

**ITÔ'S LEMMA**

Learning Objectives:

- Understand the definition of stochastic integrals and appreciate how the definition of Riemann integrals is being mimicked here.
- Be familiar with basic computations with stochastic integrals.
- Understand the basic properties of a stochastic integral.
- Be aware of the concept of the quadratic variation.
- Understand the meaning of a stochastic differential equation (SDE) - it represents equations involving stochastic integrals.
- Know thoroughly Ito's Lemma - what it means, how it is derived (as described in this lecture) and its applications. It is one key ingredient in the mathematical derivation of the Black-Scholes partial differential equation for vanilla option price.

## 1. STOCHASTIC INTEGRALS

We'll see that a solution to a generic SDE such as

$$dX = b(X, t)dt + B(X, t)dW,$$

$$X(0) = X_0$$

is of the form

$$X(t) = X_0 + \int_0^t b(X, s) ds + \int_0^t B(X, s) dW.$$

Thus, we see that in order to even discuss what a solution of an SDE means, we need to define the stochastic integral

$$\int_0^T B dW,$$

for  $B$  being some stochastic process.

The stochastic integral

$$\int_0^T G(t) dW$$

will be defined as a Riemann sum. It will turn out that  $\int_0^T G(t) dW$  is a random variable, and

$$s \mapsto \int_0^s G(t) dW$$

is a stochastic process.

Technical Preamble

Information up till time  $t$  is represented by  $\mathcal{F}_t$ . It is actually a sub- $\sigma$ -algebra of the  $\sigma$ -algebra of events  $\mathcal{F}$  from the underlying probability space.

Conditional expectation, conditional on  $\mathcal{F}_t$ , will be denoted by  $E[-|\mathcal{F}_t]$  - it is an  $\mathcal{F}_t$ -measurable random variable.

We will assume that the stochastic process  $G_t$  ( $t \in [0, T]$ ) which plays the role of the integrand in the stochastic integral is *non-anticipating*

(with respect to  $W$ ), i.e.  $G_t$  is  $\mathcal{F}_t$ -measurable for all  $t \in [0, T]$ . Intuitively, this means that the stochastic process  $G$  depends only on the information available in the sigma-algebra  $\mathcal{F}_t$  and not on the future of the Wiener process beyond time  $t$ .

Even more stringently, we will assume that the stochastic processes  $G$  are *progressively measurable*. This roughly means that  $G$  is non-anticipating and jointly measurable in both  $t$  and  $\omega$  (so that it can be integrated over time and over the probability space). To avoid infinity, we also assume that  $E[\int_0^T G^2 dt] < \infty$ .

Stochastic integrals of progressively measurable stochastic processes have nice properties (as we shall see) and they are first defined for step processes and extended to general processes by the passage of limits.

### Defining the Stochastic Integral

Suppose the time horizon is  $[0, T]$ . Let  $P$  be a partition

$$0 = t_0 < t_1 < \dots < t_m = T.$$

Let the gap size of  $P$  be

$$|P| = \max_i (t_i - t_{i-1}).$$

For each  $\lambda \in [0, 1]$  and the partition  $P$ , set

$$\tau_k = (1 - \lambda)t_k + \lambda t_{k+1}$$

and define

$$R(P, \lambda) = \sum_{k=0}^{m-1} G(\tau_k)(W(t_{k+1}) - W(t_k)),$$

We're interested in what the limit is when  $|P| \rightarrow 0$ . It turns out that, different from the classical non-stochastic case, here, each choice of  $\lambda$  will give rise to a specific class of stochastic integrals -  $\lambda = 0$  gives the Itô Integrals while  $\lambda = \frac{1}{2}$  gives the Stratonovich Integrals.

The fact that  $\lambda = 0$  and  $\tau_k = t_k$  in the case of Itô allow us to define the stochastic integral  $\int_0^T G dW$  for *nonanticipating* processes. The idea is that since we do not know what  $W$  would do on the interval  $[t_k, t_{k+1}]$ , it is best to use  $G(t_k)$  in the approximation.

We'll be concerned only with the Itô Integrals in this course. We'll shorten  $R(P, \lambda)$  to  $R(P)$ .

