# Stochastic Calculus

**MATH-GA 2903**

Yao Xiao

Spring 2023

## Contents

# 3/27 Lecture

# 1 Construction of Brownian Motion

## 1.1 Brownian Motion and Its Motivation

Loosely speaking, Brownian motion is a "random path" that satisfies the following properties.

- The path is continuous.
- At any given point in time, where the path goes next is *independent* of where the path came from.
- At any given point in time, where the path ends up in any finite time is normally distributed.

## 1.2 Symmetric Random Walk

It is always easier to start with a discrete case. Consider a symmetric random walk (SRW). Let $X_j$'s be independent and identically distributed such that $\mathbb{P}[X_j = 1] = \mathbb{P}[X_j = -1] = 1/2$. We now define $M(n)$ such that

$$M(0) = 0, \qquad M(n) = \sum_{j=1}^{n} X_j. \tag{1}$$

The process $M(n)$ is called a symmetric random walk. With each step, it either steps up or steps down one unit, and each of the two possible moves is equally likely.

### 1.2.1 Increments of the SRW

SRW has independent increments, which means that if we choose nonnegative integers $0 = n_0 < n_1 < \cdots < n_k$, then the random variables

$$M(n_1) = M(n_1) - M(n_0),\ M(n_2) - M(n_1),\ \cdots,\ M(n_k) - M(n_{k-1}) \tag{2}$$

are independent. Each of these random variables

$$M(n_{i+1}) - M(n_i) = \sum_{j=n_i+1}^{n_{i+1}} X_j \tag{3}$$

is called an increment of the random walk. It is the change in the position of the random walk between times $n_i$ and $n_{i+1}$. Increments over non-overlapping time intervals are indeed independent, since $X_j$'s are independent. Moreover,

$$\mathbb{E}[M(n_{i+1}) - M(n_i)] = \sum_{j=n_i+1}^{n_{i+1}} \mathbb{E}[X_j] = 0, \tag{4}$$

and

$$\mathrm{Var}[M(n_{i+1}) - M(n_i)] = \sum_{j=n_i+1}^{n_{i+1}} \mathrm{Var}[X_j] = n_{i+1} - n_i, \tag{5}$$

since $\mathrm{Var}[X_j] = 1$ and this holds by independence.

### 1.2.2 Martingale Property of the SRW

In order to define martingale we need the notion of $\sigma$-algebra. There is a very important, nontechnical reason to include $\sigma$-algebras in the study of stochastic processes, and that is to keep track of the amount of information. The temporal feature of a stochastic process suggests a flow of time, in which at each moment $t \geq 0$, we can talk about the past, the present, and the future, and can ask about how much an observer of the process knows about it at present, as compared to how much he know at some point in the past or will know at some time in the future.

**Definition 1.1.** A *filtration* on $(\Omega, \mathscr{F})$is a family $\{\mathscr{F}_t\}_{t\geq 0}$ of $\sigma$-algebras $\mathscr{F}_t \subseteq \mathscr{F}$, such that

$$0 \leq s \leq t \implies \mathscr{F}_s \subseteq \mathscr{F}_t, \tag{6}$$

*i.e.*, $\{\mathscr{F}_t\}_{t\geq 0}$ is increasing.

Thus informally, one can think about the $\sigma$-algebra $\mathscr{F}_t$ in the above definition as the "knowledge" available at time $t$. The fact that $\{\mathscr{F}_t\}_{t\geq 0}$ is monotonically increasing just reflects the fact that as time goes on, the amount of available information cannot decrease.

**Definition 1.2.** A stochastic process $\{Z_t\}_{t\geq 0}$ is called a *martingale* with respect to a filtration $\{\mathscr{F}_T\}_{t\geq 0}$ if

(1) $Z_t$ is $\mathscr{F}_t$-measurable for all $t$.

(2) $\mathbb{E}[|Z_t|] < \infty$ for all $t$.

(3) $\mathbb{E}[Z_t|\mathscr{F}_s] = Z_s$ for all $s \leq t$.

The first two conditions are for technical reasons, since we can only define measures on measurable sets, and we assume that the first moments are finite. The most important property in the above definition is the third property. Simply put, it states that the "best guess" for $X_t$ given information available at moment $s$ is $X_s$, and thus the process $X_t$ has neither upward or downward drift.

Now we want to check that the SRW is a martingale. First, we need to specify a filtration $\{\mathscr{F}_t\}_{t\geq 0}$. Let $\mathscr{F}_n$ be the $\sigma$-algebra (read information) generated by the set of random variables $X_1, X_2, \cdots, X_n$.

(1) $M(n)$ is $\mathscr{F}_n$-measurable since it depends only on $X_j$, $j \leq n$, *i.e.*, on the information available at time $n$.

(2) From the fact that $|M(t)| \leq t$, we can easily see that $\mathbb{E}[|M(t)|] \leq t < \infty$.

(3) Using the fact that $\mathbb{E}[M(s)|\mathscr{F}_s] = M(s)$ and the fact that the random variable $M(t) - M(s)$ is independent of $\sigma$-algebra $\mathscr{F}_s$, we can deduce that

$$\begin{aligned}\mathbb{E}[M(t)|\mathscr{F}_s] &= \mathbb{E}[M(t) - M(s) + M(s)|\mathscr{F}_s] = \mathbb{E}[M(t) - M(s)|\mathscr{F}_s] + \mathbb{E}[M(s)|\mathscr{F}_s] \\ &= \mathbb{E}[M(t) - M(s)] + M(s) = 0 + M(s) = M(s).\end{aligned} \tag{7}$$

### 1.2.3 Quadratic Variation of the SRW

The quadratic variation of a discrete stochastic process up to time $t$ is defined as

$$\langle M, M\rangle_t = \sum_{j=1}^{t}(M(j) - M(j-1))^2. \tag{8}$$

The quadratic variation up to time $t$ along the path is computed by taking all the one-step increments $M(j)-M(j-1)$ along that path, squaring these increments and then summing them up. Clearly, the SRM increments can only take values $\pm 1$, and thus $\langle M, M\rangle_t = t$. Note that in general, the quadratic variation of a stochastic process is stochastic itself. The fact that the quadratic variation in this case is deterministic is a special feature of the SRW.

### 1.3 Scaled Symmetric Random Walk

To approximate a Brownian motion, we speed up time and scale down the step size of an SRW. More precisely, we fix a positive integer $n$ and define the scaled SRW at rational points $k/n$ as

$$B^{(n)}\left(\frac{k}{n}\right) = \frac{1}{\sqrt{n}} M(k). \tag{9}$$

At all other points we define $B^{(n)}(t)$ as the linear interpolation between its values at the nearest points of the form $k/n$. The following properties of the scaled SRW could be easily proved using the corresponding properties of the SRW and we leave their proof as an exercise.

(1) Independence of increments, *i.e.*, for all rational numbers $0 = t_0 < t_1 < \cdots < t_n$ of the form $k/n$, the random variables

$$B^{(n)}(t_1) - B^{(n)}(t_0),\ B^{(n)}(t_2) - B^{(n)}(t_1),\ \cdots,\ B^{(n)}(t_n) - B^{(n)}(t_{n-1}) \tag{10}$$

are independent.

(2) $\mathbb{E}[B^{(n)}(t) - B^{(n)}(s)] = 0$, $\mathrm{Var}[B^{(n)}(t) - B^{(n)}(s)] = t - s$.

(3) $\mathbb{E}[B^{(n)}|\mathscr{F}_s] = B^{(n)}(s)$.

(4) The quadratic variation of the scaled SRW $\left\langle B^{(n)}, B^{(n)}\right\rangle_t = t$.

#### 1.3.1 Limiting Distribution of the Scaled SRW

Now the idea is to let $n \to \infty$ in the scaled SRW and in the limit we obtain something satisfying the intuitive properties outlined for Brownian motion. By definition,

$$B^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt}. \tag{11}$$

Since $M_k = \sum_{j=1}^k X_j$ has a binomial distribution with parameters $k$ and $1/2$, we can easily calculate the distribution of $B^{(n)}(t)$. In particular, one can draw the histogram of $B^{(n)}(t)$. We know that the random variable $B^{(n)}(t)$ has mean 0 and variance $t$. If we draw on top of the histogram of $B^{(n)}(t)$ the graph of normal density with mean 0 and variance $t$, we will see that the distribution of $B^{(n)}(t)$ is nearly normal. In fact, the central limit theorem asserts that $B^{(n)}(t)$ converges *in distribution* to the normal random variable $\mathscr{N}(0,t)$, but let us not use this theorem and prove the convergence of $b^{(n)}(t)$ from scratch. The whole point of this exercise is to learn a very powerful mathematical tool known as *characteristic functions*.

# 4/3 Lecture

**Definition 1.3.** Let $X$ be a random variable with distribution $\mathbb{P}$. The *characteristic function* of $X$ is defined as

$$\phi_X(u) = \mathbb{E}[e^{iuX}] = \int e^{iux} d\mathbb{P}[x] \tag{12}$$

for every $u \in \mathbb{R}$.

**Example 1.4.** (1) If $X = \mathrm{Ber}(p)$, then

$$\phi_X(u) = p(e^{iu} - 1) + 1. \tag{13}$$

(2) If $X = \mathrm{Bin}(n,p)$, then $\mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$, and

$$\phi_X(u) = (p(e^{iu} - 1) + 1)^n. \tag{14}$$

(3) If $X = \text{Pois}(\lambda)$, then $\mathbb{P}[X = k] = e^{-\lambda}\lambda^k/k!$, and

$$\phi_X(u) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} e^{iuk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{iu})^k}{k!} = \exp\left(\lambda(e^{it} - 1)\right). \tag{15}$$

(4) If $X$ is a Gaussian random variable $\mathscr{N}(m, \sigma^2)$, then

$$\phi_X(u) = \exp\left(ium - \frac{u^2\sigma^2}{2}\right). \tag{16}$$

The importance of the characteristics functions could be summarized in the following two theorems.

**Theorem 1.5.** If two random variables have the same characteristic function, then they have the same distribution. Of course, the converse also holds, in that if two random variables have the same distribution, then they have the same characteristic function.

**Theorem 1.6** (Paul Lévy theorem)**.** Let $X_n$ be a sequence of random variables on $\mathbb{R}$. Then

(1) If $X_n$ converges to $X$ in distribution, then $\phi_{X_n}(u) \to \phi_X(u)$ as $n \to \infty$ for every $u$, and in fact this convergence is uniform.

(2) If $\phi_{X_n}(u) \to \phi_X(u)$ for every $u$ pointwise, then $X_n$ converges to $X$ in distribution.

