, X_1 = 0, X_2 = 0, X_3 = 1) \\
& \{TTT\}, \quad (X_0 = 0, X_1 = 0, X_2 = 0, X_3 = 0) \\
& \emptyset \\
& \text{and all unions of the above} \\
\} &
\end{aligned}
$$

Definition 1.7 (Expectation conditional on filtration)**.** $\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}]$ is the $\mathcal{F}$ measurable random variable such that

$$\mathbb{E}^{\mathbb{P}}\left[X\,\mathbf{1}_A\right] = \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}[X|\mathcal{F}]\,\mathbf{1}_A\right] \quad \forall A \in \mathcal{F} \tag{1.1}$$

So averaging X over any event $\mathcal{F}$ can see gives the same answer whether or not X is first replaced by its conditional expectation — $\mathbb{E}[X|\mathcal{F}]$ is the $\mathcal{F}$-measurable random variable that matches X's average on every event $\mathcal{F}$ can distinguish, even though it need not match X itself.

In the coin toss example above,

$$\mathbb{E}[X_3|\mathcal{F}_1](s) = \begin{cases} 2 & \text{if } s \in \{HHH, HHT, HTH, HTT\} \\ 1 & \text{if } s \in \{THH, THT, TTH, TTT\} \end{cases}$$

Definition 1.8 (Martingale)**.** A stochastic process X_t is a martingale with respect to a filtration $\mathcal{F}_t$ of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$\mathbb{E}^{\mathbb{P}}[X_a|\mathcal{F}_b] = X_b \quad \text{for } a \geq b$$

So a martingale's expected future value, conditional on the present, is the present value — no drift that today's information can see coming.

Exercise. Show that in the coin toss example, if we had instead counted number of heads minus number of tails, X_t would have been a martingale. For example, show that $\mathbb{E}[X_3|\mathcal{F}_1]$ would have been equal to X_1.

Relaxing the equality in either direction gives the two processes that drift one way on average.

Definition 1.9 (Supermartingale and submartingale)**.** With the same setup, X_t is a *supermartingale* if

$$\mathbb{E}^{\mathbb{P}}[X_a|\mathcal{F}_b] \leq X_b \quad \text{for } a \geq b,$$

and a *submartingale* if the inequality runs the other way. A martingale is both.

The names are the wrong way round for anyone reading them as a description of the paths: a supermartingale is the one that decreases in expectation. They come from potential theory, where a function u is superharmonic when its value at a point is at least its average over any surrounding sphere. Composing such a u with a Brownian motion averages it over exactly those spheres, so $u(W_t)$ is a supermartingale precisely when u is superharmonic, and the two words describe the same inequality. In the language of chapter 3 the condition reads $\mathcal{L}u \leq 0$: a non-positive drift is what a supermartingale has.

Definition 1.10 (Stopping time)**.** A random time $\tau : \Omega \to [0, \infty]$ is a *stopping time* for the filtration $\mathcal{F}_t$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for every t. Whether it has already happened must be decidable from what is known at t, without looking ahead.

The first time a Brownian motion reaches a level, $\tau_a = \inf\{t : W_t \geq a\}$, is a stopping time — at any t you know whether it has happened. The *last* time before T that the path visits a is not, because deciding it at time t requires knowing that the path never returns. Every trading rule that can actually be followed is measurable in the first sense, which is why the condition appears wherever a strategy or an exercise decision does.

Given a stopping time, $X_t^\tau = X_{t\wedge\tau}$ is the process *stopped* at τ: it follows X until τ and is constant thereafter. Stopping preserves the martingale property, and it is the standard way to tame a process that is badly behaved only in the limit — the definition of a local martingale below is exactly that idea.

Theorem 1.11 (Optional stopping)**.** *Let X be a martingale and τ a stopping time. If τ is bounded, or if $\{X_{t\wedge\tau}\}$ is uniformly integrable, then $\mathbb{E}[X_\tau] = X_0$.*

Proof. Take first a τ with finitely many values $t_1 < \cdots < t_n = T$. The whole argument is that the event $\{\tau = t_i\}$ is known by time t_i, so the martingale property may be used on it:

$$\begin{aligned}
\mathbb{E}[X_T] &= \sum_{i=1}^{n} \mathbb{E}\big[X_T \mathbf{1}_{\{\tau=t_i\}}\big] \\
&= \sum_{i=1}^{n} \mathbb{E}\Big[\mathbb{E}[X_T \mid \mathcal{F}_{t_i}]\, \mathbf{1}_{\{\tau=t_i\}}\Big] && \text{since } \{\tau = t_i\} \in \mathcal{F}_{t_i} \\
&= \sum_{i=1}^{n} \mathbb{E}\big[X_{t_i} \mathbf{1}_{\{\tau=t_i\}}\big] && \text{the martingale property} \\
&= \mathbb{E}[X_\tau],
\end{aligned}$$

and $\mathbb{E}[X_T] = X_0$ because X is a martingale. Notice what did the work. The indicator could be taken inside the conditional expectation only because it was $\mathcal{F}_{t_i}$ measurable, and that is the definition of a stopping time — the only property of τ the calculation used. A random time that looked ahead would break the second line and nothing else.

A general bounded $\tau \le T$ is a limit of such times. Put $\tau_n = \big(\lceil 2^n \tau \rceil / 2^n\big) \wedge T$, which takes finitely many values, is a stopping time because $\{\tau_n \le k/2^n\} = \{\tau \le k/2^n\}$, and decreases to τ. Each $\mathbb{E}[X_{\tau_n}] = X_0$ by the case just proved, and $X_{\tau_n} \to X_\tau$ almost surely by continuity of the path. The same case also identifies $X_{\tau_n} = \mathbb{E}[X_T \mid \mathcal{F}_{\tau_n}]$, and a family of conditional expectations of one integrable variable is uniformly integrable, so the convergence holds in L^1 and the expectations pass to the limit.

