CHAPTER IV

Introduction to Itô Calculus

Here, we give the gist of the 'martingale and stochastic integral' method, and illustrate its use via a large number of fully-worked examples. We do not apologize for sometimes advertising the method by showing how it can obtain results which are well known and elementary. Thus, for example, we take the trouble to prove some standard results about the humble Markov chain with finite state-space. But we have also tried to bring into this chapter applications which are less elementary, and which hint at the excitement of the subject today.

TERMINOLOGY AND CONVENTIONS

R-processes and L-processes

We now use the term *R-process* on $[0, \infty)$ to signify a process all of whose paths are right-continuous on $[0, \infty)$ with limits from the left on $(0, \infty)$. Thus an R-process is what was called in Volume 1 a Skorokhod process, and what is called elsewhere a càdlàg process, or a corlol process, or whatever. An R-function or R-path on $[0, \infty)$ is defined via the obvious analogous definition.

The *L-processes* on $(0, \infty)$, all of whose paths are left-continuous with limits from the right, will now begin to feature largely in the theory.

Usual conditions, etc.

Everywhere in this chapter, we work with a set-up $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbf{P})$ *satisfying the 'usual conditions'. See (II.67).*

All martingales (and 'finite-variation processes', and 'semimartingales') will be taken to be R-processes. Because we are assuming that the usual conditions hold, this is in order. See (II.67).

We shall also always assume that a process $\{X_t : t \geqslant 0\}$ *is jointly measurable; that is, the map* $(t, \omega) \mapsto X_t(\omega)$ *is measurable with respect to* $\mathscr{B}(\mathbb{R}^+) \times \mathscr{F}$.

Recall that the process X is said to be adapted if X_t *is* $\mathscr{F}_t$*-measurable for every* $t \geqslant 0$.

Important convention about time 0

Our stochastic integrals will be defined over intervals $(0, t]$ open at 0. Thus, the value of the integral at time 0 will be 0. This differs from the convention in Dellacherie and Meyer [1]. As explained there, time 0 plays *le rôle du diable*. We consign it to Hell.

In accordance with this convention, the parameter set for our previsible processes will be the open interval $(0, \infty)$.

1. SOME MOTIVATING REMARKS

1. Itô integrals. One of our main tasks is to define the Itô integral

$$\int H\,dX,$$

where H and X are stochastic processes of appropriate classes.

We shall regard this integral as a new stochastic process, often written $H\bullet X$, and shall use the alternative notations:

$$(H\bullet X)(t,\omega)=\left(\int_{(0,t]} H_s\,dX_s\right)(\omega). \tag{1.1}$$

We shall often use differential notation, in which we can rewrite equation (1.1) as a 'stochastic differential equation':

$$d(H\bullet X)=H\,dX. \tag{1.2}$$

The theory is now essentially complete in the sense that it is known exactly what conditions need to be imposed on our integrand H and integrator X:

The essential requirement on the integrand H is that it be 'previsible'.
The integrator X must be a 'semimartingale'.

The most important example of a previsible process is provided by an *adapted L-process*. Indeed, the adapted L-processes 'generate' the previsible processes, as will be explained later. Let H be an adapted L-process. Then H_s is known to the observer at time s. The reason that H is 'previsible' is (roughly speaking) that, for a stopping time $T>0$, H_T is known immediately before time T because

$$H_T=\lim_{s\uparrow\uparrow T} H_s.$$

The simplest adapted L-process is the process

$$H=1(S,T], \tag{1.3}$$

where S and T are stopping times with $S < T$. Thus,

$$H(t, \omega) = \begin{cases} 1 & \text{if } S(\omega) < t \leqslant T(\omega), \\ 0 & \text{otherwise.} \end{cases}$$

Let M be a martingale. Then, for H as in equation (1.3), the obvious definition of $H \cdot M$ is:

$$(H \cdot M)_t = M_{T \wedge t} - M_{S \wedge t},$$

and we can easily show that *$H \cdot M$ is a martingale.*

From this simple case develops *the most fundamental property of Itô integrals*:

(*1.4*) **ITÔ INTEGRALS PRESERVE LOCAL MARTINGALES.** It will do no harm to give a precise statement of this now. The reader new to the subject will not know what Theorem 1.5 means, but it will help him or her to know that it is one of the main landmarks in our route through the subject.

