Introduction to Stochastic Calculus

Yuewu Xu, Ph.D.

School of Business

Fordham University

Fall, 2022

Overview

- Continuous-time finance is one of the most exciting and most important developments of modern finance. It has fundamentally changed how risks are priced and managed in the real world.
- Main asset pricing ideas from single-period models:
  - Mean-variance portfolio theory: Markowitz (1952).
  - Capital asset pricing model: Sharpe (1964) and Lintner (1965).
  - Arbitrage pricing theory: Ross (1976, 1977).
- Continuous-time finance models extend the insights of single-period models and provide more realistic modeling of financial markets.
- Discrete vs. continuous-time models
  - Multi-period models can be studied in discrete or continuous time.
  - Each approach has its own dis/advantages.
  - Discrete-time models are easier to understand, but sometimes are more difficult to solve.
  - Continuous-time models rely on more complicated mathematics, but can lead to more elegant and powerful solutions.
- Potential topics covered
  - Introduction to stochastic calculus.
  - Option pricing: Black-Scholes model, risk-neutral pricing.
  - Term structure of interest rates (time permits)
  - Credit risk models (time permits).
- Teaching philosophy
  - Focusing on main ideas and sketch of proofs.
  - Focusing on understanding the results: why it is true, and how to use it in practice.
  - Measure-theoretic type of proofs and conditions kept at minimum (if you are interested, read the textbook).

Part 1. Introduction to Stochastic Calculus

Main topics

- Brownian motion
- Stochastic integration
- Ito's formula
- Applications of Ito's formula

1.2 Brownian Motion

Brownian motion is one of the most widely studied continuous-time stochastic processes and is a major building block for continuous-time asset pricing models

A little history: Scatter a few grains of pollen on the surface of an apparently still beaker of liquid. Under a microscope you will observe that each grain is not still but jitters about on the liquid surface. This was first noticed by a botanist called Robert Brown in 1827. He was looking for microscopic life in a drop of water when he noticed that small grains in the water were jiggling around in a strange way - almost as if they were alive! This type of random motion is called Brownian motion after him.

Einstein showed that continual collision with water molecules causes Brownian motion. When he investigated Brownian motion at the beginning of the 20th century not all scientists believed in molecules, and Einstein was trying to demonstrate that they really did exist.

Einstein (1905) listed the following three properties of BM:

- (i) The sample paths must be continuous (based on physics);
- (ii) The increments follow a normal distribution with a variance that is proportion to the time elapsed (based on CLT);

- (iii) The increments of BM are independent, i.e., pollen grain has no memory.

Einstein could not prove that such process exists. In 1920s, Nobert Wiener proved BM exists and hence it is also called Wiener process.

Definition of Brownian Motion. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A family of random variables W_t indexed by time t (assume that $W_0 \equiv 0$) is called a Brownian Motion if it satisfies

- Continuous sample path: For each $\omega \in \Omega$, the function $W_t(\omega)$ is a continuous function of $t \geq 0$
- Independent increments: for all $0 = t_0 < t_1 < ... < t_m$ the incrementsis

  $$W_{t_1} - W_0, W_{t_2} - W_{t_1}, ..., W_{t_m} - W_{t_{m-1}}$$

  are independent of each other.
- Normality: each of the increments is normally distributed

  $$\mathbb{E}\left[W_{t_{i+1}} - W_{t_i}\right] = 0 \text{ and } \mathbb{E}\left[W_{t_{i+1}} - W_{t_i}\right]^2 = t_{i+1} - t_i.$$

A filtration for Brownian motion is defined as $\mathcal{F}_t = \sigma\left(W_s, s \leq t\right)$. That is, all observable event is based on observing the BM before t. Adaptivity means that $W\left(t\right)$ is $\mathcal{F}_t$-measurable. Independence of future increments means that for $0 \leq t < u$, $W\left(u\right) - W\left(t\right)$ is independent of $\mathcal{F}_t$.

A few important properties of Brownian motion

(1). Brownian motion is a martingale.

Proof. Let $0 \leq s < t$ be given. Then

$$\begin{aligned}
& \mathbb{E}\left[W_t | \mathcal{F}_s\right] \\
= & \mathbb{E}\left[W_t - W_s + W_s | \mathcal{F}_s\right] \\
= & \mathbb{E}\left[W_t - W_s | \mathcal{F}_s\right] + \mathbb{E}\left[W_s | \mathcal{F}_s\right] \text{ (linearity property)} \\
= & \mathbb{E}\left[W_t - W_s | \mathcal{F}_s\right] + W_s \text{ (} W_s \text{ is known given } \mathcal{F}_s\text{)} \\
= & \mathbb{E}\left[W_t - W_s\right] + W_s \text{ (independence of increments)} = W_s.
\end{aligned}$$

This property means that the best forecast of tomorrow's value is today's value. This is the idea behind the random walk model

(2). $W_t^2 - t$ is a martingale.