In the unbounded case the hypothesis does exactly the job boundedness did. The stopped process is a martingale, so $\mathbb{E}[X_{t\wedge\tau}] = X_0$ for every finite t, and $X_{t\wedge\tau} \to X_\tau$ almost surely as $t \to \infty$ on $\{\tau < \infty\}$; uniform integrability is what upgrades that to L^1 and lets the expectation follow the limit. Without it the two cannot be exchanged, and the next paragraph is a case where they cannot. □

A finite stopping time is not enough on its own, and the gap is not a technicality. The doubling strategy of chapter 4 stops a Brownian motion at the first time it reaches \$1, which happens almost surely, and collects $\mathbb{E}[X_\tau] = 1$ from a process starting at zero. What fails is uniform integrability: the paths that have not yet reached the target are ever further below it, and the loss they carry is exactly what the expectation is missing. Ruling that out is the work done by admissibility.

Stopping also supplies the weakest useful version of the martingale property.

Definition 1.12 (Local Martingale)**.** A process X is a local martingale if there are stopping times $\tau_n \uparrow \infty$ such that each stopped process $X_{t\wedge\tau_n}$ is a martingale.

"Locally a martingale" says the process is a martingale up to a sequence of stopping times that eventually exhausts the horizon, which for many processes of interest is all one can guarantee. The gap between local and true is the same gap optional stopping left open, and it is not empty: the doubling strategy has a local martingale whose expectation moves. When the process is bounded below, though, half of the property survives.

Lemma 1.13 (A non-negative local martingale is a supermartingale)**.** *If $X \ge 0$ is a local martingale then $\mathbb{E}[X_a|\mathcal{F}_b] \le X_b$ for $a \ge b$.*

Proof. Let $\tau_n \uparrow \infty$ reduce X, so that $\mathbb{E}[X_{a\wedge\tau_n}|\mathcal{F}_b] = X_{b\wedge\tau_n}$. As $n \to \infty$ the integrands converge almost surely and are non-negative, so conditional Fatou gives

$$\mathbb{E}[X_a|\mathcal{F}_b] \leq \liminf_n \mathbb{E}[X_{a\wedge\tau_n}|\mathcal{F}_b] = X_b.$$

Fatou runs in one direction only, and that is where the inequality comes from. □

So a non-negative local martingale can lose expectation but never gain it, and it is a true martingale exactly when it loses none. chapter 4 is about which side of that line a discounted price has to be on.

Definition 1.14 ($\mathcal{F}_t$ Brownian motion)**.** A Brownian motion W_t is a $\mathcal{F}_t$ Brownian motion if W_t is adapted to $\mathcal{F}_t$ and $W_{t+s} - W_t$ is independent of $\mathcal{F}_t$ for $s > 0$.

Remark (An $\mathcal{F}_t$ Brownian motion is a martingale with respect to filtration $\mathcal{F}_t$)**.**

$$\begin{aligned}
\mathbb{E}[W_a|\mathcal{F}_b] &= \mathbb{E}[W_a - W_b|\mathcal{F}_b] + \mathbb{E}[W_b|\mathcal{F}_b] \\
&= 0 + \mathbb{E}[W_b|\mathcal{F}_b] \\
&= W_b
\end{aligned}$$

1.2 Itô Calculus

Definition 1.15 (Itô Integral)**.** Let H_t be a $\mathcal{F}_t$ adapted process and W_t be a $\mathcal{F}_t$ Brownian motion. The Itô integral is defined as

$$\int_0^t H_s dW_s = \lim_{n\to\infty} \sum_{i=0}^{2^n-1} H_{t_i^n} \Delta_i W$$

This is similar to the Riemann integral. One thing to note here is that the integrand always takes the value at the start of the interval. If we make the integrand take the the average of the values at interval endpoints, we get

Definition 1.16 (Stratonovich Integral)**.**

$$\int_0^t H_s \circ dW_s = \lim_{n\to\infty} \sum_{i=0}^{2^n-1} (H_{t_i^n} + H_{t_{i+1}^n})\Delta_i W/2 \tag{1.2}$$

The two integrals usually do not give the same results, as can be checked by repeating the calculations we do in these notes for Stratonovich integral as well.

This is the first place stochastic calculus parts company with ordinary calculus, and the reason is visible in a single algebraic step. Take $H_s = W_s$, so both sums are attempts at $\int_0^t W dW$, and look at one interval. The Stratonovich term exceeds the Itô term by

$$\tfrac{1}{2}(W_{i+1} + W_i)(W_{i+1} - W_i) - W_i(W_{i+1} - W_i) = \tfrac{1}{2}(W_{i+1} - W_i)^2,$$

so summing over the partition, the gap between the two integrals is exactly half the sum of squared increments. For an ordinary differentiable function that sum vanishes as the partition refines, and the two definitions agree — which is why nobody distinguishes them in ordinary calculus. For Brownian motion it does not vanish. It converges to t.

Structure (Which integral is the geometric one)**.** The two integrals differ by more than a convention.

Stratonovich obeys the ordinary chain rule: $d\,\varphi(X) = \varphi'(X) \circ dX$, with no second order term. So a Stratonovich equation transforms correctly under a change of coordinates — write the same process in a new variable and the equation you get is the one you would have written directly. Itô does not: its correction term $\frac{1}{2}\varphi''\sigma^2 dt$ depends on the coordinate chosen, so the same process has different-looking Itô equations in different variables.

That makes Stratonovich the integral of choice whenever geometry is involved — diffusions on a sphere, on a manifold of correlation matrices, on a Lie group — because there is no privileged coordinate for the equation to be written in. It makes Itô the integral of choice in finance, where there is a privileged coordinate, namely money, and where the payoff of a martingale being a martingale is worth more than coordinate freedom.

Both are used later. The point here is only that the choice is between invariance and the martingale property, and one cannot have both.