(*1.5*) **FUNDAMENTAL THEOREM.** *If H is a locally bounded previsible process and M is a local martingale, then $H \cdot M$ exists and is a local martingale.*

We could very easily explain now what a locally bounded previsible process is. We could also easily explain what a local martingale is; indeed, let us do it:

(*1.6*) DEFINITION (local martingale): *A process M is a* local martingale *if M_0 is $\mathscr{F}_0$ measurable and there exists an increasing sequence of stopping times (T_n) with $T_n \uparrow \infty$ such that each 'stopped' process*

$$\{M_{T_n \wedge t} - M_0 : t \geqslant 0\}$$

is a martingale.

What we cannot explain in a short space is what $H \cdot M$ means in the generality of Theorem 1.5. But the discrete-time setting explains why Theorem 1.5 is true.

(*1.7*) *A discrete-time analogue.* Let $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}, \mathrm{P})$ be a discrete-time set-up, and let M be an associated martingale. Let H be a bounded process previsible in the sense that

$$Z_{n-1} = H_n \in b\mathscr{F}_{n-1} \quad (n \in \mathbb{N}).$$

Define:

$$(H \cdot M)_n = \sum_{k=1}^{n} H_k (M_k - M_{k-1}) = \sum_{k=1}^{n} Z_{k-1} (M_k - M_{k-1})$$

$$(H \cdot M)_0 = 0.$$

Then $H \cdot M$ is a martingale.

Proof. To show that a process N is a martingale, we need only show that

$$\mathbf{E}[N_n - N_{n-1} | \mathcal{F}_{n-1}] = 0, \quad n \in \mathbb{N}.$$

But, since $Z_{n-1} \in b\mathcal{F}_{n-1}$,

$$\begin{aligned}\mathbf{E}[(H \bullet M)_n - (H \bullet M)_{n-1} | \mathcal{F}_{n-1}] &= \mathbf{E}[Z_{n-1}(M_n - M_{n-1}) | \mathcal{F}_{n-1}] \\ &= Z_{n-1}\mathbf{E}[M_n - M_{n-1} | \mathcal{F}_{n-1}] = 0.\end{aligned}$$ □

As was mentioned earlier, *the general 'integrator' X will be a semimartingale.* This means that *X may be written in the form*

(1.8) $$X = X_0 + M + A,$$

where X_0 is $\mathcal{F}_0$ measurable, M is a local martingale null at 0, *and A is an adapted process with paths of finite variation, also null at* 0.

In this chapter, we present the full theory for two special cases of great importance:

(i) *the case in which $X = A$, a process with paths of finite variation*;
(ii) *the case in which the paths of X are continuous.*

This will allow us to develop many of the main applications. The general theory is given in Chapter VI.

If A is a process with paths of finite variation, then (for a bounded measurable process H) we can define $H \bullet A$ as the Stieltjes integral for each ω:

$$(H \bullet A)(t, \omega) = \int_{(0,t]} H(s, \omega)\, dA(s, \omega).$$

Though no new concept of integration is involved here, the theory is extremely useful because of what Theorem 1.5 says in this context:

(*1.9*) THEOREM. *Let H be a locally bounded previsible process, and let M be a local martingale with paths of finite variation. Then $H \bullet M$, as defined by the Stieltjes integral, is a local martingale.*