Proof. Let $0 \le s < t$ be given. Then

$$\begin{aligned}
\mathbb{E}\left[W_t^2 - t|\mathcal{F}_s\right] &= \mathbb{E}\left[(W_t - W_s + W_s)^2 - t|\mathcal{F}_s\right] \\
&= \mathbb{E}\left[(W_t - W_s)^2 + 2\left(W_t - W_s\right)W_s + W_s^2 - t|\mathcal{F}_s\right] \\
&= \mathbb{E}\left[(W_t - W_s)^2|\mathcal{F}_s\right] + 2\mathbb{E}\left[\left(W_t - W_s\right)W_s|\mathcal{F}_s\right] \\
&\quad + \mathbb{E}\left[W_s^2 - t|\mathcal{F}_s\right] \\
&= \mathbb{E}\left[(W_t - W_s)^2\right] + 2\mathbb{E}\left[W_t - W_s|\mathcal{F}_s\right]W_s + W_s^2 - t \\
&= (t - s) + W_s^2 - t = W_s^2 - s.
\end{aligned}$$

(3). For $0 \le s < t$, $cov\left(W_s, W_t\right) = s$.

Proof. The covariance of W_s and W_t is

$$\begin{aligned}
\mathbb{E}\left[W_s W_t\right] &= \mathbb{E}\left[\mathbb{E}\left[W_s W_t|\mathcal{F}_s\right]\right] = \mathbb{E}\left[W_s \mathbb{E}\left[W_t|\mathcal{F}_s\right]\right] \\
&= \mathbb{E}\left[W_s^2\right] = s.
\end{aligned}$$

Or

$$\begin{aligned}
\mathbb{E}\left[W_s W_t\right] &= \mathbb{E}\left[W_s\left[W_t - W_s + W_s\right]\right] \\
&= \mathbb{E}\left[W_s\left[W_t - W_s\right] + W_s^2\right] \\
&= \mathbb{E}\left[W_s\right]\mathbb{E}\left[W_t - W_s\right] + \mathbb{E}\left[W_s^2\right] \\
&= 0 + s = s.
\end{aligned}$$

Nondifferentiability of Brownian path

One important property of BM is that is its sample path is not a differentiable as a function of t.

- This can be understand intuitively from the fact $W(t) \sim N(0,t)$.
- The nondifferentiability result is see from the fact the fact that BM has unbounded total variation.

Total Variation [First Variation]. Let $f(t)$ be a function defined for $0 \le t \le T$. If there exists a finite $M > 0$ such that

$$\sum_{j=0}^{n-1} |f(t_{j+1}) - f(t_j)| < M,$$

for all grid $\Pi = \{t_0, t_1, ..., t_n\} : 0 = t_0 < t_1 < ... < t_n = T$, then f is said to have finite Total Variation; The smallest value M is called the Total Variation of f.

We often say the size of the grid is

$$\|\Pi\| = \max_{j=0,\dots,n-1} (t_{j+1} - t_j).$$

Example 1: What is the total variation of $f(x) = x$ for $x \in [0, 2]$.

Example 2: What is the total variation of $f(x) = x^2$ for $x \in [-2, 2]$.

Example 3: What is the total variation of $f(x) = \sin(x)$ for $x \in [0, 2\pi]$.

The Geometric interpretation of Total variation.

It should be noted that the total variation concenpt is generally applicapable to "smooth" functions. For "non-smooth" functions such as the sample path of Brownian Motion, the total variation can be shown to be infinite. In such cases, we need the concept of "quadratic variation".

Quadratic Variation. The quadratic variation of f up to time T is

$$[f, f](T) = \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2.$$

FACT: If f has a continuous derivative, then its quadratic variation is zero.

$$\begin{aligned}\sum_{j=0}^{n-1}\left[f\left(t_{j+1}\right)-f\left(t_j\right)\right]^2 &= \sum_{j=0}^{n-1}\left|f'\left(t_j^*\right)\right|^2\left(t_{j+1}-t_j\right)^2 \\ &\leq \|\Pi\|\cdot\sum_{j=0}^{n-1}\left|f'\left(t_j^*\right)\right|^2\left(t_{j+1}-t_j\right), \quad t_j^*\in\left[t_j,t_{j+1}\right],\end{aligned}$$

$$\begin{aligned}[f,f](T) &\leq \lim_{\|\Pi\|\to 0}\left[\|\Pi\|\cdot\sum_{j=0}^{n-1}\left|f'\left(t_j^*\right)\right|^2\left(t_{j+1}-t_j\right)\right] \\ &= \lim_{\|\Pi\|\to 0}\|\Pi\|\cdot\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left|f'\left(t_j^*\right)\right|^2\left(t_{j+1}-t_j\right) \\ &= \lim_{\|\Pi\|\to 0}\|\Pi\|\cdot\int_0^T\left|f'(t)\right|^2dt=0.\end{aligned}$$

Because $f'(t)$ is continuous, $\int_0^T\left|f'(t)\right|^2dt$ is finite.

Theorem [Quadratic Variation of BM]. Let W be a BM, then $[W,W](T)=T$ for all $T\geq 0$.