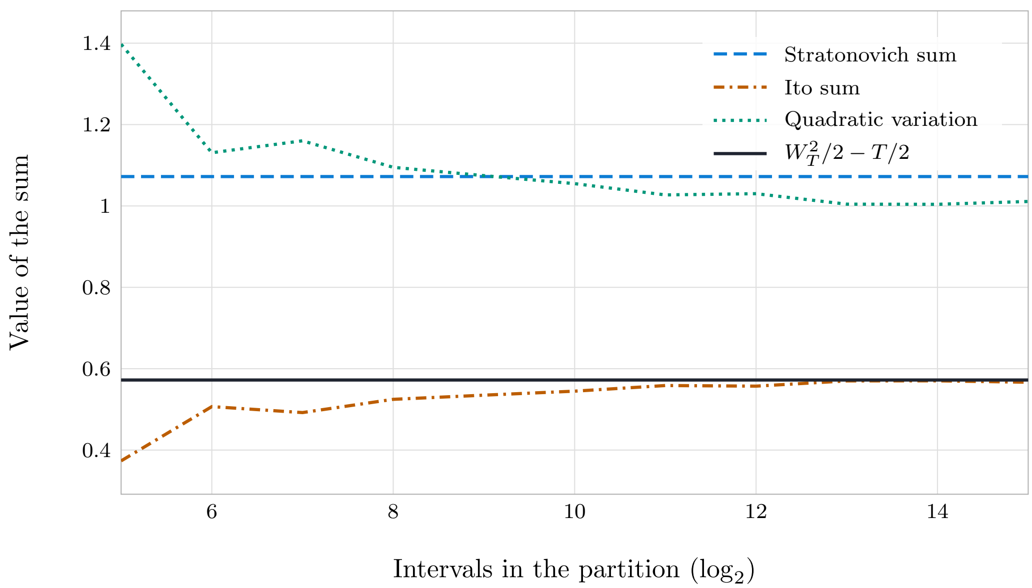

Figure 1.1: The two sums for $\int_0^t W\,dW$, computed on a single Brownian path that is refined rather than redrawn, so that what changes across the picture is the definition and not the randomness. The Stratonovich sum is exactly $W_t^2/2$ at every resolution, because it telescopes — it obeys the ordinary chain rule. The Itô sum sits half the quadratic variation below it and converges to $W_t^2/2 - t/2$. The extra $-t/2$ is not an approximation error; it is what Itô's lemma is about.
Drawn by `sums_over_path` and `brownian_path`.

Two things in the figure stand out. The Stratonovich sum is not converging to anything — it is already exactly right at thirty-two intervals, and at every coarser and finer resolution, because $\frac{1}{2}(W_{i+1} + W_i)(W_{i+1} - W_i) = \frac{1}{2}(W_{i+1}^2 - W_i^2)$ and the sum collapses. And the quadratic variation converges to t rather than to zero, which is the single fact that makes all of this necessary.

In finance the integrand is usually the quantities of assets in the portfolio and the stochastic process the price of these assets. When re-adjusting the portfolio for a specific time interval we can only use the information available at the beginning of the time interval and therefore using Itô integral would be more appropriate.

Let's compute expectations and variances of some of the basic Itô Integrals. We assume that the integrand is square-integrable i.e. $\int_0^t \mathbb{E}[H_s^2]ds < \infty$. In this case we have

The calculations below are all sums over the dyadic partition of $[0,t]$, and written out in full they do not fit on a page. Two abbreviations fix that and make the structure easier to see. Write

$$t_i^n = \frac{i\,t}{2^n}, \qquad \Delta_i W = W_{t_{i+1}^n} - W_{t_i^n}$$

for the i-th partition point and the Brownian increment across the i-th interval, so that the step length is $t_{i+1}^n - t_i^n = t/2^n$. A subscript on an expectation means conditioning on the information available then, so $\mathbb{E}_{t_i^n}[\,\cdot\,] = \mathbb{E}[\,\cdot \mid \mathcal{F}_{t_i^n}]$.

Calculation 1.17 (Expectation of Itô Integral)**.**

$$\begin{aligned}
\mathbb{E}[\int_0^t H_s dW_s|\mathcal{F}_0] &= \mathbb{E}[\lim_{n\to\infty}\sum_{i=0}^{2^n-1} H_{t_i^n}\Delta_i W|\mathcal{F}_\prime] && \text{(by definition)}\\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[H_{t_i^n}\Delta_i W|\mathcal{F}_0] && \text{(by square integrability of } H_s\text{)}\\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[\mathbb{E}[H_{t_i^n}\Delta_i W|\mathcal{F}_{t_i^n}]|\mathcal{F}_0] && \text{(iterated conditioning)}\\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[H_{t_i^n}\mathbb{E}_{t_i^n}[\Delta_i W]|\mathcal{F}_0] && (H_s \text{ is adapted to } \mathcal{F}_s)\\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[H_{t_i^n}0|\mathcal{F}_0] && (W_s \text{ is } \mathcal{F}_s \text{ Brownian})\\
&= 0
\end{aligned}$$

By iterated conditioning we also have $\mathbb{E}[\int_0^t H_s dW_s] = \mathbb{E}[\mathbb{E}[\int_0^t H_s dW_s|\mathcal{F}_0]] = \mathbb{E}[0] = 0$.