The limit is taken in the sense of expectation-square: we say that a sequence of random variables  $X_n$  converges to  $X$  if

$$\lim_{n \rightarrow \infty} E[(X - X_n)^2] = 0.$$

**Example 1.** Find the stochastic integral

$$\int_0^T dW_t.$$

*Solution.* We have

$$\begin{aligned} R(P, \lambda) &= (W(t_1) - W(t_0)) + (W(t_2) - W(t_1)) + \cdots + (W(t_n) - W(t_{n-1})) \\ &= W(T) - W(0) = W(T). \end{aligned}$$

Hence the integral  $\int_0^T dW_t$  is the normal distribution with mean 0 and variance  $T$ . Thus this integral is independent of  $\lambda$ .  $\square$

**Example 2.** Find the stochastic integral

$$\int_0^T W_t dW_t.$$

*Solution.* Let  $P^n$  be the partition of  $[0, T]$

$$0 = t_0^n < t_1^n < \cdots < t_{m_n}^n = T,$$

so that  $\lim_{n \rightarrow \infty} |P^n| = 0$ .

We have

$$\begin{aligned} R(P^n) &= W(t_0^n)(W(t_1^n) - W(t_0^n)) + W(t_1^n)(W(t_2^n) - W(t_1^n)) \\ &\quad + \cdots + W(t_{m_n-1}^n)(W(t_{m_n}^n) - W(t_{m_n-1}^n)) \\ &= \frac{1}{2}W(T)^2 - \frac{1}{2} \sum_{k=0}^{m_n-1} (W(t_{k+1}^n) - W(t_k^n))^2. \end{aligned}$$

We'll work out the term

$$T_n^1 := \sum_{k=0}^{m_n-1} (W(t_{k+1}^n) - W(t_k^n))^2.$$

We compute:

$$\begin{aligned} T_n^1 - T &= \sum_{k=0}^{m_n-1} ((W(t_{k+1}^n) - W(t_k^n))^2 - (t_{k+1}^n - t_k^n)), \\ E[(T_n^1 - T)^2] \\ &= \sum_{k=0}^{m_n-1} \sum_{j=0}^{m_n-1} E[((W(t_{k+1}^n) - W(t_k^n))^2 - (t_{k+1}^n - t_k^n))((W(t_{j+1}^n) - W(t_j^n))^2 - (t_{j+1}^n - t_j^n))]. \end{aligned}$$

When  $j \neq k$ , the two terms in the product of the summand is

$$E[((W(t_{k+1}^n) - W(t_k^n))^2 - (t_{k+1}^n - t_k^n))] \times E[\dots] = 0,$$

by the property of Brownian motions. Hence

$$\begin{aligned} E[(T_n^1 - T)^2] \\ &= \sum_{k=0}^{m_n-1} E[((W(t_{k+1}^n) - W(t_k^n))^2 - (t_{k+1}^n - t_k^n))^2] \\ &= \sum_{k=0}^{m_n-1} (t_{k+1}^n - t_k^n)^2 E\left[\left(\frac{W(t_{k+1}^n) - W(t_k^n)}{\sqrt{t_{k+1}^n - t_k^n}}\right)^2 - 1\right]^2 \\ &\leq C \sum_{k=0}^{m_n-1} (t_{k+1}^n - t_k^n)^2 \\ &\leq C|P^n|(b-a) \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ . Hence

$$T_n^1 \rightarrow T$$

as  $n \rightarrow \infty$ .

Therefore, we conclude that as  $n \rightarrow \infty$ ,

$$\lim_{n \rightarrow \infty} R(P^n) = \frac{1}{2}W(T)^2 - \frac{1}{2}T.$$

□

**Example 3.** Show that the Itô Integral

$$\int_r^s W dW = \frac{W^2(s) - W^2(r)}{2} - \frac{s - r}{2}.$$

At this juncture, it is tempting to want to evaluate  $\int_0^T W^n dW$ . But we have just seen how tedious the case  $n = 1$  already is. Just as the derivative helps in integral problems (by noticing that integration is opposite to differentiation by the Fundamental Theorem of Calculus, and since differentiation is straightforward), we'll make use of Itô's Lemma later to find  $\int_0^T W^n dW$ .