(3) If $\phi_{X_n}(u)$ converges to some function $\phi(u)$ for every $u$ pointwise, and $\phi(u)$ is continuous at 0, then there exists a unique random variable $X$, such that $\phi$ is the characteristic function of $X$ and $X_n$ converges to $X$ in distribution.

We use Part (2) of Paul Lévy theorem to prove that $B^{(n)}(t)$ converges to the normal distribution with expectation 0 and variance $t$. We have that

$$\begin{aligned} \phi_{B^{(n)}(t)}(u) &= \mathbb{E}\left[\exp\left(iuB^{(n)}(t)\right)\right] = \mathbb{E}\left[\exp\left(\frac{iuM_{nt}}{\sqrt{n}}\right)\right] = \mathbb{E}\left[\exp\left(\frac{iu}{\sqrt{n}}\sum_{j=1}^{nt} X_j\right)\right] \\ &= \left(\mathbb{E}\left[\exp\left(\frac{iuX_j}{\sqrt{n}}\right)\right]\right)^{nt} = \left(\frac{1}{2}\exp\left(\frac{iu}{\sqrt{n}}\right) + \frac{1}{2}\exp\left(\frac{-iu}{\sqrt{n}}\right)\right)^{nt}. \end{aligned} \tag{17}$$

Now we need to show that as $n \to \infty$, the above result converges to the characteristic function of the normal random variable with mean 0 and variance $t$, *i.e.*, to $e^{-u^2t/2}$. Expanding in Taylor series, we can see that

$$\frac{1}{2}\exp\left(\frac{iu}{\sqrt{n}}\right) + \frac{1}{2}\exp\left(\frac{-iu}{\sqrt{n}}\right) = 1 - \frac{u^2}{2n} + O(n^{-3/2}), \tag{18}$$

and thus

$$\phi_n(u) = \left(1 - \frac{u^2}{2n} + O(n^{-3/2})\right)^{nt} \to \exp\left(-\frac{u^2}{2n}\cdot nt\right) = \exp\left(-\frac{u^2t}{2}\right). \tag{19}$$

Therefore, we can conclude that $B^{(n)}(t)$ converges in distribution to a normal variable with expectation 0 and variance $t$. Strictly speaking, we have just demonstrated that for a fixed moment $t$, the distribution of $B^{(n)}(t)$ converges to $\mathscr{N}(0, t)$, but we did not prove the convergence of the whole stochastic process to Brownian motion. To do this, one has to further prove the convergence of the finite dimensional distributions and use *Prokhorov's theorem*.

## 1.4 Brownian Motion: A Formal Definition

A one-dimensional Brownian motion is a continuous-time stochastic process that satisfies the following properties.

- **Independence of increments:** If $t_0 < t_1 < \cdots < t_n$, then the random variables

$$B(t_0), B(t_1) - B(t_0), \cdots, B(t_n) - B(t_{n-1}) \tag{20}$$

are independent.

- **Normally distributed increments:** If $s, t \geq 0$, then

$$\mathbb{P}[B(t+s) - B(s) \in A] = \int_A \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx. \tag{21}$$

- **Continuous trajectories:** With probability 1, $B_0 = 0$ and $t \mapsto B_t$ is continuous.

# 2 Properties of the Brownian Motion

The limiting object described above is called a *standard Brownian motion.* It will be our main object of study, and the main building block in many applications.

## 2.1 Covariance of Brownian Motion

Let $0 \leq s \leq t$, then

$$\mathbb{E}[B(s)B(t)] = \mathbb{E}[B(s)(B(t) - B(s) + B(s))] = \mathbb{E}[B(s)(B(t) - B(s))] + \mathbb{E}[B^2(s)]. \tag{22}$$

By independence of increments, we have that $\mathbb{E}[B(s)(B(t)-B(s))] = \mathbb{E}[B(s)]\mathbb{E}[B(t)-B(s)] = 0$. Moreover, from the fact that $B(s)$is a normally distributed random variable with expectation 0 and variance $s$, we have that $\mathbb{E}[B^2(s)] = s$. Bringing these results into the previous expression, we have that

$$\mathbb{E}[B(t)B(s)] = 0 + s = s. \tag{23}$$

NOTE THAT IT IS IMPORTANT TO KNOW HOW TO COMPUTE THE MOMENTS OF A NORMAL RANDOM VARIABLE.

## 2.2 Non-differentiability of Brownian Paths

In the following theorem we prove that the trajectories of Brownian motion are very special, in that they are non-differentiable at every point.

**Theorem 2.1.** With probability 1, Brownian paths are not differentiable at any point.

*Proof.* Let $W = \{\omega;\ \exists t \in (0,1)$, such that $B(t)$ is differentiable at $t\}$. We want to show that $\mathbb{P}[W] = 0$. We shall use the following lemma, whose proof is fairly straightforwards and will be left as an exercise.

> *Lemma. If $f : (0,1) \to \mathbb{R}$ is differentiable at $x \in (0,1)$, then there exists $C > 0$ and $\delta > 0$, such that $f(x) - f(s) \leq C|x - s|$, for any $s \in [x - \delta, x + \delta]$.*

Now let $M_n(k) = \max_i \{|B((i+1)/n) - B(i/n)|;\ i = k-1, k, k+1\}$, and $M_n = \min_k \{M_n(k);\ k = 1, \cdots, n\}$. By the above lemma, for any $\omega \in W$, we can find $C > 0$ and $\delta > 0$, such that

$$|B(t) - B(s)| \leq C|t - s|, \qquad \forall s \in [t - \delta, t + \delta]. \tag{24}$$

Choose $n$ sufficiently large such that $n > \max\{t, 4/\delta\}$. We also choose $k$ such that $(k-1)/n \leq t \leq k/n$. The choices of $k$ and $n$ imply that $|i/n - t| < \delta$ for $i = k-1, k, k+1, k+2$. Furthermore, we can see that

$$\left|B\left(\frac{i+1}{n}\right) - B\left(\frac{i}{n}\right)\right| \leq \left|B\left(\frac{i+1}{n}\right) - B(t)\right| + \left|B\left(\frac{i}{n}\right) - B(t)\right| \leq C\left|\frac{i+1}{n} - t\right| + C\left|\frac{i}{n} - t\right| < 2C\delta < \frac{8C}{n}. \tag{25}$$

Note that $B((i+1)/n) - B(i/n)$ are independent and normally distributed with mean 0 and variance $1/n$, so that

$$\mathbb{P}\left[|M_n| \leq \frac{8C}{n}\right] \leq n\mathbb{P}\left[|M_n(k)| \leq \frac{8C}{n}\right] = n\left(\mathbb{P}\left[|z| \leq \frac{8C}{\sqrt{n}}\right]\right)^3 \leq n\left(\frac{16C}{\sqrt{2\pi n}}\right)^3 = \frac{16^3C^3}{\sqrt{2\pi n}}, \tag{26}$$

where $z$ is the standard normal. Let $A_n := \{|M_n| \leq 8C/n\}$, then the above inequality means that $\lim_{n\to\infty} \mathbb{P}[A_n] = 0$. By Fatou's lemma, we can conclude that

$$\mathbb{P}\left[\liminf_{n\to\infty} A_n\right] \leq \liminf_{n\to\infty} \mathbb{P}[A_n] = \lim_{n\to\infty} \mathbb{P}[A_n] = 0. \tag{27}$$

Finally, we note that $W \subseteq \liminf_{n\to\infty} A_n$, so $\mathbb{P}[W] = 0$, and the proof is complete. ☐

## 2.3 Scaling Properties of the Brownian Motion

If we fix time $t$, then the distribution of Brownian motion $B(t)$ is normal with expectation 0 and variance $t$. In fact, for any fixed set of times $t_1 < t_2 < \cdots < t_n$, the random variables $B(t_1), B(t_2), \cdots, B(t_n)$ are jointly normal with mean 0 and covariance given by $\mathbb{E}[B(s)B(t)] = \min\{s, t\}$.

**Theorem 2.2.** If $B(0) = 0$, then for any $\lambda > 0$, the stochastic process $B(\lambda t)/\sqrt{\lambda}$, $t \geq 0$ is a Brownian motion.

*Proof.* First, note that $X(t) = B(\lambda t)/\sqrt{\lambda}$ is a Gaussian process, *i.e.*, for any set $t_1 < t_2 < \cdots < t_n$, the joint distribution of $X(t_1), X(t_2), \cdots, X(t_n)$ is a multivariate Gaussian distribution. This property is clearly inherited from Brownian motion properties. Since normal distribution is characterized by its mean and covariance, we have to check that the mean and covariance of the process $X(t)$ coincide with those of Brownian motion. Indeed,

$$\mathbb{E}[X(t)] = \mathbb{E}\left[\frac{B(\lambda t)}{\sqrt{\lambda}}\right] = \frac{\mathbb{E}[B(\lambda t)]}{\sqrt{\lambda}} = 0, \tag{28}$$

and for $s < t$, we have that

$$\mathbb{E}[X(s)X(t)] = \mathbb{E}\left[\frac{B(\lambda s)}{\sqrt{\lambda}} \cdot \frac{B(\lambda t)}{\sqrt{\lambda}}\right] = \frac{\mathbb{E}[B(\lambda s)B(\lambda t)]}{\lambda} = \frac{\min\{\lambda s, \lambda t\}}{\lambda} = s. \tag{29}$$

Therefore, we can conclude that $B(\lambda t)/\sqrt{\lambda}$ has the same distribution as a Brownian motion, and thus a Brownian motion. The proof is thus complete. □

**Theorem 2.3.** If $B(t)$ is a Brownian motion starting at 0, then so is the process defined by $X(0) = 0$ and $X(t) = tB(1/t)$ for $t > 0$.