Calculation 1.18 (Covariance of Itô Integral)**.**

$$
\begin{aligned}
&\mathbb{E}[\int_0^t H_s dW_s \int_0^t K_s dW_s | \mathcal{F}_0] \\
&= \mathbb{E}[\lim_{n,m\to\infty} \sum_{i,j=0}^{2^n-1, 2^m-1} H_{t_i^n} \Delta_i W K_{jt/2^m} (W_{(j+1)t/2^m} - W_{jt/2^m}) | \mathcal{F}_0] \\
&\qquad\qquad\qquad\qquad \text{(by definition)} \\
&= \lim_{n,m\to\infty} \sum_{i,j=0}^{2^n-1, 2^m-1} \mathbb{E}[H_{t_i^n} \Delta_i W K_{jt/2^m} (W_{(j+1)t/2^m} - W_{jt/2^m}) | \mathcal{F}_0] \\
&\qquad\qquad\qquad\qquad \text{(by square integrability of } H_s\text{)} \\
&= \lim_{n,m\to\infty} \sum_{i,j=0}^{2^n-1, 2^m-1} \mathbb{E}[\mathbb{E}[H_{t_i^n} \Delta_i W K_{jt/2^m} (W_{(j+1)t/2^m} - W_{jt/2^m}) | \mathcal{F}_{\max(i,j)t/2^n}] | \mathcal{F}_0] \\
&\qquad\qquad\qquad\qquad \text{(iterated conditioning)} \\
&= \lim_{n\to\infty} \sum_{i,j=0}^{2^n-1} \mathbb{E}[\mathbb{E}[H_{t_i^n} \Delta_i W K_{t_j^n} \Delta_j W | \mathcal{F}_{\max(i,j)t/2^n}] | \mathcal{F}_0] \\
&\qquad\qquad\qquad\qquad \text{(assuming same partition to simplify algebra)} \\
&= \lim_{n\to\infty} \sum_{i,j=0}^{2^n-1} \begin{cases} \mathbb{E}[H_{t_i^n} K_{t_j^n} \Delta_j W \mathbb{E}_{t_i^n}[\Delta_i W] | \mathcal{F}_0] & \text{if } i > j \\ \mathbb{E}[H_{t_i^n} K_{t_i^n} \mathbb{E}_{t_i^n}[(\Delta_i W)^2] | \mathcal{F}_0] & \text{if } i = j \\ \mathbb{E}[H_{t_i^n} \Delta_i W K_{t_j^n} \mathbb{E}_{t_j^n}[\Delta_j W] | \mathcal{F}_0] & \text{if } i < j \end{cases} \\
&= \lim_{n\to\infty} \sum_{i,j=0}^{2^n-1} \begin{cases} \mathbb{E}[H_{t_i^n} K_{t_j^n} \Delta_j W 0 | \mathcal{F}_0] & \text{if } i > j \\ \mathbb{E}[H_{t_i^n} K_{t_i^n} \mathrm{Var}_{t_i^n}[\Delta_i W] | \mathcal{F}_0] & \text{if } i = j \\ \mathbb{E}[H_{t_i^n} \Delta_i W K_{t_j^n} 0 | \mathcal{F}_0] & \text{if } i < j \end{cases} \\
&= \lim_{n\to\infty} \sum_{i,j=0}^{2^n-1} \begin{cases} 0 & \text{if } i \neq j \\ \mathbb{E}[H_{t_i^n} K_{t_i^n} t/2^n | \mathcal{F}_0] & \text{if } i = j \end{cases} \\
&= \lim_{n\to\infty} \sum_{i=0}^{2^n-1} \mathbb{E}[H_{t_i^n} K_{t_i^n}\, t/2^n | \mathcal{F}_0] \\
&= \mathbb{E}[\lim_{n\to\infty} \sum_{i=0}^{2^n-1} H_{t_i^n} K_{t_i^n}\, t/2^n | \mathcal{F}_0] \\
&= \mathbb{E}[\int_0^t H_s K_s ds | \mathcal{F}_0] \qquad \text{(Riemann sum, with } t/2^n \text{ the width)}
\end{aligned}
$$

Before writing down Itô's Isometry, let's set up the two spaces it relates, since the statement is about a map between them.

Definition 1.19 (The two spaces)**.** $L^2(\Omega)$ is the set of random variables X with finite second moment, $\mathbb{E}[X^2] < \infty$. An element of it is a single random number — the value of something at one moment — and "finite second moment" is the condition that makes its variance meaningful. It is given the inner product

$$(X,Y)_{L^2(\Omega)} \equiv \mathbb{E}[XY].$$

$L^2_{ad}([0,T]\times\Omega)$ is the set of processes $X_s(\omega)$, defined for times $s \in [0,T]$ and states $\omega \in \Omega$, which are

- *adapted* — that is the subscript "ad" — so X_s is $\mathcal{F}_s$ measurable, meaning its value at time s uses only information available by then; and
- square integrable over both time and chance, $\mathbb{E}\big[\int_0^T X_s^2\,ds\big] < \infty$.

An element of it is a random *function of time* — a whole path — and it carries the inner product

$$(X,Y)_{L^2_{ad}([0,T]\times\Omega)} \equiv \mathbb{E}\left[\int_0^T X_s Y_s\,ds\right].$$

One contains random numbers, the other random paths, and both are equipped with a notion of length and angle built from an expectation. Adaptedness is doing real work in the second and is not a technical afterthought — it is what rules out an integrand that peeks at the increment it is about to be multiplied by, which is the whole reason the Itô integral has zero mean.

Remark (Itô Isometry)**.** Taking expectations of the above to remove the conditioning, the result is known as the Itô isometry. Write $\mathcal{I}(X) = \int_0^T X_s\,dW_s$ for the Itô integral over $[0,T]$. It takes a random path and returns a random number, so it is a map

$$\mathcal{I} : L^2_{ad}([0,T]\times\Omega) \longrightarrow L^2(\Omega),$$

and it preserves the inner product:

$$\big(\mathcal{I}(H), \mathcal{I}(K)\big)_{L^2(\Omega)} = (H,K)_{L^2_{ad}([0,T]\times\Omega)},$$

which written out is exactly the calculation above with the conditioning dropped:

$$\mathbb{E}\left[\int_0^T H_s\,dW_s \int_0^T K_s\,dW_s\right] = \mathbb{E}\left[\int_0^T H_s K_s\,ds\right].$$

In words: the integral carries a random path to a random number without distorting lengths or angles. Two integrands that are orthogonal — meaning $\mathbb{E}[\int H_s K_s\,ds] = 0$ — have uncorrelated integrals, and the size of an integral is the size of its integrand. That is what "isometry" means: a map preserving the geometry, in the way that a rotation of the plane preserves distances.

The isometry is not only a fact about the integral; it is what allows the integral to be defined at all. The argument is worth going through, both because the definition we started from is on shakier ground than it looks and because the same argument is used throughout analysis.