**Example 4.** *Describe the stochastic integral*

$$\int_0^T t dW_t$$

*as explicitly as you can.*

**Example 5.** *Describe the integral*

$$\int_0^T tW_t dt$$

*as explicitly as you can. (Hint: Consider the expression*

$$\frac{1}{2}((t + \delta t)^2 - t^2)W_t - \frac{1}{2}\delta t^2 W_t = tW_t \delta t.)$$

## 2. PROPERTIES OF THE STOCHASTIC INTEGRAL

Let us record down some facts that show that the stochastic integral behaves well with respect to addition, scalar multiplication and under expectation.

**Theorem 1.** For all constants  $a, b \in \mathbb{R}$  and for all  $G, H$  which are progressively measurable and square integrable, we have

(1)

$$\int_0^T (aG + bH) dW = a \int_0^T G dW + b \int_0^T H dW;$$

(2)

$$E\left[\int_0^T G dW\right] = 0;$$

(3)

$$E\left[\left(\int_0^T G dW\right)^2\right] = E\left[\int_0^T G^2 dt\right];$$

(4)

$$E\left[\left(\int_0^T G dW\right)\left(\int_0^T H dW\right)\right] = E\left[\int_0^T GH dt\right].$$

*Proof.* The linearity property follows from the linearity property of the Riemann sum and the compatibility between arithmetic operations and taking limits.

We'll assume  $G$  and  $H$  simple - the general case is then just a matter of applying limits to both sides of the equation.

Suppose  $G(t) \equiv G_k$  and  $H(t) \equiv H_k$  for  $t \in [t_k, t_{k+1})$ , and  $G_k, H_k$  are  $\mathcal{F}_k$ -measurable.

Case 2: Then

$$E\left[\int_0^T G dW\right] = \sum_{k=0}^{m-1} E[G_k(W(t_{k+1}) - W(t_k))].$$

Since  $G_k$  is  $\mathcal{F}_{t_k}$ -measurable, which is independent from  $W_{t_{k+1}} - W_{t_k}$ , therefore

$$E[G_k(W(t_{k+1}) - W(t_k))] = E[G_k]E[W(t_{k+1}) - W(t_k)] = 0.$$

Case 3 is a special case of Case 4.

Case 4: We have

$$E\left[\int_0^T G dW \int_0^T H dW\right] = \sum_{k,j=1}^{m-1} E[G_j H_k (W(t_{k+1}) - W(t_k))(W(t_{j+1}) - W(t_j))].$$

Now if  $j < k$ ,

$$\begin{aligned} & E[G_j H_k (W(t_{k+1}) - W(t_k))(W(t_{j+1}) - W(t_j))] \\ &= E[G_j H_k (W(t_{j+1}) - W(t_j))] E[(W(t_{k+1}) - W(t_k))] = 0 \end{aligned}$$

for the same reason of independence. Hence,

$$\begin{aligned} E\left[\int_0^T G dW \int_0^T H dW\right] &= \sum_{k=1}^{m-1} E[G_k H_k (W(t_{k+1}) - W(t_k))^2] \\ &= \sum_{k=1}^{m-1} E[G_k H_k] E[(W(t_{k+1}) - W(t_k))^2] = \sum_{k=1}^{m-1} E[G_k H_k] (t_{k+1} - t_k) \\ &= E\left[\sum_{k=1}^{m-1} G_k H_k (t_{k+1} - t_k)\right] = E\left[\int_0^T GH dt\right]. \end{aligned}$$

□

The integral that is usually used in QF is the Ito integral - one reason is that the process  $I_t := \int_0^t G dW$  is a martingale.

**Example 6.** Show that  $I_t := \int_0^t G dW$  is a martingale if  $G$  is a progressively measurable function for which the stochastic integral exists.