*Proof.* Fix $t_1 < t_2 < \cdots < t_n$, then clearly $X(t_1), X(t_2), \cdots, X(t_n)$ has a multivariate Gaussian distribution. We just have to check that it has the same mean and covariance structure. First of all, we have that

$$\mathbb{E}[X(t)] = \mathbb{E}[tB(1/t)] = t\mathbb{E}[B(1/t)] = 0. \tag{30}$$

Moreover, for $0 < s < t$, we can deduce that

$$\mathbb{E}[X(s)X(t)] = \mathbb{E}[sB(1/s) \cdot tB(1/t)] = st\mathbb{E}[B(1/s)B(1/t)] = st \min\{1/s, 1/t\} = s. \tag{31}$$

Therefore, $X(t)$ is indeed a Brownian motion, so the proof is complete. □

## 2.4 Quadratic Variation of Brownian Motion

Last time we computed quadratic variation of scaled SRW and it turned out to be $t$. In the following theorem we prove that the quadratic variation of the Brownian motion is also $t$. Let me again emphasize that the paths of Brownian motion are unusual in that their quadratic variation is not zero. This makes stochastic calculus different from ordinary calculus. In fact, non-zero quadratic variation is the source of the *volatility* term in the *Black-Scholes equation.*

To see how Brownian motion is different from functions we are used to in ordinary calculus, consider a function $f : \mathbb{R} \to \mathbb{R}$ that is continuously differentiable and compute its quadratic variation up to time $T$. Let us introduce the norm of the partition $0 = t_0 < t_1 < \cdots < t_n = T$, such that

$$\|\Pi\| = \max_{0 \leq j \leq n-1} (t_{j+1} - t_j). \tag{32}$$

Then, its quadratic variation can be computed as

$$\langle f, f \rangle_T = \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} (f(t_{j+1}) - f(t_j))^2. \tag{33}$$

We will use the mean value theorem which says that in each interval $(x_1, x_2)$, there exists a point $x_0$, such that

$$f(x_2) - f(x_1) = f'(x_0)(x_2 - x_1). \tag{34}$$

Applying to each interval $(t_j, t_{j+1})$, we get that there is a point $t_j^*$ in each such interval, such that

$$f(t_{j+1}) - f(t_j) = f'(t_j^*)(t_{j+1} - t_j). \tag{35}$$

Therefore, we can compute that

$$\begin{aligned}\sum_{j=0}^{n-1}(f(t_{j+1}) - f(t_j))^2 &= \sum_{j=0}^{n-1}|f'(t_j^*)|^2(t_{j+1}-t_j)^2 \le \max_{0\le j\le n-1}|t_{j+1}-t_j|\sum_{j=0}^{n-1}|f'(t_j^*)|^2(t_{j+1}-t_j) \\ &= \|\Pi\|\sum_{j=0}^{n-1}|f'(t_j^*)|^2(t_{j+1}-t_j) \to \|\Pi\|\int_0^T |f'(t)|^2 dt \to 0, \qquad \text{as } \|\Pi\| \to 0. \end{aligned} \tag{36}$$

Therefore, we have shown that continuously differentiable function has zero quadratic variation. For this reason we rarely consider quadratic variation in ordinary calculus.

**Theorem 2.4.** Let $B(t)$ be a Brownian motion. Then $\langle B, B\rangle_T = T$ almost surely.

*Proof.* Define

$$Q_\Pi = \sum_{j=0}^{n-1}(B(t_{j+1}) - B(t_j))^2. \tag{37}$$

We shall prove that $\mathbb{E}[Q_\Pi] \to T$ and $\mathrm{Var}[Q_\Pi] \to 0$ as $\|\Pi\| \to 0$. In other words, we shall prove the $L^2$ convergence of $Q_\Pi$, which is stronger than *almost sure* convergence. First, notice that

$$\mathbb{E}[(B(t_{j+1}) - B(t_j))^2] = \mathrm{Var}[B(t_{j+1}) - B(t_j)] = t_{j+1} - t_j. \tag{38}$$

□

Therefore, we can compute that

$$\mathbb{E}[Q_\Pi] = \sum_{j=0}^{n-1}\mathbb{E}[(B(t_{j+1}) - B(t_j))^2] = \sum_{j=0}^{n-1}(t_{j+1} - t_j) = T. \tag{39}$$

Now, in order to show that $\mathrm{Var}[Q_\Pi] \to 0$ as $\|\Pi\| \to 0$, note that

$$\begin{aligned}\mathrm{Var}[(B(t_{j+1}) - B(t_j))^2] &= \mathbb{E}[((B(t_{j+1}) - B(t_j))^2 - (t_{j+1}-t_j))^2] \\ &= \mathbb{E}[(B(t_{j+1}) - B(t_j))^4 - 2(t_{j+1}-t_j)(B(t_{j+1}) - B(t_j))^2 + (t_{j+1}-t_j)^2] \\ &= \mathbb{E}[(B(t_{j+1}) - B(t_j))^4] - 2(t_{j+1}-t_j)\mathbb{E}[(B(t_{j+1}) - B(t_j))^2] + (t_{j+1}-t_j)^2 \\ &= 3(t_{j+1}-t_j)^2 - 2(t_{j+1}-t_j)^2 + (t_{j+1}-t_j)^2 = 2(t_{j+1}-t_j)^2. \end{aligned} \tag{40}$$

In the previous deduction, we used the fourth moment of a normal random variable. Now, we can see that

$$\mathrm{Var}[Q_\Pi] = \sum_{j=0}^{n-1}\mathrm{Var}[(B(t_{j+1}) - B(t_j))^2] = \sum_{j=0}^{n-1}2(t_{j+1}-t_j)^2 \le 2\,\|\Pi\|\sum_{j=0}^{n-1}(t_{j+1}-t_j) = 2\,\|\Pi\|\,T, \tag{41}$$

which clearly tends to 0 as $\|\Pi\| \to 0$. Therefore, we can conclude that $\langle B, B\rangle_T = \lim_{\|\Pi\|\to 0} Q_\Pi$ which has expectation $T$ and variance 0, thus being $T$ almost surely. Our proof is thus complete.

**Remark 2.5.** $\langle B, B\rangle_t = t$ can be informally written as

$$dB(t)dB(t) = dt. \tag{42}$$

**Remark 2.6.** In addition to computing the quadratic variation of Brownian motion, we can also compute the cross variation of $B(t)$ with $t$ and the quadratic variation of $t$ itself, which are

$$\lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}(B(t_{j+1}) - B(t_j))(t_{j+1}-t_j) = 0, \qquad \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}(t_{j+1}-t_j)^2 = 0. \tag{43}$$

These can be informally written as

$$dB(t)dt = 0, \qquad dtdt = 0. \tag{44}$$

# 4/10 Lecture

## 2.5 Volatility of Geometric Brownian Motion

Let $\alpha$ and $\sigma$ be constants and define the *geometric Brownian motion* as

$$S(t) = S(0) \exp\left(\sigma B(t) + \left(\alpha - \frac{\sigma^2}{2} t\right)\right). \tag{45}$$

This is the asset-price model used in the Black-Scholes formula. We show how to use the quadratic variation of Brownian motion to identify the *volatility* $\sigma$ from a path of the process. Let $0 \leq T_1 \leq T_2$ be given and suppose that we observe the geometric Brownian motion $S(t)$ on the time interval $T_1 \leq t \leq T_2$. We may then choose a partition of this interval $T_1 = t_0 < t_1 < \cdots < t_m = T_2$ and observe the "log-return" over each interval $[t_j, t_{j+1}]$, such that

$$\log \frac{S(t_{j+1})}{S(t_j)} = \sigma\left(B(t_{j+1}) - B(t_j)\right) + \left(\alpha - \frac{\sigma^2}{2}\right)(t_{j+1} - t_j). \tag{46}$$

The sum of the squares of the log-returns, sometimes called the *realized volatility*, is

$$\begin{aligned}\sum_{j=0}^{m-1} \left(\log \frac{S(t_{j+1})}{S(t_j)}\right)^2 &= \sigma^2 \sum_{j=0}^{m-1} \left(B(t_{j+1}) - B(t_j)\right)^2 + \left(\alpha - \frac{\sigma^2}{2}\right)^2 \sum_{j=0}^{m-1} (t_{j+1} - t_j)^2 \\ &\quad + 2\sigma\left(\alpha - \frac{\sigma^2}{2}\right) \sum_{j=0}^{m-1} \left(B(t_{j+1}) - B(t_j)\right)(t_{j+1} - t_j) \to \sigma^2(T_2 - T_1). \end{aligned} \tag{47}$$

# 3 Construction of the Itô Integral

Consider an asset whose price per share is equal to $X_t$, $t \geq 0$. Also consider a portfolio that initially consists of $\Delta_0$ shares. Now consider the following trading strategy. Keep the initial position $\Delta_0$ up to time $t_1 \geq t_0 = 0$ and then rebalance the portfolio by taking position $\Delta_1$ in the asset. Keep it up to time $t_2 \geq t_1$ and then rebalance the portfolio again by taking position $\Delta_2$ in the asset. In general, we rebalance the portfolio at trading date $t_i$ by taking position $\Delta_i$ in the asset and keeping it till the next trading date $t_{i+1}$. Then we are interested in the profit $I_T(\Delta)$ of the above trading strategy at time $T$. Clearly, we have that

$$I_T(\Delta) = \Delta_0(X(t_1) - X(t_0)) + \Delta_1(X(t_2) - X(t_1)) + \cdots + \Delta_n(X(t_n) - X(t_{n-1})), \tag{48}$$

and by an analogy with the Riemann integral, we write symbolically that

$$I_T(\Delta) = \int_0^T \Delta(t) dX(t), \tag{49}$$

where $\Delta(t)$ is a piecewise constant function which is equal to $\Delta_i$ on $[t_i, t_{i+1}]$.

## 3.1 Construction of the Stochastic Integral

We fix an interval $[S, T]$ and try to make sense of

$$\int_S^T f(t, w) dX_t(w), \tag{50}$$

where $f(t, w)$ is a random function and $dX_t(w)$ refers to the increments of stochastic process $X_t$. Before we proceed we make the following clarifications.

**Remark 3.1.** (1) We restrict our attention to functions $f$, such that for any fixed $t$, the random variable $f(t, w)$ is $\mathscr{F}_t$-measurable. To explain this restriction we come back to the previous example. Position $\Delta_i$ that we take in the asset of time $t_i$, $i \geq 1$ may depend on the price history $\mathscr{F}_t$ of the asset, but it must be independent of the future behavior of the process $X_t$.

(2) We restrict our consideration to the case where $X_t$ is a Brownian motion, though the case of general stochastic process $X_t$ is quite similar.

The problem we face when trying to assign meaning to the integral (50) is that Brownian motion paths cannot be differentiated with respect to time. If $X(t)$ is a differentiable function, then we can define

$$\int_S^T f(t,w)dX(t) = \int_S^T f(t,w)X'(t)dt, \tag{51}$$

where the right-hand side is an ordinary integral with respect to time. This approach unfortunately does not work for Brownian motion as we have proved that the trajectories of Brownian motion are not differentiable.