The definition was a limit of sums $\sum_i H_{t_i}(W_{t_{i+1}} - W_{t_i})$. For each partition that sum is a perfectly good random variable. But nothing so far says the sequence of them converges as the partition refines, and a sequence of random variables does not converge merely because we would like it to. Something has to force it. The construction goes in three steps.

First, the integrands for which there is nothing to prove. Call a process *simple* if it is piecewise constant,

$$H_s = \sum_i h_i \, \mathbb{I}_{(t_i, t_{i+1}]}(s),$$

with each h_i already known at time t_i. Its integral is not defined by a limit at all; it is declared outright to be the finite sum

$$\mathcal{I}(H) = \sum_i h_i \big(W_{t_{i+1}} - W_{t_i}\big),$$

and a finite sum of products of random variables raises no questions. The covariance calculation above, applied to two simple processes, is exactly the statement that the isometry holds for these.

Second, approximate. Every $H \in L^2_{ad}([0,T] \times \Omega)$ is a limit of simple processes: there are simple $H^{(n)}$ with $\|H^{(n)} - H\| \to 0$ in that space. We would like to *define* $\mathcal{I}(H)$ as $\lim_n \mathcal{I}(H^{(n)})$ — but that limit has to be shown to exist before it can be used as a definition.

Third, the isometry supplies it. The integral is linear on simple processes, so the difference of two of the approximating integrals is the integral of the difference, and the isometry applies to it:

$$\big\|\mathcal{I}(H^{(n)}) - \mathcal{I}(H^{(m)})\big\|_{L^2(\Omega)} = \big\|\mathcal{I}(H^{(n)} - H^{(m)})\big\|_{L^2(\Omega)} = \big\|H^{(n)} - H^{(m)}\big\|_{L^2_{ad}([0,T]\times\Omega)}.$$

The right hand side tends to zero, because a sequence that converges is Cauchy. So the left hand side does too, which says the integrals $\mathcal{I}(H^{(n)})$ form a Cauchy sequence in $L^2(\Omega)$. And $L^2(\Omega)$ is *complete* — every Cauchy sequence in it has a limit inside it — so the limit exists. The same estimate shows it does not depend on which approximating sequence was chosen, so the definition is unambiguous.

Remark (What the isometry is doing)**.** It is a bridge between two things we can measure. On one side we can make the integrands close, because that is how the approximation was built. On the other we need the integrals to be close, because that is what makes the limit exist. The isometry says those are the same statement.

Without an estimate of that kind the approximating integrals could oscillate forever without settling, and there would be no integral to speak of — the definition would not fail to be useful, it would fail to be a definition.

With $H = K$, the above two results imply

$$\begin{aligned} \mathbb{E}[\int_0^t H_s dW_s] &= 0 \\ \mathrm{Var}[\int_0^t H_s dW_s] &= \mathbb{E}[\int_0^t H_s^2 ds]. \end{aligned}$$

One key feature of Itô calculus is that the second order differentials are not equivalent to 0. In ordinary calculus we have

$$\begin{aligned}
\int_0^t X(t)(dt)^2 &\equiv \lim_{n\to\infty} \sum_{i=0}^{2^n-1} X(it/2^n)((i+1)t/2^n - it/2^n)^2 \\
&= \lim_{n\to\infty} \sum_{i=0}^{2^n-1} X(it/2^n)(t/2^n)^2 \\
&\le \max_{0<s<t}(X(s)) \lim_{n\to\infty} \sum_{i=0}^{2^n-1} (t/2^n)^2 \\
&\le \max_{0<s<t}(X(s))t \lim_{n\to\infty} \sum_{i=0}^{2^n-1} (t/2^n) \\
&= 0
\end{aligned}$$

In stochastic calculus we can check the expectation and variance of the distribution of the integral:

Calculation 1.20 (Integral of second order time differential)**.**

$$\begin{aligned}
\mathbb{E}[\int_0^t X_t(dt)^2] &\equiv \mathbb{E}[\lim_{n\to\infty} \sum_{i=0}^{2^n-1} X_{it/2^n}((i+1)t/2^n - it/2^n)^2] \\
&= \lim_{n\to\infty} \sum_{i=0}^{2^n-1} \mathbb{E}[X_{it/2^n}](t/2^n)^2 \\
&\le \max_{0<s<t}(|\mathbb{E}[X_s]|) \lim_{n\to\infty} \sum_{i=0}^{2^n-1} (t/2^n)^2 \\
&\le \max_{0<s<t}(|\mathbb{E}[X_s]|)t \lim_{n\to\infty} \sum_{i=0}^{2^n-1} (t/2^n) \\
&= 0
\end{aligned}$$

Variance can be similarly computed to be 0 if the stochastic process has finite mean and variance at all times. Therefore the distribution is 0 with probability 1.

However the integral does not necessarily vanish when taken with respect to second order differential of Brownian motion. We have

Calculation 1.21 (Expectation of integral of second order Brownian motion differential)**.**

$$
\begin{aligned}
\mathbb{E}[\int_0^t H_s(dW_s)^2 \mid \mathcal{F}_0] &= \mathbb{E}[\lim_{n\to\infty}\sum_{i=0}^{2^n-1} H_{t_i^n}\Delta_i W^2 \mid \mathcal{F}_0] \\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[H_{t_i^n}\Delta_i W^2 \mid \mathcal{F}_0] \\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[\mathbb{E}[H_{t_i^n}\Delta_i W^2 \mid \mathcal{F}_{t_i^n}] \mid \mathcal{F}_0] \\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[H_{t_i^n}\mathbb{E}[\Delta_i W^2 \mid \mathcal{F}_{t_i^n}] \mid \mathcal{F}_0] \\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[H_{t_i^n} t/2^n \mid \mathcal{F}_0] \\
&= \mathbb{E}[\int_0^t H_s dt \mid \mathcal{F}_0]
\end{aligned}
$$

We therefore have $\mathbb{E}[\int_0^t H_s(dW_s)^2 - \int_0^t H_s dt] = 0$. So the two integrals have same expectation. Let's see if their difference has any variance.