*Solution:*

We need to show that

$$E[I_T | \mathcal{F}_t] = I_t.$$

Write

$$\begin{aligned} I_T &= I_t + \int_t^T G_s dW_s \\ &= I_t + \int_t^T G_s d(W_s - W_t) \\ &= I_t + \int_0^{T-t} \tilde{G}_u d\tilde{W}_u, \end{aligned}$$

where  $\tilde{W}_u := W_{u+t} - W_t$ , which is Brownian, and  $\tilde{G}_u := G_{u+t}$ , which is progressively measurable, with respect to the filtration  $\mathcal{F}_{u+t}$ ,  $u \geq 0$ .

By the above Theorem,

$$E\left[\int_t^T G_s dW_s \mid \mathcal{F}_t\right] = E\left[\int_0^{T-t} \tilde{G}_u d\tilde{W}_u\right] = 0.$$

Hence,

$$E[I_T \mid \mathcal{F}_t] = I_t.$$

The following concept plays a fundamental role in the theory behind variance derivatives.

**Definition 1.** Let  $f : [0, T] \rightarrow \mathbb{R}$  be a function. The quadratic variation of  $f$  up to time  $T$  is defined to be

$$[f, f](T) = \lim_{|\Pi| \rightarrow 0} \sum_{i=1}^n (f(t_i) - f(t_{i-1}))^2,$$

where  $\Pi : 0 = t_0 < t_1 < \dots < t_n = T$  is a partition of  $[0, T]$  and  $|\Pi|$  is the mesh size of  $\Pi$ .

The map  $t \mapsto [f, f](t)$  is a function on  $[0, T]$  and for a stochastic process  $(X_t, t \in [0, T])$ , its quadratic variation  $([X, X]_t, t \in [0, T])$  is a stochastic process over  $[0, T]$ .

**Proposition 1.** Let  $W$  denote a Brownian motion over  $[0, T]$ . Show that

$$[W, W](T) = T$$

almost surely.

*Proof.* Let

$$Q_\Pi = \sum_{i=1}^n (W(t_i) - W(t_{i-1}))^2.$$

We have

$$E[Q_\Pi] = \sum_{i=1}^n E[(W(t_i) - W(t_{i-1}))^2] = \sum_{i=1}^n (t_i - t_{i-1}) = T.$$

We have

$$\begin{aligned}
 \text{Var}[Q_{\Pi}^2] &= \sum_{i=1}^n \text{Var}[(W(t_i) - W(t_{i-1}))^2] \\
 &= \sum_{i=1}^n E[((W(t_i) - W(t_{i-1}))^2 - t_i + t_{i-1})^2] \\
 &= \sum_{i=1}^n E[(W(t_i) - W(t_{i-1}))^4] - 2(t_i - t_{i-1})E[(W(t_i) - W(t_{i-1}))^2] + (t_i - t_{i-1})^2 \\
 &= \sum_{i=1}^n 3(t_i - t_{i-1})^2 - 2(t_i - t_{i-1})^2 + (t_i - t_{i-1})^2 \\
 &= \sum_{i=1}^n 2(t_i - t_{i-1})^2 \\
 &\leq 2|\Pi| \sum_{i=1}^n (t_i - t_{i-1}) \\
 &\quad 2|\Pi|T \rightarrow 0
 \end{aligned}$$

as  $|\Pi| \rightarrow 0$ .

Hence,  $[W, W](T) = T$  almost surely.

□

*Remark:* The first-order or linear variation is the simply the variation. We showed earlier that the Brownian motion does not have bounded variation. The proposition shows that it however has bounded second variation.

**Example 7.** Show that the quadratic variation of the stochastic integral

$$I_t := \int_0^t G_s dW_s \quad (t \in [0, T])$$

up to time  $t$  is

$$[I, I](t) = \int_0^t G_s^2 ds.$$

## 3. ITO'S FORMULA/LEMMA

Suppose  $X$  is a stochastic process which satisfies

$$X(s) = X(r) + \int_r^s F dt + \int_r^s G dW$$

for all  $0 \leq r \leq s \leq T$ . Then we say that  $X$  satisfies the stochastic differential equation

$$dX = F dt + G dW$$

for all  $0 \leq t \leq T$ .

Note that the symbols  $dX$ ,  $dt$  and  $dW$  has no formal meaning - they are just alternative expressions for the stochastic integral equation.