Just like in the definition of the usual Riemann integral $\int_S^T f(t)dt$, where $f(t)$ is a deterministic function, we start with a definition for a simple class of functions $f$ and then extend by some approximation procedure. Assume that $\Pi = \{t_0, \cdots, t_n\}$ is a partition of $[S,T]$, *i.e.*, $S = t_0 < t_1 < \cdots < t_n = T$, and that $f(t,w)$ is constant in $t$ on each subinterval $[t_j, t_{j+1})$. Such a process $f(t,w)$ is called a *simple process*. We start with defining integral (50) for simple processes. Consider the interval $[t_0,t_1)$, on which $f(t,w) = e_1(w)$ is a random quantity, but independent of $t$. Therefore, it is natural to assume that

$$\int_{t_0}^{t_1} f(t,w)dB_t(w) = \int_0^1 e_1(w)dB_t(w) = e_1(w)(B(t_1) - B(t_0)). \tag{52}$$

Applying this procedure to intervals $[t_1,t_2), [t_2,t_3), \cdots, [t_{n-1},t_n)$, we can obtain that

$$\int_S^T f(t,w)dB_t(w) = e_1(w)(B(t_1) - B(t_0)) + e_2(w)(B(t_2) - B(t_1)) + \cdots + e_n(w)(B(t_n) - B(t_{n-1})). \tag{53}$$

Now, to extend this definition of stochastic integral (50) for general processes $f(t,w)$, we approximate it with simple processes in a similar way that we approximate continuous functions by stepwise constant functions in the theory of Riemann integration. But without any further assumption on approximating functions $e_i(w)$, our definition of the integral leads to difficulties. Here is an example of what kind of difficulties we can expect. Consider

$$\int_0^T B_t dB_t. \tag{54}$$

Riemann integral is just a limit of Riemann sums, such that

$$\int_S^T f(t)dt \approx \sum_i f(t_i^*)(t_{i+1} - t_i), \tag{55}$$

where $t_i^*$ is any point on the interval $[t_i, t_{i+1}]$. When the length of the longest interval in the partition tends to zero, we will obtain the correct limit. Note that it was not important what point $t_i^*$ we take. For instance, it could be $t_i^* = t_i$ (left point approximation) or $t_i^* = t_{i+1}$ (right point approximation) or something else. Let us try to do something similar for integral (54). The left point approximation can be written as

$$I_1 \sim \sum_i B(t_i)(B(t_{i+1}) - B(t_i)), \tag{56}$$

and the right point approximation can be written as

$$I_2 \sim \sum_i B(t_{i+1})(B(t_{i+1}) - B(t_i)). \tag{57}$$

From the independence of increments of Brownian motion and the fact that $\mathbb{E}[B(t_{i+1}) - B(t_i)] = 0$ and $\mathbb{E}[B(t_1)] = 0$, we have that

$$\mathbb{E}[I_1] = \sum_i \mathbb{E}[B(t_i)(B(t_{i+1}) - B(t_i))] = \sum_i \mathbb{E}[B(t_i)]\mathbb{E}[B(t_{i+1}) - B(t_i)] = 0, \tag{58}$$

$$\mathbb{E}[I_2] = \sum_i \mathbb{E}[B(t_{i+1})(B(t_{i+1}) - B(t_i))] = \sum_i \left(\mathbb{E}[B^2(t_{i+1})] - \mathbb{E}[B(t_{i+1})B(t_i)]\right) = \sum_i (t_{i+1} - t_i) = T, \tag{59}$$

where in right point approximation, the result follows by observing that $\mathbb{E}[B^2(t_{i+1})] = t_{i+1}$ (quadratic variation) and $\mathbb{E}[B(t_{i+1})B(t_i)] = \min(t_{i+1}, t_i) = t_i$. Consequently, we can see that depending on the choice of the point $t_i^*$ in the approximation, we can get very different results, unlike in Riemann integrals where such a choice does not matter. Now since the function $f(t,w)$ is $\mathscr{F}_t$-measurable, it is reasonable to choose the approximating simple function to be $\mathscr{F}_t$-measurable as well, so the left point approximation $t_i^* = t_i$ would be our only choice. This leads to the definition of the *Itô integrals.*

**Remark 3.2.** If for each $t \geq 0$, the random variable $f(t,w)$ is $\mathscr{F}_t$-measurable, then we say that the process $f(t,w)$ is $\mathscr{F}_t$*-adapted.* For instance, if $\mathscr{F}_t$ is a filtration of Brownian motion, then the process $f_t(t,w) = B(t/2)$ is $\mathscr{F}_t$-adapted, while $f_t(t,w) = B(2t)$ is not.

## 3.2 Properties of the Itô Integral for Simple Processes

The Itô integral (50) is defined as the gain from treading in the martingale $B_t$. A martingale has no tendency to rise or fall, and hence it is to be expected that

$$I_t(f) = \int_0^t f(t,w)dB_t \tag{60}$$

also has no tendency to rise or fall.

**Theorem 3.3.** Itô integral is a martingale.

*Proof.* It suffices to prove for simple processes, since the same conclusion for general stochastic processes can be obtained up to taking limits. Let $W(t)$ be a simple process, $0 \leq s < t \leq T$ be given, and $\Pi$ be a partition. Assume that there exists partition points $t_l < t_k$, such that $s \in [t_l, t_{l+1})$ and $t \in [t_k, t_{k+1})$. Let $D_j = W(t_{j+1}) - W(t_j)$ for for $0 \leq j \leq k-1$ and $D_k = W(t) - W(t_k)$. Thus we split the sum as

$$I(t) = \sum_{j=0}^{l-1} \Delta_j D_j + \Delta_l D_l + \sum_{j=l+1}^{k-1} \Delta_j D_j + \Delta_k D_k. \tag{61}$$

We investigate the four summands respectively. For the first summand, since $t_l \leq s$, $\Delta_j D_j$ is $\mathscr{F}_s$-measurable for all $0 \leq j \leq l-1$. Hence, we have that

$$\mathbb{E}[\Delta_j D_j | \mathscr{F}_s] = \Delta_j D_j. \tag{62}$$

For the second summand, we can compute that

$$\mathbb{E}[\Delta_l D_l | \mathscr{F}_s] = \Delta_l \mathbb{E}[W(t_{l+1})|\mathscr{F}_s] - \Delta_l \mathbb{E}[W(t_l)|\mathscr{F}_s] = \Delta_l(W(s) - W(t_l)). \tag{63}$$

For the third summand, we can compute for each $l+1 \leq j \leq k-1$ that

$$\begin{aligned}\mathbb{E}[\Delta_j D_j | \mathscr{F}_s] &= \mathbb{E}[\mathbb{E}[\Delta_j D_j | \mathscr{F}_{t_j}]|\mathscr{F}_s] = \mathbb{E}[\Delta_j \mathbb{E}[W(t_{j+1})|\mathscr{F}_{t_j}] - \Delta_j \mathbb{E}[W(t_j)|\mathscr{F}_{t_j}]|\mathscr{F}_s] \\ &= \mathbb{E}[\Delta_j(W(t_j) - W(t_j))|\mathscr{F}_s] = 0. \end{aligned} \tag{64}$$

The fourth summand is similar to the third, which would vanish in expectation when conditioned on $\mathscr{F}_s$. Combining the previous results, we can then conclude that

$$\mathbb{E}[I(t)|\mathscr{F}_s] = \sum_{j=0}^{l-1} \Delta_j D_j + \Delta_l(W(s) - W(t_l)) = I(s), \tag{65}$$

which implies that $I(t)$ is a martingale, so the proof is complete. □

Now, since $I_t(f)$ is a martingale and $I_0 = 0$, we have that $\mathbb{E}[I_t(f)] = 0$ for all $t \geq 0$. It then follows that $\mathrm{Var}[I_t(f)] = \mathbb{E}[I_t^2(f)]$, which can be evaluated by the formula in the following theorem.

**Theorem 3.4.** The Itô integral satisfies that

$$\mathbb{E}[I_t^2(f)] = \mathbb{E}\left[\int_0^t f^2(s,w)ds\right]. \tag{66}$$

*Proof.* For the simplicity of notation we introduce $\Delta B_i = B(t_{i+1}) - B(t_i)$ and $e_i = e_i(w)$. Then by definition, we can write that

$$I_t(f) = \int_0^t f(s,w)dB_s = \sum_i e_i \Delta B_i. \tag{67}$$

Squaring on both sides, we will obtain that

$$I_t^2(f) = \sum_i e_i^2 (\Delta B_i)^2 + 2\sum_{i<j} e_i e_j \Delta B_i \Delta B_j. \tag{68}$$

Now by taking expectation on both side, we can see that

$$\begin{aligned}\mathbb{E}[I_t^2(f)] &= \sum_i \mathbb{E}[e_i^2(\Delta B_i)^2] + 2\sum_{i<j}\mathbb{E}[e_ie_j\Delta B_i\Delta B_j] = \sum_i \mathbb{E}[e_i^2]\mathbb{E}[(\Delta B_i)^2] + 2\sum_{i<j}\mathbb{E}[e_ie_j]\underbrace{\mathbb{E}[\Delta B_i]}_{=0}\underbrace{\mathbb{E}[\Delta B_j]}_{=0} \\ &= \sum_i \mathbb{E}[e_i^2]\mathrm{Var}[\Delta B_i] = \sum_i \mathbb{E}[e_i^2](t_{i+1}-t_i) = \mathbb{E}\left[\sum_i e_i^2(t_{i+1}-t_i)\right] = \mathbb{E}\left[\int_0^t f^2(s,w)ds\right],\end{aligned} \tag{69}$$

as desired, so the proof is complete. □

**Theorem 3.5.** The quadratic variation of the stochastic integral

$$\left\langle \int_0^t f(t,w)dB_t, \int_0^t f(t,w)dB_t \right\rangle_T = \int_0^T f^2(t,w)dt = \sum_i e_i^2 \Delta t_i. \tag{70}$$

*Proof.* We take $\Pi = \{t_0, \cdots, t_n\}$ to be a partition of $[0,T]$, *i.e.*, $0 = t_0 < t_1 < \cdots < t_n = T$, then the quadratic variation can be computed as

$$\begin{aligned}Q_\Pi = \left\langle \int_0^t f(t,w)dB_t, \int_0^t f(t,w)dB_t \right\rangle_T &= \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left(\int_0^{t_{j+1}} f(t,w)dB_t - \int_0^{t_j} f(t,w)dB_t\right)^2 \\ &= \lim_{\|\Pi\|\to 0}\sum_{j=0}^{n-1}\left(e_j(w)(B_{t_{j+1}} - B_{t_j})\right)^2.\end{aligned} \tag{71}$$