Proof. For any finite partition Π of $[0,T]$, denote by $Q=\sum_{j=0}^{n-1}\left(W_{t_{j+1}}-W_{t_j}\right)^2$,

and we want to show that $Q\to T$ as $\|\Pi\|\to 0$. We can show that

$$\mathbb{E}[Q]=T \text{ and } \lim_{\|\Pi\|\to 0}\mathbb{E}[Q-T]^2=0.$$

[Note: This means that Q converges to T in $\mathcal{L}^2$-norm]. We have

$$\begin{aligned}&\mathbb{E}\left[\left(W_{t_{j+1}}-W_{t_j}\right)^2\right]=Var\left[W_{t_{j+1}}-W_{t_j}\right]=t_{j+1}-t_j \\ \Rightarrow\quad &\mathbb{E}[Q]=\mathbb{E}\left[\sum_{j=0}^{n-1}\left(W_{t_{j+1}}-W_{t_j}\right)^2\right]=\sum_{j=0}^{n-1}\left(t_{j+1}-t_j\right)=T.\end{aligned}$$

Moreover,

$$Var\left[\left(W_{t_{j+1}}-W_{t_j}\right)^2\right]=\mathbb{E}\left[\left(\left(W_{t_{j+1}}-W_{t_j}\right)^2-\left(t_{j+1}-t_j\right)\right)^2\right]$$

$$\begin{aligned}=\quad &\mathbb{E}\left[\left(W_{t_{j+1}}-W_{t_j}\right)^4\right]-2\left(t_{j+1}-t_j\right)\mathbb{E}\left[\left(W_{t_{j+1}}-W_{t_j}\right)^2\right] \\ &+\left(t_{j+1}-t_j\right)^2 \quad =2\left(t_{j+1}-t_j\right)^2.\end{aligned}$$

Note $\mathbb{E}\left[\left(W_{t_{j+1}}-W_{t_j}\right)^4\right]=3\left(t_{j+1}-t_j\right)^2$. Therefore,

$$\begin{aligned} Var\left(Q\right) &= \sum_{j=0}^{n-1} Var\left[\left(W_{t_{j+1}}-W_{t_j}\right)^2\right]=\sum_{j=0}^{n-1} 2\left(t_{j+1}-t_j\right)^2 \\ &\leq \sum_{j=0}^{n-1}\left\|\Pi\right\| 2\left(t_{j+1}-t_j\right)=2\left\|\Pi\right\| T. \end{aligned}$$

Therefore, $\lim_{\|\Pi\|\to 0} Var\left(Q\right)=0$ and $\lim_{\|\Pi\|\to 0} Q=\mathbb{E}\left(Q\right)=T$. These two properties imply the result in the Theorem■

Corollary. $\left[W,W\right]\left(T\right)=T\Rightarrow TV\left(W\right)=\infty$ almost surely.

Proof. We prove by contradiction. Suppose $TV\left(W\right)<\infty$, then

$$\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left(W_{t_{j+1}}-W_{t_j}\right)^2 \leq \lim_{\|\Pi\|\to 0}\left[\max_{t_j\in\Pi}\left|W_{t_{j+1}}-W_{t_j}\right|\right]\sum_{j=0}^{n-1}\left|W_{t_{j+1}}-W_{t_j}\right|$$

$$\begin{aligned} &\leq \lim_{\|\Pi\|\to 0}\max_{t_j\in\Pi}\left|W_{t_{j+1}}-W_{t_j}\right|\sum_{j=0}^{n-1}\left|W_{t_{j+1}}-W_{t_j}\right| \\ &\leq \lim_{\|\Pi\|\to 0}\max_{t_j\in\Pi}\left|W_{t_{j+1}}-W_{t_j}\right| TV\left(W\right)=0. \end{aligned}$$

This contradicts with $\left[W,W\right]\left(T\right)=T$.■

In the above derivation, we have used

$$\begin{aligned} \mathbb{E}\left[\left(W_{t_{j+1}}-W_{t_j}\right)^2\right] &= t_{j+1}-t_j, \\ Var\left[\left(W_{t_{j+1}}-W_{t_j}\right)^2\right] &= 2\left(t_{j+1}-t_j\right)^2. \end{aligned}$$

Review: L^2 distance, Second Moment, etc. Convergence in L^2.

Now we provide a brief introduction to the so-called L^2-distance, which will be needed for studying stochastic integrals.

The L^2-distance is an extension to the Euclidean distance in $\mathbb{R}^3$:

$$\|x-y\|=\sqrt{\sum_{i=1}^{3}\left(x_i-y_i\right)^2} \text{ for } x=\left(x_1,x_2,x_3\right), \text{ and } y=\left(y_1,y_2,y_3\right).$$

And we have the familiar concept of convergence in terms of the Euclidean distance.

The L^2-distance is a natural extension of the Euclidean distance to the infinite dimensional space (space of functions):

$$\|f-g\|_{L^2} = \sqrt{\int |f(z)-g(z)|^2 dz}.$$

To reconcile this with the Euclidean distance, just note that integration sign is just a continuous sum. The L^2 distance enjoys most of the properties of the Euclidean distance.

For random variables X and Y, the L^2-distance can be defined as

$$\|X-Y\|_{L^2} = \sqrt{\mathbb{E}|X-Y|^2}.$$

Hence for the study of random variables, the L^2-distance is important, (which is equivalent to analyzing Variances).

Stochastic Integration

In cintinuous-time finance, a basic task is to compute continuous trading gains/losses. Stochastic integration is such a tool.

Assume that we have a trading strategy H_t which represents the number of shares we hold for an asset at any time t. Assume that the price of the asset moves according to a Brownian motion W_t. Then the proft/loss of this continuous trading is represented by stochastic integral $\int_0^T H(t)\, dW(t)$.