Calculation 1.22 (Variance of integral of second order Brownian motion differential)**.**

$$
\begin{aligned}
&\mathbb{E}[\Big(\int_0^t H_s(dW_s)^2 - \int_0^t H_s dt\Big)^2|\mathcal{F}_0] \\
&= \mathbb{E}[\lim_{n\to\infty}\Big(\sum_{i=0}^{2^n-1} H_{t_i^n}\big(\Delta_i W^2 - ((i+1)t/2^n - it/2^n)\big)\Big)^2|\mathcal{F}_0] \\
&= \mathbb{E}[\lim_{n\to\infty}\Big(\sum_{i=0}^{2^n-1} H_{t_i^n}\big(\Delta_i W^2 - t/2^n\big)\Big)^2|\mathcal{F}_0] \\
&= \mathbb{E}[\lim_{n\to\infty}\sum_{i,j=0}^{2^n-1} H_{t_i^n}H_{t_j^n}(\Delta_i W^2 - t/2^n)((\Delta_j W)^2 - t/2^n)|\mathcal{F}_0] \\
&= \lim_{n\to\infty}\sum_{i,j=0}^{2^n-1} \mathbb{E}[H_{t_i^n}H_{t_j^n}(\Delta_i W^2 - t/2^n)((\Delta_j W)^2 - t/2^n)|\mathcal{F}_0] \\
&= \lim_{n\to\infty}\sum_{i,j=0}^{2^n-1} \mathbb{E}[\mathbb{E}[H_{t_i^n}H_{t_j^n}(\Delta_i W^2 - t/2^n)((\Delta_j W)^2 - t/2^n)|\mathcal{F}_{\max(i,j)t/2^n}]|\mathcal{F}_0] \\
&= \lim_{n\to\infty}\sum_{i,j=0}^{2^n-1} \begin{cases} \mathbb{E}[H_{t_i^n}H_{t_j^n}\big((\Delta_j W)^2 - t/2^n\big)\mathbb{E}[\Delta_i W^2 - t/2^n|\mathcal{F}_{t_i^n}]|\mathcal{F}_0] & \text{if } i > j \\ \mathbb{E}[(H_{t_i^n})^2\mathbb{E}[(\Delta_i W^2 - t/2^n)^2|\mathcal{F}_{t_i^n}]|\mathcal{F}_0] & \text{if } i = j \\ \mathbb{E}[H_{t_i^n}H_{t_j^n}\big((\Delta_i W)^2 - t/2^n\big)\mathbb{E}[\Delta_j W^2 - t/2^n|\mathcal{F}_{t_j^n}]|\mathcal{F}_0] & \text{if } i < j \end{cases} \\
&= \lim_{n\to\infty}\sum_{i,j=0}^{2^n-1} \begin{cases} \mathbb{E}[H_{t_i^n}H_{t_j^n}\big((\Delta_j W)^2 - t/2^n\big)0|\mathcal{F}_0] & \text{if } i > j \\ \mathbb{E}[(H_{t_i^n})^2\mathbb{E}[\Delta_i W^4 + (t/2^n)^2 - 2(W_{(i+1)t/2^n} - W_{it/2^n})^2 t/2^n|\mathcal{F}_{t_i^n}]|\mathcal{F}_0] & \text{if } i = j \\ \mathbb{E}[H_{t_i^n}H_{t_j^n}\big((\Delta_i W)^2 - t/2^n\big)0|\mathcal{F}_0] & \text{if } i < j \end{cases} \\
&= \lim_{n\to\infty}\sum_{i,j=0}^{2^n-1} \begin{cases} 0 & \text{if } i \neq j \\ \mathbb{E}[(H_{t_i^n})^2(3(t/2^n)^2 + (t/2^n)^2 - 2(t/2^n)^2)|\mathcal{F}_0] & \text{if } i = j \end{cases} \\
&= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} \mathbb{E}[(H_{t_i^n})^2(2(t/2^n)^2)|\mathcal{F}_0] \\
&= \int_0^t 2\mathbb{E}[(H_s)^2|\mathcal{F}_0](dt)^2 \\
&= 0
\end{aligned}
$$

We therefore have $\mathbb{E}[\int_0^t H_s(dW_s)^2 - \int_0^t H_s dt] = 0$ and $\mathrm{Var}[\int_0^t H_s(dW_s)^2 - \int_0^t H_s dt] = 0$ and therefore $\mathbb{P}(\int_0^t H_s(dW_s)^2 = \int_0^t H_s dt) = 1$ i.e.

$$\int_0^t H_s(dW_s)^2 = \int_0^t H_s dt \quad \text{almost surely}$$

Remark (Itô Differentials)**.** The above result is also written in differential notation as $(dW)^2 = dt$. We can similarly check that $dtdW = 0$ and $(dW)^3 = 0$. Therefore we have the algebra on differentials:

$$(dt)^2 = 0$$
$$(dW)^2 = dt$$

The identity $(dW)^2 = dt$ is a statement about Brownian motion in particular, and chapter 2 needs the corresponding statement for processes that are not Brownian motion. The object it names is the accumulated sum of squared increments, which is what the figure above was plotting.

Definition 1.23 (Quadratic variation)**.** Let X be a continuous adapted process and let Π_n be a sequence of partitions $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ of $[0,t]$ whose mesh $\max_i(t_{i+1}^n - t_i^n)$ tends to zero. The *quadratic variation* of X is

$$\langle X\rangle_t = \lim_{n\to\infty} \sum_{i=0}^{k_n-1} \left(X_{t_{i+1}^n} - X_{t_i^n}\right)^2,$$

the limit taken in probability, when it exists and does not depend on the choice of partitions.

For a continuous local martingale the limit always exists. That is the Doob–Meyer decomposition, and stated in full generality — for submartingales, with no continuity assumed — its proof is a long argument about weak compactness. The case these notes use is the one whose proof is short, and it is short because it is the argument that built the Itô integral, run a second time.