To this end, we consider the quantity

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{j=0}^{n-1}\left(e_j(w)(B_{t_{j+1}}-B_{t_j})\right)^2 - \sum_{j=0}^{n-1} e_j(w)^2(t_{j+1}-t_j)\right)^2\right] \\ &= \mathbb{E}\left[\left(\sum_{j=0}^{n-1}\underbrace{e_j(w)}_{=:e_j}{}^2\left((\underbrace{B_{t_{j+1}}-B_{t_j}}_{=:\Delta B_j})^2 - (\underbrace{t_{j+1}-t_j}_{=:\Delta t_j})\right)\right)^2\right] \\ &= \sum_j \mathbb{E}\left[e_j^4\left((\Delta B_j)^2 - \Delta t_j\right)^2\right] + 2\sum_{i<j}\mathbb{E}\left[e_i^2e_j^2\left((\Delta B_i)^2-\Delta t_i\right)\left((\Delta B_j)^2-\Delta t_j\right)\right] \\ &= \sum_j \mathbb{E}\left[e_j^4\right]\mathbb{E}\left[\left((\Delta B_j)^2-\Delta t_j\right)^2\right] + 2\sum_{i<j}\mathbb{E}\left[e_i^2e_j^2\right]\underbrace{\mathbb{E}\left[(\Delta B_i)^2-\Delta t_i\right]}_{=0}\underbrace{\mathbb{E}\left[(\Delta B_j)^2-\Delta t_j\right]}_{=0} \\ &= \sum_j \mathbb{E}\left[e_j^4\right]\mathbb{E}\left[(\Delta B_j)^2 - \mathbb{E}\left[(\Delta B_j)^2\right]\right] = \sum_j \mathbb{E}\left[e_j^4\right]\mathrm{Var}\left[(\Delta B_j)^2\right] = \sum_j \mathbb{E}\left[e_j^4\right]\cdot 2\left(\Delta t_j\right)^2 \le 2\|\Pi\|\sum_j \mathbb{E}\left[e_j^4\right]\Delta t_j.\end{aligned} \tag{72}$$

Now, if $\|\Pi\| \to 0$, we have that $\sum_j \mathbb{E}\left[e_j^4\right]\Delta t_j \to \int_0^T \mathbb{E}[f^4(t,w)]dt < \infty$, so that the expected value as computed above must vanish. As a result, the quadratic variation is shown to be

$$\sum_i e_i^2 \Delta t_i = \int_0^T f^2(t,w)dt, \tag{73}$$

as desired, and the proof is complete. □

### 3.3 Itô Integral for General Functions

We now introduce the class of general functions $f(t, w)$ for which is Itô integral will be defined.

**Definition 3.6.** Let $V = V(S, T)$ be the class of functions $f(t, w) : [0, \infty) \times \Omega \to \mathbb{R}$, such that $f(t, w)$ is $\mathscr{F}_t$-adapted and $\int_S^T f^2(t, w)dt < \infty$. The Itô integral is defined for $V(S, T)$.

We claim that each function $f \in V(S, T)$ can be approximated by a sequence $\{\phi_n\}_{n\geq 1}$ of simple functions (or equivalently, by a sequence of simple processes) in the sense that

$$\mathbb{E}\left[\int_S^T (f - \phi_n)^2 dt\right] \to 0, \qquad \text{as } n \to \infty. \tag{74}$$

The approximation is done in three steps.

(1) Approximate bounded continuous functions with simple functions. Let $g \in V$ be bounded, with every trajectory $g(\cdot, w)$ ($w$ is fixed and $t$ changes) being continuous. Then, there exists a sequence of simple functions $\phi_n \in V$, such that

$$\mathbb{E}\left[\int_S^T (g - \phi_n)^2 dt\right] \to 0, \qquad \text{as } n \to \infty. \tag{75}$$

(2) Approximate bounded functions with bounded continuous functions. Let $h \in V$ be bounded, then there exists a sequence of bounded continuous functions $g_n$, such that

$$\mathbb{E}\left[\int_S^T (h - g_n)^2 dt\right] \to 0, \qquad \text{as } n \to \infty. \tag{76}$$

(3) Approximate general functions with bounded functions. Let $f \in V$, then there exists a sequence of bounded functions $h_n$, such that

$$\mathbb{E}\left[\int_S^T (f - h_n)^2 dt\right] \to 0, \qquad \text{as } n \to \infty. \tag{77}$$

Putting together the three steps above, we get that for any function $f(t, w) \in V$, there exists a sequence of simple functions $\phi_n(t, w)$, such that (74) holds. We then define the Itô integral of the function $f(t, w)$ as

$$I(f) = \int_S^T f(t, w)dB_t = \lim_{n\to\infty} I(\phi_n). \tag{78}$$

The limit exists since by Theorem 3.4, we can deduce that

$$\mathbb{E}\left[(I(\phi_n) - I(\phi_m))^2\right] = \mathbb{E}\left[\int_S^T (\phi_n - \phi_m)^2 dt\right] \leq \mathbb{E}\left[\int_S^T (f - \phi_n)^2 dt\right] + \mathbb{E}\left[\int_S^T (f - \phi_m)^2 dt\right] \to 0, \tag{79}$$

as $m, n \to \infty$. Thus, the sequence of random variables $\left\{\int_S^T \phi_n(t, w)dB_t\right\}$ forms a Cauchy sequence in $L_2(\Omega, \mathscr{F}, \mathbb{P})$. Since $L_2(\Omega, \mathscr{F}, \mathbb{P})$ is a complete space, there exists a limit of $I(\phi_n)$ as an element of it. This limit is by definition the Itô integral $I(f)$.

**Example 3.7.** Consider the Itô integral

$$\int_0^T B_t dB_t. \tag{80}$$

By definition, we can write that

$$\int_0^T B_t dB_t = \lim_{n\to\infty} \int_0^T \phi_n(t, w)dB_t, \tag{81}$$

where $\phi_n$ are $\mathscr{F}_t$-adapted simple functions such that

$$\mathbb{E}\left[\int_0^T (\phi_n - B_t)^2 dt\right] \to 0, \qquad \text{as } n \to \infty. \tag{82}$$

We now approximate $f(t,w) = B_t$ by partitioning $[0,T]$ into $[0,t_1],[t_1,t_2],\cdots,[t_{n-1},T]$ and defining $\phi_n(t,w) = B(t_i)$ for $t \in [t_i, t_{i+1}]$. We check that

$$\begin{aligned}\mathbb{E}\left[\int_0^T (\phi_n - B_t)^2 dt\right] &= \mathbb{E}\left[\sum_i \int_{t_i}^{t_{i+1}} (\phi_n - B_t)^2 dt\right] = \mathbb{E}\left[\sum_i \int_{t_i}^{t_{i+1}} (B(t_i) - B_t)^2 dt\right] \\ &= \sum_i \int_{t_i}^{t_{i+1}} \mathbb{E}\left[(B(t_i) - B(t))^2\right] dt = \sum_i \int_{t_i}^{t_{i+1}} (t - t_i) dt = \sum_i \frac{(t_{i+1} - t_i)^2}{2}.\end{aligned} \tag{83}$$

By defining $M_n = \max_i (t_{i+1} - t_i)$, we can thus see that

$$\mathbb{E}\left[\int_0^T (\phi_n - B_t)^2 dt\right] = \sum_i \frac{(t_{i+1} - t_i)^2}{2} \leq \frac{M_n}{2} \sum_i (t_{i+1} - t_i) = \frac{M_n T}{2} \to 0. \tag{84}$$

Therefore, now it suffices to compute

$$\int_0^T \phi_n dB_t = \sum_i B(t_i) \underbrace{(B(t_{i+1}) - B(t_i))}_{=:\Delta B_i} = \sum_i B(t_i) \Delta B_i. \tag{85}$$

Note that we have the identity

$$\Delta B_i^2 = B^2(t_{i+1}) - B^2(t_i) = (\Delta B_i)^2 + 2B(t_{i+1})B(t_i) - 2B^2(t_i) = (\Delta B_i)^2 + 2B(t_i)\Delta B_i. \tag{86}$$

Summing both sides over $i$, we would then obtain that

$$\begin{aligned} B^2(T) &= \sum_i \Delta B_i^2 = \sum_i (\Delta B_i)^2 + 2\sum_i B(t_i)\Delta B_i = \sum_i (\Delta B_i)^2 + 2\int_0^T \phi_n dB_t \\ \implies \int_0^T \phi_n dB_t &= \frac{1}{2}\left(B^2(T) - \sum_i (\Delta B_i)^2\right) \to \frac{1}{2}\left(B^2(T) - T\right), \end{aligned} \tag{87}$$

where the limit is in the $L_2$ sense. Therefore, we can conclude that

$$\int_0^T B_t dB_t = \frac{B^2(T) - T}{2}. \tag{88}$$

## 4/17 Lecture

# 4 Itô Formula

During the last lecture we defined Itô integral

$$\int_0^t f(s,w) dB_s = I(f) \tag{89}$$

for stochastic processes $f(t,w)$ that are $\mathscr{F}_t$-adapted and square-integrable. We defined $I(f)$ to be the limit in $L^2(\Omega, \mathscr{F}, \mathbb{P})$ of $I(\phi_n)$ for any sequence of approximating simple stochastic processes $\phi_n$. Approximating sequence was defined as satisfying

$$\mathbb{E}\left[\int_0^t (\phi_n - f)^2 dt\right] \to 0, \qquad \text{as } n \to \infty. \tag{90}$$

Thus the procedure of calculating Itô integral from the definition is rather work consuming. Let us step back a second and take a look at the Riemann integral. Even though it is defined as the limit of Riemann sums, in practice one never does that. Instead, one uses the fundamental theorem of calculus and the chain rule. For instance, to compute

$$I(t) = \int_0^t s \exp\left(-\frac{s^2}{2}\right) ds, \tag{91}$$

we notice that $\left(-\exp\left(-s^2/2\right)\right)' = s\exp\left(-s^2/2\right)$, and thus

$$I(t) = \int_0^t s \exp\left(-\frac{s^2}{2}\right) ds = \int_0^t \left(-\exp\left(-\frac{s^2}{2}\right)\right)' ds = -\exp\left(-\frac{s^2}{2}\right)\bigg|_{s=0}^{t} = 1 - \exp\left(-\frac{t^2}{2}\right). \tag{92}$$

In general, it is desirable to have some analog of the chain rule in the case of Itô integral (to avoid taking the limit of $I(\phi_n)$). The analog for the chain rule is the Itô formula.

## 4.1 Itô Formula for Brownian Motion

If Brownian motion *were* differentiable, then the chain rule would give

$$\frac{df(B_t)}{dt} = \frac{df}{dx}\bigg|_{x=B_t} \frac{dB_t}{dt}. \tag{93}$$

However, the problem we face is the fact that the Brownian motion $B_t$ is non-differentiable, and thus the chain rule above would not apply.