To motivate the above, discretize the time interval $[0, T]$ into n small intervals $0 = t_0 < t_1 < \cdots < t_n = T$. It is easy to see that if we only trade at the grid points t_i, then the trading profit/loss is represented by

$$\sum_i H_{t_i} \left(W_{t_{i+1}} - W_{t_i} \right).$$

If we want to allow continuous trading, then we will have to find the limit of the above sum of random variables. If the limit exist (as the grid becomes finer and finer) in some sense, then we can difine this limit as the stochastic intergral

$$\int_0^T H(t)\,dW(t) := \lim \sum_i H_{t_i}\left(W_{t_{i+1}} - W_{t_i}\right). \text{ [in some sense]}$$

Remark: This definition has a clear finance interpretation: If we interpret H as trading strategy, and $W_{t_{i+1}} - W_{t_i}$ as changes in stock price, the the stochastic integral representations the total trading profit/loss.

Remark: For H_t to be a trading strategy, H_t needs to be $\mathcal{F}_t$ measurable. Otherwise the trading strategies is not executable (why?)

Remark: The convergence of the defining sum is in the sense of L^2-distance. It is also called mean-square, quadratic-mean, second moment, etc. This is a "better" distance to use. I gave a short review on what L^2 -distance is in class, so I will not repeat here.

Definition (Stochastic Integral in Ito sense). If

$$\sum_i H_{t_i}\left(W_{t_{i+1}} - W_{t_i}\right)$$

convergence to a random elelemt (variable) I in the sense of L^2-distance:

$$\lim \sum_{t_i \in \Pi} H_{t_i}\left(W_{t_{i+1}} - W_{t_i}\right) \xrightarrow{L^2} I, \text{ as size of grid } \Pi \to 0. \tag{1}$$

then random variable I is called the Ito Integral, and we define

$$\int_0^T H(t)\,dW(t) \equiv I.$$

Note that, by the definition of L^2-distance, (1) can be written in the more familiar form:

$$\mathbb{E}\left[\sum_{t_i \in \Pi} H_{t_i}\left(W_{t_{i+1}} - W_{t_i}\right) - I\right]^2 \to 0, \text{ as grid size } \Pi \to 0.$$

Hence, it only involves computation of second moment of random variables!

Remark: What is the meaning L^2-distance? Why is L^2-distance a good distance is good to work with. Review earlier lectures.

Remark: The limit is well defined, meaning it does not depend on the partition.

Remark: We can NOT define stochastic integrals this way,

$$\int_0^T H_t dW_t = \int_0^T H_t \left(\frac{dW_t}{dt}\right) dt,$$

because the sample path of BM is not differentiable.

Now we look at a typical example (often asked in the job interviews).

Example (Important!): Compute $\int_0^T W(t)\, dW(t)$ from the definition:

We set up a grid t_i as in the definition. Note that here $H(t) = W(t)$. The left end points are

$$\begin{array}{cc} W(0) & \text{if } 0 \le t < \frac{T}{n} \\ W\left(\frac{T}{n}\right) & \text{if } \frac{T}{n} \le t < \frac{2T}{n} \\ & \dots \\ W\left(\frac{(n-1)T}{n}\right) & \text{if } \frac{(n-1)T}{n} \le t < T. \end{array}$$

For ease of notation, we set $W_j = W\left(\frac{jT}{n}\right)$. To evaluate the stochastic integral, we need to study

$$\int_0^T W(t)\, dW(t) = \lim_{n\to\infty} \sum_{j=0}^{n-1} W_j \left[W_{j+1} - W_j\right].$$

Now we proceed to evaluate this limit. Note that the sum can be written as

$$\sum_{j=0}^{n-1} W_j \left[W_{j+1} - W_j\right] = \sum_{j=0}^{n-1} W_{j+1} W_j - \sum_{j=0}^{n-1} W_j^2$$

The first term contains the cross product, so we can use formula

$$ab = \frac{1}{2}a^2 + \frac{1}{2}b^2 - \frac{1}{2}(a-b)^2$$

to rewrite it as as

$$\sum_{j=0}^{n-1} W_{j+1} W_j = \frac{1}{2}\sum_{j=0}^{n-1} W_{j+1}^2 + \frac{1}{2}\sum_{j=0}^{n-1} W_j^2 - \frac{1}{2}\sum_{j=0}^{n-1} (W_{j+1} - W_j)^2,$$

Plus in to conclude that

$$\sum_{j=0}^{n-1} W_j \left(W_{j+1} - W_j\right) = \frac{1}{2}\sum_{j=0}^{n-1} \left[W_{j+1}^2 - W_j^2\right] - \frac{1}{2}\sum_{j=0}^{n-1} \left[W_{j+1} - W_j\right]^2.$$

The first summation is a telescoping sum, and the second term is qudratic variation of BM. So we conclude

$$\begin{aligned}\sum_{j=0}^{n-1} W_j \left(W_{j+1} - W_j\right) &= \frac{1}{2}W_n^2 - \frac{1}{2}\sum_{j=0}^{n-1}\left(W_{j+1} - W_j\right)^2 \\ &\rightarrow \frac{1}{2}W^2(T) - \frac{1}{2}T, \text{ as } n \rightarrow \infty.\end{aligned}$$

Hence we get

$$\int_0^T W(t)\, dW(t) = \frac{1}{2}W^2(T) - \frac{1}{2}T. \blacksquare$$

In ordinary calculus, if g is a differentiable function with $g(0) = 0$, then

$$\int_0^T g(t)\, dg(t) = \int_0^T g(t)\, g'(t)\, dt = \frac{1}{2}g^2(t)\,|_0^T = \frac{1}{2}g^2(T).$$

The extra term $-\frac{1}{2}T$ comes from the nonzero quadratic variation of $W(t)$.