Theorem 1.24 (The bracket of a continuous square integrable martingale)**.** *Let M be a continuous martingale with $M_0 = 0$ and $\mathbb{E}[M_t^2] < \infty$ for every t. Then the sums of squared increments converge in L^2, so $\langle M\rangle_t$ exists. It is continuous, non-decreasing, adapted and zero at the origin; $M^2 - \langle M\rangle$ is a martingale; and $\langle M\rangle$ is the only process with those properties.*

Proof. Write $Q_n(t) = \sum_i \left(M_{t_{i+1}^n} - M_{t_i^n}\right)^2$ for the sum over the partition Π_n of $[0,t]$. Everything follows from one piece of algebra. Since $b^2 - a^2 = (b-a)^2 + 2a(b-a)$, summing over the partition telescopes the left side and gives

$$M_t^2 = Q_n(t) + 2\sum_i M_{t_i^n}\left(M_{t_{i+1}^n} - M_{t_i^n}\right). \tag{1.3}$$

The last term is a martingale in t for each fixed partition, being a sum of increments each multiplied by a factor known before its increment occurs — the same predictability that gives the Itô integral zero mean. Taking expectations, $\mathbb{E}[Q_n(t)] = \mathbb{E}[M_t^2]$ at every resolution, so the sums are not merely converging to the right object, they have the right expectation all the way along.

That also identifies what Q_n is: by (1.3) it is M_t^2 minus a discrete stochastic integral of M against itself, so its convergence is the convergence of that integral, which is the construction of the Itô integral and nothing new. Refining the partition changes the integral by an integral of the difference between M and its piecewise-constant approximation, the isometry bounds that in L^2 by the size of the approximation error, and continuity of the path drives the error to zero uniformly on $[0,t]$. So $\{Q_n(t)\}$ is Cauchy in L^2; L^2 is complete, so the limit exists; and the same bound shows it is independent of the sequence of partitions. Passing to the limit in (1.3) leaves $M^2 - \langle M\rangle$ as an L^2 limit of martingales, hence a martingale.

The path properties come along with it. Each Q_n is non-decreasing in t, and a limit of non-decreasing functions is non-decreasing; adaptedness is inherited term by term; and continuity follows from the same L^2 bound, which controls the increase of $\langle M\rangle$ across an interval by the oscillation of M on it.

Uniqueness needs no new machinery. If A and A' both qualify then $A-A' = (M^2-A')-(M^2-A)$ is a difference of martingales, hence a martingale; it is also continuous, zero at the origin, and of finite variation, being a difference of two non-decreasing functions. By the lemma below such a process has zero quadratic variation, and a continuous martingale starting at zero whose bracket vanishes is identically zero. So $A = A'$. $\square$

Two things about the scope of that. Localisation extends it from square integrable martingales to local martingales in the usual way — stop at each τ_n, take the bracket of the stopped process, and observe that the brackets agree where they overlap because each is a limit of the same sums. And it does not require M to be an Itô integral, which matters, because the Cantor-function example below is a martingale it covers and an integral it does not. What we do still need to check is which parts of a process contribute.

Lemma 1.25 (Only the martingale part contributes)**.** *Let A be continuous and of finite variation on $[0,t]$, and let X be continuous with quadratic variation $\langle X\rangle_t$. Then $\langle A\rangle_t = 0$ and $\langle X+A\rangle_t = \langle X\rangle_t$.*

Proof. Write V for the total variation of A on $[0,t]$ and $\omega(\delta)$ for its modulus of continuity. Then

$$\sum_i (A_{t^n_{i+1}} - A_{t^n_i})^2 \le \max_i |A_{t^n_{i+1}} - A_{t^n_i}| \sum_i |A_{t^n_{i+1}} - A_{t^n_i}| \le \omega(\operatorname{mesh}\Pi_n)\, V,$$

and A is uniformly continuous on the compact interval, so $\omega(\operatorname{mesh}\Pi_n) \to 0$. For the second claim expand the square and bound the cross term by Cauchy–Schwarz,

$$\Big|2\sum_i (X_{t^n_{i+1}} - X_{t^n_i})(A_{t^n_{i+1}} - A_{t^n_i})\Big| \le 2\Big(\sum_i (\Delta X_i)^2\Big)^{1/2}\Big(\sum_i (\Delta A_i)^2\Big)^{1/2},$$

whose second factor tends to zero while the first stays bounded. $\square$

So the drift of a diffusion is invisible to $\langle X\rangle$ — which is the content of $(dt)^2 = 0$ and $dt\,dW = 0$ — and the quadratic variation measures the martingale part alone. For the martingale part itself the two calculations above have already done the work: if $X_t = \int_0^t H_s\,dW_s$ then the increments are $\Delta X_i \approx H_{t_i}\Delta W_i$, and summing their squares is summing $H_{t_i}^2(\Delta W_i)^2$, which we showed converges to $\int_0^t H_s^2\,ds$. Hence

$$X_t = \int_0^t H_s\,dW_s \quad\Longrightarrow\quad \langle X\rangle_t = \int_0^t H_s^2\,ds. \tag{1.4}$$

One consequence recurs whenever two candidate martingales have to be shown equal: a continuous local martingale M with $M_0 = 0$ and $\langle M\rangle = 0$ vanishes identically. Indeed $M^2 - \langle M\rangle = M^2$ is then a non-negative local martingale, hence a supermartingale, so $\mathbb{E}[M_t^2] \le M_0^2 = 0$.

Definition 1.26 (The differential of the quadratic variation)**.** Since almost every path of $\langle X\rangle$ is a continuous non-decreasing function starting at zero, it defines a measure on $[0,\infty)$ by $d\langle X\rangle\big((s,u]\big) = \langle X\rangle_u - \langle X\rangle_s$, and $\int_0^t H_s\,d\langle X\rangle_s$ means the Lebesgue–Stieltjes integral of H against that measure, computed path by path.