If M is a (path-) continuous local martingale, then the paths of M generally will not have finite variation. Indeed, the only paths of finite variation will be constant! Thus the integral $H \bullet M$ (where H is a locally bounded previsible process) is a true extension of the Stieltjes integral. The very existence of the integral is inextricably tied up with its calculus, that is, with the *integration-by-parts* formula and the *pièce de résistance* of the theory, *Itô's formula.*

2. Integration by parts. The most important integral associated with a local martingale M is the integral $\int_{(0,t]} M_{s-} dM_s$. The adapted L-process $M_- = \{M_{s-} : s > 0\}$ is previsible and also locally bounded, so the integral exists. More generally, if X and Y are semimartingales, then the Itô integral $\int X_{s-} dY_s$ may be defined. Now the integral $\int X_{s-}dY_s$ is analogous to a sum of the form

(2.1) $$\sum x_{k-1}(y_k - y_{k-1}).$$

The summation-by-parts formula for such sums:

(2.2) $$x_n y_n - x_0 y_0 = \sum_{k=1}^{n} x_{k-1}(y_k - y_{k-1}) + \sum_{k=1}^{n} y_{k-1}(x_k - x_{k-1}) + \sum_{k=1}^{n} (x_k - x_{k-1})(y_k - y_{k-1})$$

suggests the fundamental *integration-by-parts formula for semimartingales*:

(2.3) $$X_t Y_t - X_0 Y_0 = \int_{(0,t]} X_{s-}\,dY_s + \int_{(0,t]} Y_{s-}\,dX_s + \int_{(0,t]} dX_s\,dY_s.$$

But what sense are we to make of the last term in (2.3)?

It is easy to believe (and to prove!) that if X and Y have paths of finite variation, then the correct interpretation is as follows:

(2.4) $$\int_{(0,t]} dX_s\,dY_s = \sum_{0<s\leq t} (X_s - X_{s-})(Y_s - Y_{s-}).$$

What happens if X and Y are path-continuous martingales? To gain insight into this situation, and into more general situations, we again look at a discrete analogue.

(2.5) *A discrete-time analogue.* Let $M = \{M_n\colon n \geq 0\}$ and $N = \{N_n\colon n \geq 0\}$ be martingales on a set-up $(\Omega, \mathscr{F}, \{\mathscr{F}_n\}, \mathbf{P})$ such that for all n, $\mathbf{E}M_n^2 < \infty$, $\mathbf{E}N_n^2 < \infty$. Take $x_k = M_k(\omega)$, $y_k = N_k(\omega)$ in (2.2) to see that

$$M_n N_n - M_0 N_0 = \sum_{k=1}^{n} M_{k-1}(N_k - N_{k-1}) + \sum_{k=1}^{n} N_{k-1}(M_k - M_{k-1}) + \sum_{k=1}^{n} \Delta M_k \Delta N_k,$$

where $\Delta M_k = M_k - M_{k-1}$ and $\Delta N_k = N_k - N_{k-1}$. Now (compare (1.6)):

$$U_n = \sum_{k=1}^{n} M_{k-1}(N_k - N_{k-1})$$

defines a martingale U because

$$\mathbf{E}[U_n - U_{n-1} \mid \mathscr{F}_{n-1}] = M_{n-1}\mathbf{E}[N_n - N_{n-1} \mid \mathscr{F}_{n-1}] = 0.$$

Hence, if we put

$$[M, N]_n = \sum_{k=1}^{n} \Delta M_k \Delta N_k,$$

then

$$M_n N_n - M_0 N_0 - [M, N]_n \textit{ is a martingale.}$$

In particular, on taking $N=M$, we see that

$$V_n = M_n^2 - M_0^2 - \sum_{k=1}^{n} (M_k - M_{k-1})^2 \textit{ is a martingale}$$

because V is the 'stochastic integral':

$$V_n = \sum_{k=1}^{n} 2M_{k-1}(M_k - M_{k-1}). \qquad \square$$

With the above discrete-time analogue very much in mind, we shall now develop the interpretation of (2.3) for the case when $X=Y=B$, a Brownian motion on $\mathbb{R}$. This is the only case we shall consider in this section. Since we shall be providing all details for the general continuous local martingale later, a sketched argument will suffice here. Let us write for $t \geqslant 0$,