The above result fot the stochastic integral can be recasted as

$$W^2(T) = 2\int_0^T W(t)\, dW(t) + [W, W](T).$$

Note: Although the definition of stochastic integrals in (1) works for general integrand H_t, a popular alternative way of achieving the same goal is to define stochastic integral for simple process first, and then use a limiting process to extend the definition to general processes. This approach has the great advantage that there no limiting process is needed for stochastic integrals with integrands of simple processes. *This is what we will do in the follows.*

Ito Integral for simple process. A simple process is such that there exists a partition $\{t_0, t_1, ..., t_n\}$ of $[0, T]$ for which $H(t)$ is constant in t on each subinterval $[t_j, t_{j+1})$. Assume that $t_k \leq t \leq t_{k+1}$, then the Ito integral for simple process $H(t)$ is defined as

$$\int_0^t H(u)\, dW(u) = \sum_{j=0}^{k-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right] + H_{t_k}\left[W_t - W_{t_k}\right]. \tag{2}$$

Notice that there is no limit process in the definition.

Theorem. The Ito integral for the simple process is a martingale.

Proof. Suppose $0 \leq s \leq t \leq T$, $s \in [t_l, t_{l+1})$, $t \in [t_k, t_{k+1})$, $t_l < t_k$, then

$$\begin{aligned} I(t) &= \int_0^t H(u)\, dW(u) \\ &= \sum_{j=0}^{l-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right] + H_{t_l}\left[W_{t_{l+1}} - W_{t_l}\right] \\ &\quad + \sum_{j=l+1}^{k-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right] + H_{t_k}\left[W_t - W_{t_k}\right]. \end{aligned}$$

We must show

$$\mathbb{E}\left[I(t)\,|\mathcal{F}(s)\right] = I(s) \equiv \sum_{j=0}^{l-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right] + H_{t_l}\left(W_s - W_{t_l}\right).$$

We examine each of the four terms.

The first term $\sum_{j=0}^{l-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right]$ is $\mathcal{F}(s)$ measurable (why?), so

$$\mathbb{E}\left[\sum_{j=0}^{l-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right]|\mathcal{F}(s)\right] = \sum_{j=0}^{l-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right].$$

The second term satisfies

$$\mathbb{E}\left[H_{t_l}\left[W_{t_{l+1}} - W_{t_l}\right]|\mathcal{F}(s)\right] = H_{t_l}\left(\mathbb{E}\left[W_{t_{l+1}}|\mathcal{F}(s)\right] - W_{t_l}\right) = H_{t_l}\left(W_s - W_{t_l}\right).$$

The third and fourth term has zero conditional expectations: for $t_j \geq t_{l+1} > s$,

$$\begin{aligned} \mathbb{E}\left\{H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right]|\mathcal{F}(s)\right\} &= \mathbb{E}\left\{\mathbb{E}\left[H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right]|\mathcal{F}(t_j)\right]|\mathcal{F}(s)\right\} \\ &= \mathbb{E}\left\{H_{t_j}\left(\mathbb{E}\left[W_{t_{j+1}}|\mathcal{F}(t_j)\right] - W_{t_j}\right)|\mathcal{F}(s)\right\} \\ &= \mathbb{E}\left\{H_{t_j}\left(W_{t_j} - W_{t_j}\right)|\mathcal{F}(s)\right\} \\ &= 0. \end{aligned}$$

Therefore,

$$\mathbb{E}\left\{\sum_{j=l+1}^{k-1} H_{t_j}\left[W_{t_{j+1}} - W_{t_j}\right] + H_{t_k}\left[W_t - W_{t_k}\right]|\mathcal{F}(s)\right\} = 0.$$

Puting together, we obtain the martingale property of $I(t)$.

Remark (Important): Because $I(t)$ is a martingale and $I(0) = 0$, we have $\mathbb{E}I(t) = 0$ for all $t > 0$. It follows that $Var\left[I(t)\right] = \mathbb{E}\left[I^2(t)\right]$.

Theorem (Ito Isometry). The Ito integral for the simple process satisfies

$$\mathbb{E}\left[I(t)\right]^2 = \mathbb{E}\int_0^t H^2(u)\, du.$$

Note that the right-hand side can be interpreted as $\mathbf{E}[H]^2$ for suitably defined $\mathbf{E}$.

Proof. **Denote** $D_j = W_{t_{j+1}} - W_{t_j}$ for $j = 0, ..., k-1$ and $D_k = W_t - W_{t_k}$. Then

$$I(t) = \sum_{j=0}^{k} H_{t_j} D_j \text{ and } I^2(t) = \sum_{j=0}^{k} H_{t_j}^2 D_j^2 + 2 \sum_{0 \le i < j \le k} H_{t_i} H_{t_j} D_i D_j.$$

$$\begin{aligned} \mathbb{E}\left[H_{t_j}^2 D_j^2\right] &= \mathbb{E}\left\{\mathbb{E}\left[H_{t_j}^2 D_j^2 | \mathcal{F}_{t_j}\right]\right\} \\ &= \mathbb{E}\left\{H_{t_j}^2 \mathbb{E}\left[D_j^2 | \mathcal{F}_{t_j}\right]\right\} \\ &= \mathbb{E}\left[H_{t_j}^2 \left(t_{j+1} - t_j\right)\right], \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left[H_{t_i} H_{t_j} D_i D_j\right] &= \mathbb{E}\left\{\mathbb{E}\left[H_{t_i} H_{t_j} D_i D_j | \mathcal{F}_{t_j}\right]\right\} \\ &= \mathbb{E}\left\{H_{t_i} H_{t_j} D_i \mathbb{E}\left[D_j | \mathcal{F}_{t_j}\right]\right\} \\ &= \mathbb{E}\left\{H_{t_i} H_{t_j} D_i \cdot 0\right\} = 0. \end{aligned}$$