**Theorem 2** (Ito's Formula/Lemma). *Suppose  $X$  satisfies*

$$dX = Fdt + GdW.$$

*Suppose  $u : \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$  is infinitely differentiable. Let*

$$Y(t) := u(X(t), t).$$

*Then  $Y$  satisfies*

$$dY = \left( \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} F + \frac{1}{2} \frac{\partial^2 u}{\partial x^2} G^2 \right) dt + \frac{\partial u}{\partial x} G dW.$$

Before we can prove the Formula, we need to prove establish simpler cases first.

**Lemma 1.**  $d(W^2) = 2W dW + dt$

*Proof.* This actually means

$$W^2(r) - W^2(s) = 2 \int_r^s W dW + (s - r).$$

We have displayed this equation in an example from the last section.  $\square$

**Lemma 2.**  $d(tW) = W dt + t dW$

*Proof.* We use notation on partitions from the previous section.

From first principles,

$$\int_0^r t dW = \lim_{n \rightarrow \infty} \sum_{k=0}^{m_n-1} t_k^n (W(t_{k+1}^n) - W(t_k^n)),$$

$$\int_0^r W dt = \lim_{n \rightarrow \infty} \sum_{k=0}^{m_n-1} W(t_{k+1}^n) (t_{k+1}^n - t_k^n),$$

the second equation being the theorem about the convergence of ordinary Riemann sums to the Riemann integral, since

$$t \mapsto W(t)$$

is a continuous function. Adding the two equations together, one gets

$$\int_0^r t dW + \int_0^r W dt = rW(r).$$

□

**Theorem 3** (Ito Product Rule). *Suppose*

$$\begin{cases} dX_1 = F_1 dt + G_1 dW \\ dX_2 = F_2 dt + G_2 dW \end{cases} \quad (0 \leq t \leq T),$$

for square integrable functions  $F_1, F_2, G_1, G_2$ . Then

$$d(X_1 X_2) = X_2 dX_1 + X_1 dX_2 + G_1 G_2 dt.$$

*Proof.* We will establish this in the case that  $X_1(0) = X_2(0) = 0$  and  $F_i, G_i$  are time-independent. With these assumptions,

$$X_i(t) = F_i t + G_i W(t), \quad (t \geq 0, i = 1, 2).$$

Then

$$\begin{aligned} & \int_0^r X_2 dX_1 + X_1 dX_2 + G_1 G_2 dt \\ &= \int_0^r (X_1 F_2 + X_2 F_1) dt + \int_0^r (X_1 G_2 + X_2 G_1) dW + \int_0^r G_1 G_2 dt \\ & \quad + \int_0^r ((F_1 t + G_1 W) F_2 + (F_2 t + G_2 W) F_1) dt \\ & \quad + \int_0^r ((F_1 t + G_1 W) G_2 + (F_2 t + G_2 W) G_1) dW + G_1 G_2 r \\ &= F_1 F_2 r^2 + (G_1 F_2 + G_2 F_1) \left( \int_0^r W dt + \int_0^r t dW \right) + 2G_1 G_2 \int_0^r W dW + G_1 G_2 r \end{aligned}$$

$$\begin{aligned}
&= F_1 F_2 r^2 + (G_1 F_2 + G_2 F_1) r W(r) + G_1 G_2 W^2(r) \\
&= X_1(r) X_2(r).
\end{aligned}$$

□

*Proof of Ito's Formula/Lemma.* We assume

$$dX = Fdt + GdW.$$

This is done in 3 steps.