$$t_k^n \equiv t \wedge (k2^{-n}).$$

We have:

$$B(t)^2 - B(0)^2 = \sum_{k\geqslant 1} 2B(t_{k-1}^n)(B(t_k^n) - B(t_{k-1}^n)) + [B]_t^n,$$

where

(2.6) $$[B]_t^n \equiv \sum_{k\geqslant 1} (B(t_k^n) - B(t_{k-1}^n))^2.$$

For each t, there are only finitely many non-zero terms in (2.6), and these are independent random variables. Moreover, since $B_{s+h} - B_s$ has the normal distribution of mean zero and variance h,

$$B_{s+h} - B_s \sim N(0, h),$$

we have:

$$\mathbf{E}[(B_{s+h} - B_s)^2] = h, \quad \operatorname{var}[(B_{s+h} - B_s)^2] = h^2 \operatorname{var}(\xi^2),$$

where $\xi \sim N(0, 1)$. Hence, for fixed t, $[B]_t^n$ has mean t and variance $O(2^{-n})$, so that:

(2.7) $$[B]_t^n \to t \text{ with probability } 1.$$

More is true. By an obvious modification of our argument for the discrete-time analogue, we can show that for each n,

(2.8) $$V_t^n \equiv B(t)^2 - B(0)^2 - [B]_t^n = \sum_{k=1}^{n} 2B(t_{k-1}^n)(B(t_k^n) - B(t_{k-1}^n))$$

is a martingale. But

$$V_t^n - V_t^{n+1} = [B]_t^{n+1} - [B]_t^n,$$

so that $[B]^{n+1} - [B]^n$ is a martingale. Doob's L^2 inequality (II.52.6) now allows us to strengthen (2.7) to the following result, quoted at I.11.1:

(*2.9*) (*Lévy's quadratic variation result*). *With probability* 1, $[B]^n_t \to t$ *uniformly on compact intervals.*

See Exercise 2.17. It now follows from (2.8) that, with probability 1,

(2.10) $$\lim_n \sum_{k=1}^{n} 2B(t^n_{k-1})(B(t^n_k)-B(t^n_{k-1})) = (B(t)^2-t)-(B(0)^2-0),$$

uniformly on compact intervals.

The correct interpretation of (2.3) when $X=Y=B$ is now clear:

(*2.11*) *The Itô integral* $\int_{(0,t]} 2B_{s-}\,dB_s$ *is the martingale*:

(2.12) $$\int_{(0,t]} 2B_{s-}\,dB_s = (B(t)^2-t)-(B(0)^2-0), \text{ and}$$

(2.13) $$\int_{(0,t]} (dB_s)^2 = [B]_t = t.$$

We sometimes write (2.13) in differential notation:

(2.14) $$(dB_t)^2 = dt.$$

Of course, since B is a continuous process, the B_{s-} in (2.12) is equal to B_s. However, it was worth leaving the '$s-$' in (2.12) to emphasize that the integral must be interpreted as the limit at (2.10). This situation is quite different from that in Riemann integration. Thus it is tautological that:

(2.15) $$\lim \sum (B(t^n_{k-1})+B(t^n_k))(B(t^n_k)-B(t^n_{k-1})) = B(t)^2 - B(0)^2,$$

which corresponds to the Stratonovich integral (written in this book with a ∂):

$$\int_{(0,t]} 2B_s\,\partial B_s = B(t)^2 - B(0)^2.$$

The Stratonovich integral is useful, particularly in applications to stochastic differential geometry. But the Itô integral is the one we need at present because (as you are aware by now!) the Itô integral preserves local martingales. Until further notice, we shall deal only with the Itô integral.