Therefore,

$$\mathbb{E} I^2(t) = \sum_{j=0}^{k} \mathbb{E}\left[H_{t_j}^2 D_j^2\right] = \sum_{j=0}^{k-1} \mathbb{E}\left[H_{t_j}^2\right]\left(t_{j+1} - t_j\right) + \mathbb{E}\left[H_{t_k}^2\right]\left(t - t_k\right).$$

But H_{t_j} is constant on the interval $[t_j, t_{j+1})$, and hence each term above can be represented by

$$H_{t_j}^2 \left(t_{j+1} - t_j\right) = \int_{t_j}^{t_{j+1}} H^2(u)\, du.$$

Thus

$$\begin{aligned} \mathbb{E} I^2(t) &= \sum_{j=0}^{k-1} \mathbb{E} \int_{t_j}^{t_{j+1}} H^2(u)\, du + \mathbb{E} \int_{t_k}^{t} H^2(u)\, du \\ &= \mathbb{E}\left[\sum_{j=0}^{k-1} \int_{t_j}^{t_{j+1}} H^2(u)\, du + \int_{t_k}^{t} H^2(u)\, du\right] \\ &= \mathbb{E}\left[\int_0^t H^2(u)\, du\right]. \end{aligned}$$

Ito Integral for General Integrands

We now move to define Ito integral $\int_0^T H(t)\, dW(t)$ for integrands $H(t)$ that are allowed to vary continuously and also to jump. That is $H(t)$ is no longer a simple process. We do assume that $H(t)$, $t \ge 0$, is adapted to the filtration $\mathcal{F}(t_j)$ and the square-integrability condition

$$\mathbb{E} \int_0^T H^2(t)\, dt < \infty.$$

Suppose there is a sequence $H_n(t)$ of simple processes such that as $n \to \infty$, these processes "converge" to the continuously varying $H(t)$. Here "converge" means that

$$\lim_{n\to\infty} \mathbb{E} \int_0^T |H_n(t) - H(t)|^2 \, dt = 0.$$

For each $H_n(t)$, the Ito integral $\int_0^t H_n(u) \, du$ has already been defined for $0 \le t \le T$. We define the Ito integral for the continuously varying integrand $H(t)$ by the formula

$$\int_0^t H(u) \, dW(u) = \lim_{n\to\infty} \int_0^t H_n(u) \, dW(u), \quad 0 \le t \le T.$$

Here the limit is in L^2 sense.

Theorem. Ito integral $I(t) = \int_0^t H(u) \, dW(u)$ thus defined have the following properties.

(i) (Continuity). As a function of t, the paths of $I(t)$ are continuous.

(ii) (Adaptivity). For each t, $I(t)$ is $F(t)$-measurable.

(iii) (Linearity) If $I_1(t) = \int_0^t H_1(u) \, dW(u)$ and $I_2(t) = \int_0^t H_2(u) \, dW(u)$, then

$$\begin{aligned} I_1(t) + I_2 &= \int_0^t [H_1(u) + H_2(u)] \, dW(u), \\ cI(t) &= \int_0^t cH(u) \, dW(u) \text{ for constant } c \end{aligned}$$

(iv) (Martingale) $I(t)$ is a martingale.

(v) (Ito Isometry) $\mathbb{E} I^2(t) = \mathbb{E} \int_0^t H^2(u) \, du$.

(vi) (Quadratic Variation) $[I, I](t) = \int_0^t H^2(u) \, du$.

Remark: We computed the stochastic integral of $\int_0^T W_t dW_t$ using definition (1). Now we show that our computation also holds for the alternative definition using simple processes. We need to justify $\lim_{n\to\infty} \mathbb{E} \int_0^T |H_n(t) - W(t)|^2 \, dt = 0$. This is because

$$
\begin{aligned}
\mathbb{E}\int_0^T |H_n(t) - W(t)|^2 \, dt &= \mathbb{E}\sum_j \int_{t_j}^{t_{j+1}} (W_j - W_s)^2 ds \\
&= \sum_j \int_{t_j}^{t_{j+1}} \mathbb{E}(W_j - W_s)^2 ds \\
&= \sum_j \int_{t_j}^{t_{j+1}} (s - t_j) ds \\
&= \sum_j \frac{1}{2}(t_{j+1} - t_j)^2 \to 0.
\end{aligned}
$$

1.4 Ito's Formula

Recall the Newton-Leibniz formula in calculus:

$$
G(x) - G(a) = \int_a^x g(y)dy. \text{ (here } G(x)' = g(x)\text{).}
$$

This formula links the operations of differentiation and integration: one is the "inverse" of the other.

- In one direction, if we can find G, then it offers a way to calculate the integral on the right hand side
- In the other direction, we can express function $G(x)$ through its derivatives $g(x)$.

Question: How to prove the Newton-Leibniz formula?