Remark (What $d\langle X\rangle_t$ is and is not)**.** It is tempting to read $d\langle X\rangle_t$ as "some multiple of dt", on the grounds that the only thing a differential in t can be proportional to is dt. Two things hide inside that reading.

First, the multiple is random. For $X = \int H\,dW$, equation (1.4) gives $d\langle X\rangle_t = H_t^2\,dt$, and H_t is an adapted process, not a number: the density is a random variable that the path itself determines, and it can depend on the same noise that drives X. That density is exactly what a stochastic volatility model models.

Second, there need not be a multiple at all. A measure carries no obligation to be absolutely continuous with respect to Lebesgue measure. Let A be the Cantor function, continuous and non-decreasing with $A' = 0$ almost everywhere, and set $M_t = B_{A_t}$ for a Brownian motion B. Then M is a continuous martingale in the time-changed filtration with $\langle M\rangle = A$, so its quadratic variation increases while having no density: all of the randomness arrives on a Lebesgue-null set of times. Writing an equation $dX_t = \mu_t\,dt + \sigma_t\,dW_t$ rules such a process out. The ansatz is not only a choice of drift and volatility; it is already the assumption that the quadratic variation accumulates at a rate.

What the bracket *means* has to wait for the theorem that establishes it, which is chapter 2's: every continuous local martingale turns out to be a Brownian motion watched on a clock, and $\langle M\rangle$ is the clock's reading. Until then it is enough that it measures accumulated randomness, and that a diffusion's drift contributes none of it.

Two processes have a covariation for the same reason one has a variation.

Definition 1.27 (Covariation)**.** For continuous adapted X and Y,

$$\langle X, Y\rangle_t = \lim_{n\to\infty} \sum_{i=0}^{k_n-1} \big(X_{t^n_{i+1}} - X_{t^n_i}\big)\big(Y_{t^n_{i+1}} - Y_{t^n_i}\big) = \tfrac{1}{4}\big(\langle X+Y\rangle_t - \langle X-Y\rangle_t\big),$$

the second equality being the polarisation identity. It is symmetric and bilinear, and $\langle X, X\rangle = \langle X\rangle$.

In differential notation the product of two differentials *is* the covariation, $dX\,dY = d\langle X, Y\rangle$, and the algebra of the remark above is the special case

$$dt\,dt = 0, \qquad dW\,dt = 0, \qquad dW^i\,dW^j = \rho_{ij}\,dt,$$

for correlated Brownian motions W^i, since $\langle W^i, W^j\rangle_t = \rho_{ij}t$.

From the ordinary calculus we have the Taylor expansion of a function as

$$f((x) + \boldsymbol{\delta x}) = f(\boldsymbol{x}) + f'(\boldsymbol{x})\cdot\boldsymbol{\delta x} + \frac{1}{2}(\boldsymbol{\delta x})^T \cdot f'' \cdot \boldsymbol{\delta x} + \ldots$$

where f' is the gradient and f'' is the Hessian. When taking the limit $\delta x \to dx$ we often ignore the second and higher order differentials $(dx)^2, (dx)^3, \ldots$ giving the usual chain rule. In stochastic calculus, however, as we have seen second order differentials need not be 0. And we have:

Remark (Itô's Lemma)**.**

$$df(\boldsymbol{x}) = f'(\boldsymbol{x})\cdot\boldsymbol{dx} + (\boldsymbol{dx})^T \cdot f'' \cdot \boldsymbol{dx} \tag{1.5}$$

and in the special case when f is a function of t and a stochastic process X_t.

$$df(t, X_t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial X_t}dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial X_t^2}(dX_t)^2$$

where the second order term is $(dX_t)^2 = d\langle X\rangle_t$, so the lemma applies to any continuous process whose quadratic variation exists, and not only to functions of W. When X_t is a Brownian motion W_t we have

$$df(t, X_t) = \Big(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W_t^2}\Big)dt + \frac{\partial f}{\partial W_t}dW_t$$

Definition 1.28 (Itô Diffusion)**.** An Itô diffusion is a stochastic process satisfying a stochastic differential equation of the form

$$d\boldsymbol{X}_t = \underbrace{\boldsymbol{\mu}(t, \boldsymbol{X}_t)dt}_{\text{drift}} + \underbrace{\boldsymbol{\sigma}(t, \boldsymbol{X}_t)\cdot d\boldsymbol{W}_t}_{\text{diffusion}} \tag{1.6}$$

The diffusion coefficient in the equation above is also known as volatility.

A stochastic integral against a Brownian motion always has zero drift, but its expectation can still fail to be constant if the integrand is wild enough, which is why the conclusion is local.

Remark. An Itô diffusion with zero drift,

$$d\boldsymbol{X}_t = \boldsymbol{\sigma}(t, \boldsymbol{X}_t)\cdot d\boldsymbol{W}_t,$$

is a local martingale.

Example 1.2 (Geometric Brownian Motion)**.**

$$\frac{dX_t}{X_t} = \mu dt + \sigma dW_t \tag{1.7}$$

$$\begin{aligned}
d\log(X_t) &= \frac{dX_t}{X_t} - \frac{1}{2}\frac{1}{X_t^2}(dX_t)^2 \qquad \text{(by Itô's Lemma)}\\
&= \mu dt + \sigma dW_t - \frac{1}{2}\sigma^2 dt\\
&= \big(\mu - \frac{1}{2}\sigma^2\big)dt + \sigma dW_t\\
\log(X_t) - \log(X_0) &= \int_0^t \big(\mu - \frac{1}{2}\sigma^2\big)ds + \int_0^t \sigma\, dW_s\\
&= \big(\mu - \frac{1}{2}\sigma^2\big)t + \sigma W_t\\
X_t &= X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}
\end{aligned}$$

Because of its simplicity Geometric Brownian Motion has been widely used in finance as a model for evolution of stock prices.

References

- Protter, P. (2005). *Stochastic Integration and Differential Equations.* Springer.