**Theorem 4.1** (Itô formula for Brownian motion)**.** Let $f(t,x)$ be a function for which the partial derivatives $f_t(f,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous, and let $B_t$ be a Brownian motion. Then for every $T \geq 0$, we have

$$f(T, B_T) = f(0, B_0) + \int_0^T f_t(t, B_t)dt + \int_0^T f_x(t, B_t)dB_t + \frac{1}{2}\int_0^T f_{xx}(t, B_t)dt. \tag{94}$$

*Proof.* From the Taylor series expansion, we can obtain that

$$\begin{aligned} f(t_{j+1}, x_{j+1}) - f(t_j, x_j) &= f_t(t_j, x_j)(t_{j+1} - t_j) + f_x(t_j, x_j)(x_{j+1} - x_j) + \frac{1}{2} f_{tt}(t_j, x_j)(t_{j+1} - t_j)^2 \\ &+ \frac{1}{2} f_{xx}(t_j, x_j)(x_{j+1} - x_j)^2 + f_{tx}(t_j, x_j)(t_{j+1} - t_j)(x_{j+1} - x_j) + \text{higher order terms.} \end{aligned} \tag{95}$$

Let $\Pi = \{t_0, \cdots, t_n\}$ be a partition of the interval $[0,T]$. Applying this formula with $x_{j+1} = B(t_{j+1})$ and $x_j = B(t_j)$ and summing over $j$, we thus have that

$$\begin{aligned} f(T, B_T) - f(0, B_0) &= \sum_{j=0}^{n-1} \left(f(t_{j+1}, B(t_{j+1})) - f(t_j, B(t_j))\right) \\ &= \sum_{j=0}^{n-1} f_t(t_j, B(t_j))(t_{j+1} - t_j) + \sum_{j=0}^{n-1} f_x(t_j, B(t_j))(B(t_{j+1}) - B(t_j)) + \frac{1}{2}\sum_{j=0}^{n-1} f_{tt}(t_j, B(t_j))(t_{j+1} - t_j)^2 \\ &+ \frac{1}{2}\sum_{j=0}^{n-1} f_{xx}(t_j, B(t_j))(B(t_{j+1}) - B(t_j))^2 + \sum_{j=0}^{n-1} f_{tx}(t_j, B(t_j))(t_{j+1} - t_j)(B(t_{j+1}) - B(t_j)) + \text{higher order terms.} \end{aligned} \tag{96}$$

We investigate each summand respectively. For the first summand, we can see that

$$\sum_{j=0}^{n-1} f_t(t_j, B(t_j))(t_{j+1} - t_j) \to \int_0^T f_t(t, B_t)dt, \qquad \text{as } \|\Pi\| \to 0. \tag{97}$$

For the second summand, we can see that

$$\sum_{j=0}^{n-1} f_x(t_j, B(t_j))(B(t_{j+1}) - B(t_j)) \to \int_0^T f_x(t, B_t) dB_t, \qquad \text{as } \|\Pi\| \to 0. \tag{98}$$

For the third summand, let $a_j = f_{tt}(t_j, B(t_j))$, so we can write that

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{j=0}^{n-1} f_{tt}(t_j, B(t_j))(t_{j+1} - t_j)^2\right)^2\right] &= \sum_j \mathbb{E}\left[a_j^2 (\Delta t_j)^4\right] + 2\sum_{i<j} \mathbb{E}\left[a_i a_j (\Delta t_i)^2 (\Delta t_j)^2\right] \\ &= \sum_j \mathbb{E}\left[a_j^2\right] (\Delta t_j)^4 + 2\sum_{i<j} \mathbb{E}\left[a_i a_j\right] (\Delta t_i)^2 (\Delta t_j)^2 \to 0, \qquad \text{as } \|\Pi\| \to 0. \end{aligned} \tag{99}$$

Therefore, the third summand vanishes. For the fourth summand, let $b_j = f_{xx}(t_j, B(t_j))$, so we can write that

$$\sum_{j=0}^{n-1} f_{xx}(t_j, B(t_j))(B(t_{j+1}) - B(t_j))^2 = \sum_{j=0}^{n-1} b_j (\Delta B_j)^2. \tag{100}$$

We then consider

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{j=0}^{n-1} b_j (\Delta B_j)^2 - \sum_{j=0}^{n-1} b_j \Delta t_j\right)^2\right] &= \sum_j \mathbb{E}\left[b_j^2 \left((\Delta B_j)^2 - \Delta t_j\right)^2\right] + 2\sum_{i<j} \mathbb{E}\left[b_i b_j \left((\Delta B_i)^2 - \Delta t_i\right)\left((\Delta B_j)^2 - \Delta t_j\right)\right] \\ &= \sum_j \mathbb{E}\left[b_j^2\right] \mathbb{E}\left[(\Delta B_j)^4 - 2(\Delta B_j)^2 \Delta t_j + (\Delta t_j)^2\right] + 2\sum_{i<j} \mathbb{E}\left[b_i b_j\right] \mathbb{E}\left[(\Delta B_i)^2 - \Delta t_i\right] \mathbb{E}\left[(\Delta B_j)^2 - \Delta t_j\right] \\ &= \sum_j \mathbb{E}\left[b_j^2\right] \left(3(\Delta t_j)^2 - 2(\Delta t_j)^2 + (\Delta t_j)^2\right) = 2\sum_j \mathbb{E}\left[b_j^2\right] (\Delta t_j)^2 \to 0, \qquad \text{as } \|\Pi\| \to 0. \end{aligned} \tag{101}$$

Thus the fourth summand converges to

$$\int_0^T b dt = \int_0^T f_{xx}(t, B_t) dt. \tag{102}$$

Finally for the fifth summand, let $c_j = f_{tx}(t_j, B(t_j))$, so we can write that

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{j=0}^{n-1} f_{tx}(t_j, B(t_j))(t_{j+1} - t_j)(B(t_{j+1}) - B(t_j))\right)^2\right] &= \sum_j \mathbb{E}\left[c_j^2 (\Delta t_j)^2 (\Delta B_j)^2\right] + 2\sum_{i<j} \mathbb{E}\left[c_i c_j \Delta t_i \Delta t_j \Delta B_i \Delta B_j\right] \\ = \sum_j \mathbb{E}\left[c_j^2\right] (\Delta t_j)^2 \mathbb{E}\left[\Delta B_j^2\right] + 2\sum_{i<j} \mathbb{E}\left[c_i c_j\right] \Delta t_i \Delta t_j \mathbb{E}\left[\Delta B_i\right] \mathbb{E}\left[\Delta B_j\right] &= \sum_j \mathbb{E}\left[c_j^2\right] (\Delta t_j)^3 \to 0, \qquad \text{as } \|\Pi\| \to 0. \end{aligned} \tag{103}$$

Therefore, the fifth summand vanishes. Putting these summand together, we finally obtain that

$$f(T, B_T) = f(0, B_0) + \int_0^T f_t(t, B_t) dt + \int_0^T f_x(t, B_t) dB_t + \frac{1}{2} \int_0^T f_{xx}(t, B_t) dt, \tag{104}$$

as desired, so the proof is complete. □

**Remark 4.2.** One often writes Itô formulain the differential form, such that

$$df(t, B_t) = f_t(t, B_t) dt + f_x(t, B_t) dB_t + \frac{1}{2} f_{xx}(t, B_t) dt. \tag{105}$$

Yet the mathematically meaningful form of the Itô formula is (94). This is because we have precise definitions for all terms appearing on the right-hand side: the first and the third terms are ordinary Riemann integrals, while the second term is an Itô integral as we have defined.

**Example 4.3.** In the last lecture, we have shown that

$$\int_0^T B_t dB_t = \frac{B_T^2}{2} - \frac{T}{2}. \tag{106}$$

Let us show how Itô formula simplifies the computation of this Itô integral. Take $f(t,x) = \frac{x^2}{2}$, then the Itô formula says that

$$df(t,B_t) = f_t(t,B_t)dt + f_x(t,B_t)dB_t + \frac{1}{2}f_{xx}(t,B_t)dt = 0dt + B_t dB_t + \frac{1}{2}\cdot 1dt. \tag{107}$$

By integration, we can thus obtain that

$$\int_0^T B_t dB_t = \int_0^T df(t,B_t) - \frac{1}{2}\int_0^T dt = (f(T,B_T) - f(0,B_0)) - \frac{1}{2}(T-0) = \frac{B_T^2}{2} - \frac{T}{2}, \tag{108}$$

same as the result we have obtained earlier but much simpler to compute.

**Example 4.4.** Let $\beta_k(t) = \mathbb{E}\left[B_t^k\right]$. We will use the Itô formula to prove the recursive relation

$$\beta_k(t) = \frac{1}{2}k(k-1)\int_0^t \beta_{k-2}(s)ds, \qquad \text{for } k \geq 2. \tag{109}$$

Note that we can choose $f(t,x) = x^k$, then the Itô formula says that

$$df(t,B_t) = f_t(t,B_t)dt + f_x(t,B_t)dB_t + \frac{1}{2}f_{xx}(t,B_t)dt = 0dt + kB_t^{k-1}dB_t + \frac{1}{2}k(k-1)B_t^{k-2}dt. \tag{110}$$

Written in the integral form and taking expectation, this gives

$$\beta_k(t) = \mathbb{E}\left[B_t^k\right] = k\mathbb{E}\left[\int_0^t B_s^{k-1}dB_s\right] + \frac{1}{2}k(k-1)\mathbb{E}\left[\int_0^t B_s^{k-2}ds\right]. \tag{111}$$

Since $B_s^{k-1}$ is $\mathscr{F}_s$-adapted, we have that

$$\mathbb{E}\left[\int_0^t B_s^{k-1}dB_s\right] = 0. \tag{112}$$

Therefore, we can compute that

$$\beta_k(t) = \frac{1}{2}k(k-1)\mathbb{E}\left[\int_0^t B_s^{k-2}ds\right] = \frac{1}{2}k(k-1)\int_0^t \mathbb{E}\left[B_s^{k-2}\right]ds = \frac{1}{2}k(k-1)\int_0^t \beta_{k-2}(s)ds. \tag{113}$$

## 4.2 Itô Formula for General Diffusion

We extend the Itô formula to stochastic processes more general than Brownian motion.

**Definition 4.5.** An *Itô process* is a stochastic process $X_t$ on $(\Omega, \mathscr{F}, \mathbb{P})$ of the form

$$X_t = X_0 + \int_0^t \mu(s,w)ds + \int_0^t \nu(s,w)dB_s, \tag{114}$$

where $\nu \in V$ and the event $\mathbb{P}\left[\int_0^t \nu^2(s,w)ds < \infty, \forall t \geq 0\right] = 1$. We also assume that $\mu$ is $\mathscr{F}_t$-adapted and the event $\mathbb{P}\left[\int_0^t |\mu(s,w)|ds < \infty, \forall t \geq 0\right] = 1$.