Sketch of Proof (Important). The idea is to use Taylor expansion or its equivalents for partition $a = x_0 < x_1 < ... < x_n = x$:

$$
\begin{aligned}
G(x) - G(a) &= \sum_{i=0}^{n-1} G(x_{i+1}) - G(x_i) \\
&= \sum_{i=0}^{n-1} G'(\xi_i^*)[x_{i+1} - x_i] \\
&\to \int_a^x g(y)dy. \text{ (here } G(x)' = g(x)\text{).}
\end{aligned}
$$

Hence

$$
G(x) - G(a) = \int_a^x g(y)dy
$$

In continuous time finance, we often need to deal with functions of Brownian Motion (or more generally, functions of Ito processes).

Example. The Black-Scholes option pricing formula for European call options, if found, would be a function that depends on the (current) stock price $S(t)$

$$C(S(t), K, T, r, \sigma) \equiv C(S(t)) \text{ [other parameters are omitted]}$$

Hence if we want to derivative the BS formula, we need to be able to handle functions of BM or functions of Ito processes.

The Goal: find an expression for $f(W(t))$ where $f(x)$ is a smooth function, and $W(t)$ is a BM.

Example. In the example in the previous lecture, we obtained

$$W^2(T) = 2\int_0^T W(t)\, dW(t) + [W, W](T).$$

The above formula is special case of the following Ito formula for BM:

Ito's Formula: If function f is twice continuously differentiable, then

$$f(W(T)) - f(W(0)) = \int_0^T f'(W(u))\, dW(u) + \frac{1}{2}\int_0^T f''(W(u))\, du.$$

The integral form has precise mathematical meaning because we have formally defined Ito integral. The differential form only has intuitive but imprecise meaning

Derivation. Let $\Pi = \{t_0, t_1, ..., t_n\}$ be a partition of $[0, T]$. We can express $f(W_T)$

$$f(W_T) - f(W_0) = \sum_{j=0}^{n-1} \left[f\left(W_{t_{j+1}}\right) - f\left(W_{t_j}\right) \right].$$

Since $f(x)$ is twice continuously differentiable, Taylor expansion gives us that

$$f(x) - f(y) = f'(y)(x - y) + \frac{1}{2} f''(y)(x - y)^2 + R(f, x, y).$$

This yields

$$f(W_T) - f(W_0) = \sum_{j=0}^{n-1} \left[f\left(W_{t_{j+1}}\right) - f\left(W_{t_j}\right) \right]$$

$$= \sum_{j=0}^{n-1} f'\left(W_{t_j}\right)\left(W_{t_{j+1}} - W_{t_j}\right) + \frac{1}{2}\sum_{j=0}^{n-1} f''\left(W_{t_j}\right)\left[W_{t_{j+1}} - W_{t_j}\right]^2 + R$$

If we let $\|\Pi\| \to 0$, then

$$\begin{aligned} & f(W_T) - f(W_0) \\ = \; & \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} f(W_{t_j})\left(W_{t_{j+1}} - W_{t_j}\right) + \frac{1}{2}\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f''\left(W_{t_j}\right)\left[W_{t_{j+1}} - W_{t_j}\right]^2 \\ = \; & \int_0^T f'(W_t)\, dW_t + \frac{1}{2}\int_0^T f''(W_t)\, dt. \end{aligned}$$

Note that we have ignored the error term as it can be justified to converge to 0.

A slight extension of the above Ito's formula is the following:

Theorem (Ito formula for Brownian motion). Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous, and let $W(t)$ be a Brownian motion. Then $\forall$ $T \geq 0$,

$$f(T, W_T) = f(0, W_0) + \int_0^T f_t(t, W_t)\, dt + \int_0^T f_x(t, W_t)\, dW_t + \frac{1}{2}\int_0^T f_{xx}(t, W_t)\, dt. \tag{3}$$

The only difference in the proof is that we need to expand $f(t,x)$ both with respect to t and x.

Derivation. Let $\Pi = \{t_0, t_1, ..., t_n\}$ be a partition of $[0,T]$. For a general function $f(t,x)$, Taylor's expansion says that

$$\begin{aligned} & f(t_{j+1}, x_{j+1}) - f(x_j, x_j) \\ = \; & f_t(t_j, x_j)(t_{j+1} - t_j) + f_x(t_j, x_j)(x_{j+1} - x_j) \\ & + \frac{1}{2} f_{xx}(t_j, x_j)(x_{j+1} - x_j)^2 + f_{tx}(t_j, x_j)(t_{j+1} - t_j)(x_{j+1} - x_j) \\ & + \frac{1}{2} f_{tt}(t_j, x_j)(t_{j+1} - t_j)^2 + R. \end{aligned}$$

We replace x_{j+1} and x_j by $W_{t_{j+1}}$ and W_{t_j}, respectively, and sum:

$$f(W_T) - f(W_0) = \sum_{j=0}^{n-1} \left[f\left(t_{j+1}, W_{t_{j+1}}\right) - f\left(t_j, W_{t_j}\right) \right]$$

$$= \sum_{j=0}^{n-1} f_t\left(t_j, W_{t_j}\right)(t_{j+1}-t_j) + \sum_{j=0}^{n-1} f_x\left(t_j, W_{t_j}\right)\left(W_{t_{j+1}}-W_{t_j}\right)$$

$$+\frac{1}{2}\sum_{j=0}^{n-1} f_{xx}\left(t_j, W_{t_j}\right)\left[W_{t_{j+1}}-W_{t_j}\right]^2$$
$$+\sum_{j=0}^{n-1} f_{tx}\left(t_j, W_{t_j}\right)(t_{j+1}-t_j)\left(W_{t_{j+1}}-W_{t_j}\right)$$
$$+\frac{1}{2}\sum_{j=0}^{n-1} f_{tt}\left(t_j, W_{t_j}\right)[t_{j+1}-t_j]^2 + R$$