The reason for the discrepancy between (2.10) and (2.15) is of course that the Brownian motion path has non-zero 'quadratic variation' $[B]$. As was explained at I.11, this implies:

(*2.16*) LEMMA. *Almost all Brownian motion paths have infinite variation on every time interval.*

Proof. Suppose that b is a continuous function of bounded variation on $[0, t]$. Then:

$$\sum_{k=1}^{n}(b(t_k^n)-b(t_{k-1}^n))^2 \leqslant S^n(t)V^n(t)$$

where

$$S^n(t) \equiv \sup_k |b(t_k^n)-b(t_{k-1}^n)|,$$

$$V^n(t) \equiv \sum_k |b(t_k^n)-b(t_{k-1}^n)|.$$

Now $V^n(t)$ is bounded by the total variation of b on $[0, t]$, and since b is uniformly continuous on $[0, t]$, $S^n(t) \to 0$ as $n \to \infty$. Hence:

$$\lim \sum (b(t_k^n)-b(t_{k-1}^n))^2 = 0,$$

for all t. This contradicts (2.9). □

(*2.17*) **Exercise.** Prove (2.9). *Hint.* Let $U_n(t) \equiv [B]_t^{n+1} - [B]_t^n$. Then, by II.43.3,

$$\mathbf{P}[\sup_{s\leqslant t}|U_n(s)| \geqslant \varepsilon] \leqslant \varepsilon^{-2}\mathbf{E}[\sup_{s\leqslant t}|U_n(s)|^2] \leqslant 4\varepsilon^{-2}\mathbf{E}[U_n(t)^2].$$

Show that $\mathbf{E}[U_n(t)^2] = O(2^{-n})$. □

3. Itô's formula for Brownian motion. The celebrated *Itô's formula for Brownian motion* says that for $f \in C^2$,

(3.1) $$f(B_t)-f(B_0) = \int_{(0,t]} f'(B_s)\,dB_s + \int_{(0,t]} \tfrac{1}{2}f''(B_s)\,ds.$$

This will be derived from (2.11), to which it reduces when $f(x)=x^2$. The first integral in (3.1), an Itô integral, is a suitable limit:

(3.2) $$\lim \sum f'(B(t_{i-1}^n))(B(t_i^n)-B(t_{i-1}^n)).$$

The second integral in (3.1) is a Riemann integral for each ω.

We often write (3.1) in differential form:

(3.3) $$df(B_t) = f'(B_t)dB_t + \tfrac{1}{2}f''(B_t)dt,$$

and regard (3.3) as a Taylor expansion with the rules:

(3.4) $$(dB_t)^2 = dt, \quad (dB_t)^3 = 0.$$

A useful extension of (3.3) is the following: for $f \in C^{1,2}$:

(3.5) $$df(t,B_t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial B}dB_t + \tfrac{1}{2}\frac{\partial^2 f}{\partial B^2}dt,$$

so that we need to supplement (3.4) with the rules:

(3.6) $$(dt)^2=0, \quad dtdB=0.$$

For calculations and estimates, we need martingales, and Itô's formula can provide us with them. We shall see many illustrations of this.

We saw in McKean's proof (§ I.16) of the iterated logarithm law and in calculation of the distribution of hitting times in § I.9 how important it is that, for $\theta \in \mathbb{R}$, the '*Brownian exponential*'

$$M_t^\theta = \exp(\theta B_t - \tfrac{1}{2}\theta^2 t)$$

is a martingale. If we apply (3.5) with $f(t, x) = \exp(\theta x - \frac{1}{2}\theta^2 t)$, we obtain:

$$dM_t^\theta = M_t^\theta(-\tfrac{1}{2}\theta^2 dt + \theta dB_t + \tfrac{1}{2}\theta^2 dt) = \theta M_t^\theta dB_t,$$

so that M_t^θ, as a stochastic integral relative to B, is certainly a local martingale. How to prove by Itô calculus that it is a true martingale, we shall see later.