The Itô process is often written in the differential form as

$$dX_t = \mu dt + \nu dB_t, \tag{115}$$

where $\mu dt$ is called the *drift*, and $\nu dB_t$ is called the *volatile part.*

**Theorem 4.6** (Itô's lemma)**.** Let $X_t$ be an Itô process given by (114). Let $g(t,x) \in C^2([0,\infty) \times \mathbb{R}])$, *i.e.*, $g$ is twice continuously differentiable on $[0,\infty) \times \mathbb{R}$. Then $Y_t = g(t, X_t)$ is again an Itô process, and we have that

$$dY_t = g_t(t, X_t)dt + g_x(t, X_t)dX_t + \frac{1}{2}g_{xx}(t, X_t)(dX_t)^2, \tag{116}$$

We note that

$$(dX_t)^2 = (\mu dt + \nu dB_t)^2 = \mu^2(dt)^2 + \nu^2(dB_t)^2 + 2\mu\nu dt dB_t = 0 + \nu^2 dt + 0 = \nu^2 dt. \tag{117}$$

Substituting this expression into the Itô's lemma, we thus have that

$$dY_t = g_t(t, X_t)dt + \frac{1}{2}g_x(t, X_t)(\mu dt + \nu dB_t) + g_{xx}(t, X_t)(\nu^2 dt) = \left(g_t + \mu g_x + \frac{\nu^2}{2}g_{xx}\right)dt + \nu g_x dB_t. \tag{118}$$

Again, the first term is the new *drift*, and the second term is the new *volatility term*.

## 4.3 Stochastic Differential Equations

The general stochastic differential equations can be written in the form

$$X_t = X_0 + \int_0^t b(s,w)ds + \int_0^t \sigma(s,w)dB_s. \tag{119}$$

This is stochastic since $\int_0^t \sigma(s,w)dB_s$ is a random component, and also $b = b(s,w)$ can be a random function. The solution to a stochastic differential equation is not a function, but a family of functions. In the short differential form, we can also write that

$$dX_t = bdt + \sigma dB_t, \qquad X(0) = X_0. \tag{120}$$

**Black-Scholes SDE.** The **Black-Schole equation** is given by

$$dS_t = rS_t dt + \sigma S_t dB_t, \tag{121}$$

with the initial condition $S(0) = S_0$, and the term $\sigma S_t dB_t$ represents uncertainty. To interpret this equation, $r$ is the risk-free interest rate (assumed constant), $\sigma$ is the volatility (also assumed constant), $B_t$ is a standard Brownian motion, and $S_t$ is the stock price. We first consider the ordinary differential equation without uncertainty, such that

$$dS_t = rS_t dt, \qquad S(0) = S_0. \tag{122}$$

This would solve to $S_t = S_0 \exp(rt)$. In other words, investing $S_0$ at time 0 would result in $S_0 \exp(rt)$ at time $t$. This is simply a formula for return on risk-free investment. Now Black-Scholes equation adds uncertainty to it. Let $Y_t = g(t, S_t) = \log S_t$, then the Itô formula gives that

$$dY_t = \left(0 + rS_t \cdot \frac{1}{S_t} + \frac{(\sigma S_t)^2}{2} \cdot \frac{-1}{S_t^2}\right)dt + \sigma S_t \cdot \frac{1}{S_t}dB_t = \left(r - \frac{\sigma^2}{2}\right)dt + \sigma dB_t. \tag{123}$$

Converting back to the integral form, this means that

$$\log S_t - \log S_0 = \left(r - \frac{\sigma^2}{2}\right)(t-0) + \sigma(B_t - B_0) = \left(r - \frac{\sigma^2}{2}\right)t + \sigma B_t \implies S_t = S_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right)t + \sigma B_t\right). \tag{124}$$

The solution to the Black-Scholes equation implies that $S_t \geq 0$ as long as $S_0 \geq 0$. The result above allows to calculate the prices for *vanilla European options*. In particular, the call price

$$c(t, S_0) = \exp(-rt)\,\mathbb{E}\left[(S_t - K)^+\right] = S_0 N(d_+(t, S_0)) - K\exp(-rt)\,N(d_-(t, S_0)), \tag{125}$$

where $N$ is the cumulative standard normal distribution and

$$d_\pm(t,x) = \frac{1}{\sigma\sqrt{t}}\left(\log\frac{x}{K} + \left(r \pm \frac{\sigma^2}{2}\right)t\right), \qquad N(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy. \tag{126}$$

**Ornstein-Uhlenbeck SDE.** To be done...

## 4/24 Lecture

# 5 Girsanov's Theorem

Recall that in standard probability theory, given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a nonnegative random variable $Z$ such that $\mathbb{E}[Z] = 1$, we can define a new probability measure $\tilde{\mathbb{P}}$ as

$$\tilde{\mathbb{P}}(A) = \int_A Z(w) dP(w), \qquad \forall A \in \mathscr{F}. \tag{127}$$

Let $\tilde{\mathbb{E}}$ denote the expectation taken under this new measure, then the expectations for any random variable $X$ under these two probability measures satisfy

$$\tilde{E}[Z] = \mathbb{E}[XZ]. \tag{128}$$

Conversely, if $\mathbb{P}[Z > 0] = 1$, then

$$\tilde{E}\left[\frac{X}{Z}\right] = \mathbb{E}[X]. \tag{129}$$

We say that $Z$ is the *Radon-Nikodỳm derivative* of $\tilde{\mathbb{P}}$ with respect to $\mathbb{P}$, and we write that

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}. \tag{130}$$

Now we extend this idea to stochastic processes. Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a filtration $\mathscr{F}_t$, we define the Radon Nikodỳm derivative as

$$Z(t) = \tilde{\mathbb{E}}[Z|\mathscr{F}_t]. \tag{131}$$

It is clear that $Z(t)$ is a martingale under $\mathbb{P}$, such that

$$\mathbb{E}[Z(t)|\mathscr{F}_s] = \mathbb{E}[\mathbb{E}[Z|\mathscr{F}_t]|\mathscr{F}_s] = \mathbb{E}[Z|\mathscr{F}_s] = Z(s). \tag{132}$$

**Lemma 5.1.** Let $Y$ be an $\mathscr{F}_t$-measurable random variable, then $\tilde{E}[Y] = \mathbb{E}[YZ(t)]$.

*Proof.* We can see that

$$\tilde{E}[Y] = \mathbb{E}[YZ] = \mathbb{E}[\mathbb{E}[YZ|\mathscr{F}_t]] = \mathbb{E}[Y\mathbb{E}[Z|\mathscr{F}_t]] = \mathbb{E}[YZ(t)]. \tag{133}$$

□

**Lemma 5.2.** Let $Y$ be an $\mathscr{F}_t$-measurable random variable, then $Z(s)\tilde{\mathbb{E}}[Y|\mathscr{F}_s] = \mathbb{E}[YZ(t)|\mathscr{F}_s]$.

*Proof.* By definition of conditional expectation, we need to show that

$$\int_A Y d\tilde{\mathbb{P}} = \int_A \frac{1}{Z(s)} \mathbb{E}[YZ(t)|\mathscr{F}_s] d\tilde{\mathbb{P}}, \qquad \forall A \in \mathscr{F}_s. \tag{134}$$

Note that

$$\begin{aligned}\int_A \frac{1}{Z(s)} \mathbb{E}[YZ(t)|\mathscr{F}_s] d\tilde{\mathbb{P}} &= \tilde{\mathbb{E}}\left[\mathbb{1}_A \frac{1}{Z(s)} \mathbb{E}[YZ(t)|\mathscr{F}_s]\right] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[YZ(t)|\mathscr{F}_s]] = \mathbb{E}[\mathbb{E}[\mathbb{1}_A YZ(t)|\mathscr{F}_s]] \\ &= \mathbb{E}[\mathbb{1}_A YZ(t)] = \tilde{\mathbb{E}}[\mathbb{1}_A Y] = \int_A Y d\tilde{\mathbb{P}},\end{aligned} \tag{135}$$

so the proof is complete. □

Before we prove Girsanov's theorem, we first state (without proof) Lévy's theorem.

**Theorem 5.3** (Lévy's theorem)**.** Let $M(t)$ be a martingale relative to $\mathscr{F}_t$. Assume that $M(0) = 0$ and $M(t)$ has continuous paths. Furthermore, the quadratic variation $[M, M](t) = t$ for all $t \geq 0$. Then $M(t)$ is a Brownian motion.

Now we state and prove Girsanov's theorem.

**Theorem 5.4** (Grisanov's theorem)**.** Let $B(t)$ be a Brownian motion relative to $\mathscr{F}_t$, $0 \leq t \leq T$, and $\Theta(t)$ be an adapted process. Define

$$Z(t) = \exp\left(-\int_0^t \Theta(u)dB(u) - \frac{1}{2}\int_0^t \Theta^2(u)du\right), \tag{136}$$

$$\tilde{B}(t) = B(t) + \int_0^t \Theta(u)du, \tag{137}$$

and assume that

$$\mathbb{E}\left[\int_0^T \Theta^2(u)Z^2(u)du\right] < \infty. \tag{138}$$

Set $Z = Z(t)$, then $\mathbb{E}[Z] = 1$ and under $\tilde{\mathbb{P}}$ as defined before, $\tilde{B}(t)$ is a Brownian motion.