As $\|\Pi\| \to 0$, we have

$$\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f_t\left(t_j, W_{t_j}\right)(t_{j+1}-t_j) = \int_0^T f_t\left(t, W_t\right)dt,$$

$$\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1} f_x\left(t_j, W_{t_j}\right)\left(W_{t_{j+1}}-W_{t_j}\right) = \int_0^T f_x\left(t, W_t\right)dW_t,$$

$$\lim_{\|\Pi\|\to 0}\frac{1}{2}\sum_{j=0}^{n-1} f_{xx}\left(t_j, W_{t_j}\right)\left[W_{t_{j+1}}-W_{t_j}\right]^2 = \int_0^T f_{xx}\left(t, W_t\right)dt.$$

Note that, in the third term, it almost like we replace $\left[W_{t_{j+1}}-W_{t_j}\right]^2$ by $(t_{j+1}-t_j)$. This is where the quadratic variation comes in.

The fourth term and the fifth term all converges to 0: For the fourth term

$$\lim_{\|\Pi\|\to 0}\left|\sum_{j=0}^{n-1} f_{tx}\left(t_j, W_{t_j}\right)(t_{j+1}-t_j)\left(W_{t_{j+1}}-W_{t_j}\right)\right|$$

$$\le \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left|f_{tx}\left(t_j, W_{t_j}\right)\right|\cdot(t_{j+1}-t_j)\cdot\left|\left(W_{t_{j+1}}-W_{t_j}\right)\right|$$

$$\le \lim_{\|\Pi\|\to 0}\max_{0\le k\le n-1}\left|\left(W_{t_{k+1}}-W_{t_k}\right)\right|\cdot\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left|f_{tx}\left(t_j, W_{t_j}\right)\right|\cdot(t_{j+1}-t_j)$$

$$= 0\cdot\int_0^T\left|f_{tx}\left(t, W(t)\right)\right|dt = 0.$$

The fifth term is treated similarly:

$$\begin{aligned}
& \lim_{\|\Pi\|\to 0}\left|\frac{1}{2}\sum_{j=0}^{n-1} f_{tt}\left(t_j, W_{t_j}\right)\left(t_{j+1}-t_j\right)^2\right| \\
\leq & \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\frac{1}{2}\left|f_{tt}\left(t_j, W_{t_j}\right)\right|\cdot\left(t_{j+1}-t_j\right)^2 \\
\leq & \frac{1}{2}\lim_{\|\Pi\|\to 0}\max_{0\leq k\leq n-1}\left|t_{k+1}-t_k\right|\cdot\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left|f_{tt}\left(t_j, W_{t_j}\right)\right|\cdot\left(t_{j+1}-t_j\right) \\
= & \frac{1}{2}\cdot 0\cdot\int_0^T\left|f_{tt}\left(t, W_t\right)\right|dt = 0.
\end{aligned}$$

The (higher-order) error terms likewise contribute zero to the final answer.■

Differential form of Ito's formula:

$$df\left(t, W_t\right) = f_t\left(t, W_t\right)dt + f_x\left(t, W_t\right)dW_t + \frac{1}{2}f_{xx}\left(t, W_t\right)dt.$$

Note that the differential form's exact meaning is given by the integral form

$$f\left(T, W_T\right) - f\left(0, W_0\right) = \int_0^T f_t\left(t, W_t\right)dt + \int_0^T f_x\left(t, W_t\right)dW_t + \frac{1}{2}\int_0^T f_{xx}\left(t, W_t\right)dt. \tag{4}$$

As a convenience device, people use the following "multiplication table" to help memorize the operation

	dt	$dW(t)$
dt	0	0
$dW(t)$	0	dt

In other words, people write:

$$dW_t dW_t = dt,\ dt dW_t = dW_t dt = 0,\ dt dt = 0.$$

Example: If $f(x) = \frac{1}{2}x^2$, we have

$$\begin{aligned}
\frac{1}{2}W_T^2 &= f\left(W_T\right) - f\left(W_0\right) \\
&= \int_0^T f'\left(W_t\right)dW_t + \frac{1}{2}\int_0^T f''\left(W_t\right)dt \\
&= \int_0^T W_t dW_t + \frac{1}{2}T
\end{aligned}$$

Hence

$$\int_0^T W_t dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T.$$

Discussions: *It should be noted that, Ito's formula is the key tool for computing the stochastic integrals (as demonstrated by the above example). To do this, you will need to find the appropriate function f such that after applying Ito's formula for $f(W_t)$, the term appears is the stochastic integral you want to compute. If you compare how you compute Riemann integral using Newton-Leibniz formula to stochastic integrals, the requirement for f is the same: it is the anti-derivative of the integrand.*

Note that in some finance books use intuitive treatment of the differential of BM: For example, in Ingersoll (1987), $dW(t)$ is defined as

$$dW(t) = \lim_{\Delta t \to 0} \sqrt{\Delta t}\epsilon, \ \epsilon \sim N(0,1).$$

The above Ito's formula can be extended to the so-called Ito processes.