Exponential martingales have a very prominent rôle in the theory, and it is important to be able to solve equations such as

$$dM_t^\theta = \theta M_t^\theta dB_t, \quad M_0^\theta = 1.$$

We show how to solve such equations in § 19 for the finite variation case, and in § 37 for the continuous case. See Doléans [2] for the general case.

4. A rough plan of the chapter. The remainder of the chapter is divided into five parts as follows.

Part 2 introduces some essential concepts and methods. Previsible processes make their first appearance as the integrands of our theory. (We shall get to know them better, and see them feature in other starring rôles, in Chapter VI.) Fundamental Lemma 5.3 establishes the most basic result on preservation of the local martingale property by Itô integration. Extensions of this lemma, either by continuity arguments or by 'localization', are what lead to the Fundamental Theorem 1.5. The substantial portion of Part 2 devoted to localization is absolutely essential. We have tried to make it run as smoothly as possible, so don't shirk your duty!

Part 3 develops the elementary stochastic calculus of finite-variation processes to a point where we can do calculations relating to a Markov chain with finite state–space. This should help attune your intuition to a new way of thinking about results you know well, and thereby sharpen it for more challenging things. (You could skip Part 3 on a first reading, but we would not recommend this.) There is a lot more to the calculus of finite-variation processes, as we shall see when we come to study dual previsible projections in Chapter VI.

Part 4 presents the so-called L^2 theory of stochastic integrals. This superbly elegant theory due to Kunita and (S.) Watanabe extends the Fundamental Lemma 5.3 by exploiting various Hilbert-space isometries.

Part 5 completes the development of the stochastic integral with respect to a continuous semimartingale, and proves Itô's formula for this context.

Part 6, 'Applications of Itô's formula', contains Lévy's Theorem, the Cameron–Martin–Girsanov Theorem, Tanaka's formula, Trotter's Theorem, and many others of the results which make the subject come to life.

2. SOME FUNDAMENTAL IDEAS: PREVISIBLE PROCESSES, LOCALIZATION, etc

Previsible processes

5. Basic integrands $Z(S, T]$. An Itô integral will be a *process*

$$H\bullet X = \int H_s \, dX_s,$$

where H is an '*integrand*', and X is an '*integrator*'. All integrators will be special kinds of R-processes. We shall see that:

The only sensible integrands are previsible processes.
The only sensible integrators are semimartingales.

As is usual in integration theory, we build things up in stages, starting with integrands of a particularly simple type, namely, step-functions (linear combinations of indicator functions). Of course, for our stochastic step-functions, both the intervals of constancy and the values taken on those intervals must be allowed to depend on ω in an intelligent way.

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbf{P})$ be a filtered probability space satisfying the usual conditions. Let S and T be stopping times, $S \leqslant T$, and let $Z \in \mathrm{b}\mathscr{F}_S$. (Recall that $\mathscr{F}_S$ is the pre-S σ-algebra, and $\mathrm{b}\mathscr{F}_S$ is the space of bounded $\mathscr{F}_S$-measurable functions on Ω.) The most basic type of integrand is the process

$$H = Z(S, T]. \tag{5.1}$$

This signifies that, for $t < \infty$,

$$H(t, \omega) = \begin{cases} Z(\omega) & \text{if } S(\omega) < t \leqslant T(\omega), \\ 0 & \text{otherwise.} \end{cases}$$

Note that $T(\omega)$ may be infinite. Then for any integrator X we make the obvious definition:

$$(H\bullet X)_t \equiv \int_{(0,t]} H_s \, dX_s \equiv \int_0^t H_s \, dX_s \equiv Z(X_{T\wedge t} - X_{S\wedge t}). \tag{5.2}$$

In full, (5.2) means:

$$(H\bullet X)(t, \omega) = Z(\omega)(X_{T(\omega)\wedge t}(\omega) - X_{S(\omega)\wedge t}(\omega)).$$