*Proof.* It is clear that $\tilde{B}(0) = 0$ and $\tilde{B}(t)$ has continuous paths. Also, $[\tilde{B}, \tilde{B}](t) = [B, B](t) = t$, since the term $\int_0^t \Theta(u)du$ contributes zero quadratic variation. Therefore by Lévy's theorem, we only need to show that $\tilde{B}(t)$ is a martingale under $\tilde{\mathbb{P}}$ in order to conclude that $\tilde{B}(t)$ is a Brownian motion under $\tilde{\mathbb{P}}$. Let

$$X(t) = -\int_0^t \Theta(u)dB(u) - \frac{1}{2}\int_0^t \Theta^2(u)du, \tag{139}$$

then by Itô's formula, we have that

$$\begin{aligned} dZ(t) = d\left(\exp(X(t))\right) &= \exp(X(t))dX(t) + \frac{1}{2}\exp(X(t))(dX(t))^2 \\ &= \exp(X(t))\left(-\Theta(t)dB(t) - \frac{1}{2}\Theta^2(t)dt\right) + \frac{1}{2}\exp(X(t))\Theta^2(t)dt = -\Theta(t)Z(t)dB(t). \end{aligned} \tag{140}$$

Therefore, $Z(t)$ is a martingale under $\mathbb{P}$, so we have that $Z(t) = \mathbb{E}[Z|\mathscr{F}_t]$, which means that $Z(t)$ can be seens as a Radon-Nikodỳm derivative process and the previous lemmas can be applied. Next we show that $\tilde{B}(t)Z(t)$ is a martingale under $\mathbb{P}$. Note that

$$\begin{aligned} d\left(\tilde{B}(t)Z(t)\right) &= \tilde{B}(t)dZ(t) + Z(t)d\tilde{B}(t) + d\tilde{B}(t)dZ(t) \\ &= -\tilde{B}(t)\Theta(t)Z(t)dB(t) + Z(t)\left(dB(t) + \Theta(t)dt\right) - \left(dB(t) + \Theta(t)dt\right)\Theta(t)Z(t)dB(t) \\ &= -\tilde{B}(t)\Theta(t)Z(t)dB(t) + Z(t)dB(t) + Z(t)\Theta(t)dt - \Theta(t)Z(t)(dB(t))^2 - \Theta^2(t)Z(t)dtdB(t) \\ &= -\tilde{B}(t)\Theta(t)Z(t)dB(t) + Z(t)dB(t) + Z(t)\Theta(t)dt - \Theta(t)Z(t)dt \\ &= \left(-\tilde{B}(t)\Theta(t) + 1\right)Z(t)dB(t). \end{aligned} \tag{141}$$

Finally, we can see that

$$\tilde{\mathbb{E}}[\tilde{B}(t)|\mathscr{F}_s] = \frac{1}{Z(s)}\mathbb{E}[\tilde{B}(t)Z(t)|\mathscr{F}_s] = \frac{1}{Z(s)}\tilde{B}(s)Z(s) = \tilde{B}(s), \tag{142}$$

so indeed $\tilde{B}(t)$ is a martingale under $\tilde{\mathbb{P}}$, and the proof is complete by Lévy's theorem. □

# 5/1 Lecture

# 6 Stochastic Calculus and Partial Differential Equations

Throughout this section, we will be using the following lemma.

**Lemma 6.1.** Assume we are given a random variable $X$ on $(\Omega, \mathscr{F}, \mathbb{P})$ and a filtration $(\mathscr{F}_t)_{t\geq 0}$. Then $\mathbb{E}[X|\mathscr{F}_t]$ is a martingale with respect to filtration $(\mathscr{F}_t)_{t\geq 0}$.

*Proof.* The proof follows directly from the towering property of conditional expectation. □

**Corollary 6.2.** Let $X_t$ be a Markov process and $\mathscr{F}_t$ be the natural filtration associated with this process. Then according to the above lemma, for any function $V$, the process $\mathbb{E}[V(X_T)|\mathscr{F}_t]$ is a martingale, and applying Markov property, we get that $\mathbb{E}[V(X_T)|X_t]$ is a martingale. In the following we often write $\mathbb{E}[V(X_T)|X_t]$ as $\mathbb{E}_{X_t=x}[V(X_T)]$.

## 6.1 Expected Value of Payoff

Assume that $X_t$ is a stochastic process satisfying the following stochastic differential equation

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dB_t, \tag{143}$$

or in the integral form

$$X_t - X_0 = \int_0^t a(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s. \tag{144}$$

Let $u(t, x) = \mathbb{E}_{X_t=x}[V(X_T)]$ be the expected value of some payoff $V$ at maturity $T > t$ given that $X_t = x$. Then the expected value of payoff $u(t, x)$ solves

$$u_t + a(t, x)u_x + \frac{1}{2}\sigma^2(t, x)u_{xx} = 0, \qquad t < T, \tag{145}$$

with $u(T, x) = V(x)$. By the corollary we can conclude that $u(t, x)$ is a martingale. Applying Itô's lemma, we get

$$du(t, X_t) = u_t dt + u_x dX_t + \frac{1}{2}u_{xx}(dX_t)^2 = \left(u_t + au_x + \frac{1}{2}\sigma^2 u_{xx}\right) dt + \sigma u_x dB_t. \tag{146}$$

Since $u(t, x)$ is a martingale, the drift term must be zero, and thus $u(t, x)$ solves

$$u_t + au_x + \frac{1}{2}\sigma^2 u_{xx} = 0, \tag{147}$$

and $u(T, x) = \mathbb{E}_{X_T=x}[V(X_T)] = V(x)$.

## 6.2 Feynman-Kac Formula

Suppose that we are interested in a suitably "discounted" final-time payoff of the form

$$u(t, x) = \mathbb{E}_{X_t=x}\left[\exp\left(-\int_t^T b(s, X_s)ds\right) V(X_T)\right], \tag{148}$$

for some specific function $b(t, X_t)$. We will show that $u$ then solves

$$u_t + a(t, x)u_x + \frac{1}{2}\sigma^2(t, x)u_{xx} - b(t, x)u = 0. \tag{149}$$

with $u(T, x) = V(x)$. The fact that $u(T, x) = V(x)$ is clear from the definition of the function $u$. Therefore, let us focus on the proof of the differential equation. Our strategy is to apply the corollary and thus we have to find some martingale involving $u(t, x)$. For this reason, we consider

$$\begin{aligned}\exp\left(-\int_0^t b(s, X_s)ds\right) u(t, x) &= \exp\left(-\int_0^t b(s, X_s)ds\right) \mathbb{E}_{X_t=x}\left[\exp\left(-\int_t^T b(s, X_s)ds\right) V(X_T)\right] \\ &= \mathbb{E}_{X_t=x}\left[\exp\left(-\int_0^T b(s, X_s)ds\right) V(X_T)\right]. \end{aligned} \tag{150}$$

According to the corollary, the above is a martingale. Applying Itô's lemma, we get

$$d\left(\exp\left(-\int_0^t b(s,X_s)ds\right)u(t,x)\right)$$
$$= \left(u_t + a(t,x)u_x + \frac{1}{2}\sigma^2(t,x)u_{xx} - b(t,x)u\right)\exp\left(-\int_0^t b(s,X_s)ds\right)dt + \exp\left(-\int_0^t b(s,X_s)ds\right)u_x dB_t. \tag{151}$$

Since the drift of a martingale must be zero, we obtain the desired differential equation.

**Example 6.3** (Black-Scholes SDE)**.** We assume that the underlying (stock, for instance) follows a geometric Brownian motion. That is, in the risk-neutral measure, it satisfies the SDE

$$dS_t = rS_t dt + \sigma S_t dB_t, \tag{152}$$

where $r$ is the risk-free rate which we assume to be constant. The payoff of the European option at maturity $T$ is known and is equal to $V(S_T)$. Then to find the value of the option at some earlier time $t < T$, we have to compute

$$\mathbb{E}_{S_t=x}\left[\exp(-rt)V(S_T)\right]. \tag{153}$$

From the Feynman-Kac formula, we conclude that $u(t,x)$ solves the partial differential equation

$$u_t + rxu_x + \frac{1}{2}\sigma^2 x^2 u_{xx} - ru = 0, \tag{154}$$

with $u(T,x) = V(x)$. As above is the famous Black-Scholes SDE.

## 6.3 Running Payoff

Now suppose that we are interested in

$$u(t,x) = \mathbb{E}_{X_t=x}\left[\int_t^T b(s,X_s)ds\right], \tag{155}$$

for some specified function $b(t,x)$. First of all, let us find the final-time condition for $u(t,x)$. Clearly, $u(T,x) = 0$. Our next step is to find a martingale involving $u(t,x)$ so as to use the corollary. Note that

$$u(t,x) + \int_0^t b(s,X_s)ds = \mathbb{E}_{X_t=x}\left[\int_t^T b(s,X_s)ds\right] + \int_0^t b(s,X_s)ds = \mathbb{E}_{X_t=x}\left[\int_0^T b(s,X_s)ds\right], \tag{156}$$

which is a martingale. Applying Itô's lemma, we get

$$u_t + a(t,x)u_x + \frac{1}{2}\sigma^2(t,x)u_{xx} + b(t,x) = 0, \tag{157}$$

analogous to previous deductions (using the fact that the drift of a martingale must be zero).

## 6.4 Boundary Value Problems and Exit Times

In previous examples, we were interested in the expectation of the form $\mathbb{E}_{X_t=x}[V(X_T)]$, that is, the expectation of some payoff at specified maturity $T$. Now let us assume that we are given a region $D \subseteq \mathbb{R}$ and a process $X_t$ starting from some point $x \in D$. Let

$$\tau(x) = \min(T, \inf\{t;\ X_t \notin D\}). \tag{158}$$

That is, $\tau(x)$ is the first time $X_t$ exits from the region $D$ if prior to $T$, otherwise $\tau(x) = T$. Assume that at exit time $\tau$, the payoff of an optionis given by the function $V$. We are interested in the fair price of such an option at some earlier time $t$, *i.e.*, in the following quantity

$$u(t,x) = \mathbb{E}_{X_t=x}[V(\tau, X_\tau)]. \tag{159}$$

We will see that just like in the previous examples, $u(t,x)$ solves a partial differential equation, but in contrast, the PDE must be solve inside the region $D$ with suitable boundary data. The key to derivation of the PDE is the following lemma.

**Lemma 6.4.** Let $\mathbb{E}[\tau] < \infty$. Then $\mathbb{E}[V(X_\tau)|\mathscr{F}_t]$ is a martingale with respect to filtration $(\mathscr{F}_{t\wedge\tau})_{t\geq 0}$.

Applying Itô's lemma, we get

$$du(t,x) = \left(u_t + au_x + \frac{1}{2}\sigma^2 u_{xx}\right) dt + u_x \sigma dB_t. \tag{160}$$

By the lemma above, $u(t,x)$ is a martingale and thus has no drift term. Thus $u(t,x)$ solves the following PDE

$$u_t + au_x + \frac{1}{2}\sigma^2 u_{xx} = 0, \tag{161}$$

with boundary conditions $u(t,x) = V(t,x)$ for $x \in \partial D$ and $u(T,x) = V(T,x)$ for $x \in D$. An application is the distribution of the first arrivals. Consider $\mathbb{E}_{X_t=x}[\mathbb{1}_{\tau<T}] = \mathbb{P}_{X_t=x}[\tau < T]$. According to above, it suffices to solve

$$u_t + au_x + \frac{1}{2}\sigma^2 u_{xx} = 0, \tag{162}$$

with boundary condition $u = 1$ on $x \in \partial D$.

---

**Last Modified: May 8, 2023.**