Step 1: We first show that

$$d(X^m) = mX^{m-1}dX + \frac{1}{2}m(m-1)X^{m-2}G^2dt,$$

for  $m = 0, 1, 2, \dots$ . This has been done for  $m = 0, 1, 2$ . The general case follows from the Ito Product Rule and induction:

$$\begin{aligned}
d(X^m) &= d(XX^{m-1}) \\
&= Xd(X^{m-1}) + X^{m-1}dX + (m-1)X^{m-2}G^2dt \\
&= X((m-1)X^{m-2}dX + \frac{1}{2}(m-1)(m-2)X^{m-3}G^2dt) \\
&\quad + (m-1)X^{m-2}G^2dt + X^{m-1}dX \\
&= mX^{m-1}dX + \frac{1}{2}m(m-1)X^{m-2}G^2dt.
\end{aligned}$$

Step 2: Next we establish the Formula for  $u(t, x) = f(x)g(t)$ , where  $f$  and  $g$  are polynomials. Then by the Ito Product Rule,

$$\begin{aligned}
d(u(X, t)) &= d(f(X)g(t)) \\
&= f(X)d(g(t)) + g(t)d(f(X)) \\
&= f(X)g'(t)dt + g(t)(f'(X)dX + \frac{1}{2}f''(X)G^2dt) \\
&= \frac{\partial u}{\partial t}dt + \frac{\partial u}{\partial x}dX + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}G^2dt.
\end{aligned}$$

Since the Ito Formula respects linearity, it is now established for all polynomial functions  $u(t, x)$ .

Step 3: A infinitely differentiable function can be approached uniformly by a sequence of polynomial functions in the limit. Hence the Formula also holds for infinitely differentiable functions. □

**Example 8.** Let  $X = W$ ,  $u(x, t) = e^{\lambda x - \frac{\lambda^2 t}{2}}$ ,  $F \equiv 0$ ,  $G \equiv 1$ . Then

$$\begin{aligned} d(e^{\lambda W(t) - \frac{\lambda^2 t}{2}}) &= \left(-\frac{\lambda^2}{2} e^{\lambda W(t) - \frac{\lambda^2 t}{2}} + \frac{\lambda^2}{2} e^{\lambda W(t) - \frac{\lambda^2 t}{2}}\right) dt + \lambda e^{\lambda W(t) - \frac{\lambda^2 t}{2}} dW \\ &= \lambda e^{\lambda W(t) - \frac{\lambda^2 t}{2}} dW. \end{aligned}$$

In other words,  $Y$  satisfies the stochastic differential equation

$$\begin{aligned} dY &= \lambda Y dW, \\ Y(0) &= 1. \end{aligned}$$

This is how you may remember Ito's Lemma in the form of the following heuristic:

$$dt^2 = 0, \quad dt dW = 0, \quad dW^2 = dt.$$

**Example 9.** Find the stochastic differential equation satisfied by  $Y_t = \sin W_t$ .

**Example 10.** Solve (for) the Arithmetic Brownian Motion

$$dS_t = \mu dt + \sigma dW_t.$$

**Example 11.** Solve (for) the Geometric Brownian Motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t.$$

**Example 12.** Consider the SDE

$$dS_t/S_t = \mu_t dt + \sigma_t dW_t,$$

where  $\mu_t, \sigma_t$  are stochastic processes (assumed sufficiently nice). Show that the quadratic variation of  $\log S_t$  over  $[0, T]$  is

$$[\log S, \log S](T) = \int_0^T \sigma_t^2 dt.$$

(For simplicity, assume that  $\mu_t, \sigma_t$  are independent from  $W_t$ ).

This is the link between quadratic variation and variance derivatives that was mentioned earlier. The ‘underlying asset’ of variance derivatives is the term  $\int_0^T \sigma_t^2 dt$ . The equation above shows that this can also be expressed roughly as the sum of squares of changes in  $\log S_t$  (i.e. the daily stock price returns, squared), suitable weighted (theoretically, by  $\delta t = T/n$ ; in practice, this results in averaging). Take a look at the contract specification of variance or volatility derivatives for details.

**Example 13.** *The Vasicek interest rate model is the following:*

$$dr_t = (\alpha - \beta r_t)dt + \sigma dW_t.$$

*Verify that*

$$r_t = e^{-\beta t} r_0 + \frac{\alpha}{\beta} (1 - e^{-\beta t}) + \sigma e^{-\beta t} \int_0^t e^{\beta s} dW_s$$

*solves the model SDE. (Hint: Set  $f(t, x) = e^{-\beta t} r_0 + \frac{\alpha}{\beta} (1 - e^{-\beta t}) + \sigma e^{-\beta t} x$ .)*