Learning Report on:

Stochastic Calculus: An Introduction with Applications

By Gregory F. Lawler

Colin Wan

January 1, 2020

This is a self-studying, learning report on Stochastic Calculus using *Stochastic Calculus: An Introduction with Applications* by Gregory F. Lawler. Other references used are, but not limited to, *Real Analysis: Modern Techniques and Their Applications* by Gerald B. Folland, *Almost None of the Theory of Stochastic Processes* by Cosma Rohilla Shalizi, *Brownian Motion, Martingales, and Stochastic Calculus* by Jean-François Le Gall.

0 Background Knowledge

0.1 Basic Measure Theory and Notations

Definition 0.1. ***σ-algebra***
A non-empty subset $\mathcal{A}$ of Ω is a σ-algebra if: for $E \in \Omega$

1. $E \in \mathcal{A} \implies E^c \in \mathcal{A}$
2. $\forall E_1, E_2, ... \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$

Remarks

1. $\Omega \in \mathcal{A}$ & $\emptyset \in \mathcal{A}$
2. $\forall E_1, E_2, ... \in \mathcal{A} \implies \bigcap_{i=1}^{\infty} E_i \in \mathcal{A}$

Definition 0.2. ***measure***
A measure μ on Ω with σ algebra $\mathcal{A}$ is a function $\mu : \mathcal{A} \to [0, \infty)$ if:

1. $\mu(\emptyset) = 0$
2. $\mu(\bigcup_{i=1}^{\infty} E_i) = \Sigma_{i=1}^{\infty} \mu(E_i)$

Additionally, a measure $\mathcal{P}$ is a probability measure if $\mathcal{P}(\Omega) = 1$

Definition 0.3. ***Random Variable***

Given $(\Omega, \mathcal{A}, \mathcal{P})$*, a random variable is a function* $X : \Omega \to \mathbb{R}$ *s.t.* $\forall x \in \mathbb{R}\ \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$

Remarks

1. *A random variable* X *that satisfy* $\forall x \in \mathbb{R}\ \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$ is called *measurable by* $\mathcal{A}$

Definition 0.4. ***Filtration***

If $X_1, X_2....$ *is a sequence of random variables, then the associated* ***filtration*** *is the collection* $\mathcal{F}_n$ *where* $\mathcal{F}_n$ *denote the information in* $X_1, X_2....X_n$

To illustrate by an example: let's go with a simple coin flipping, and we are interested in the results of two flips. Then $\Omega = \{HH, TT, HT, TH\}$ *At time 0, we know nothing about the outcome after two flips, therefore the information contained in* $\mathcal{F}_0 = \{\emptyset, \Omega\}$ *At time 1, after 1 flip, we can observe the result of first flip and know more about the experiment. Hence, we know these events:* $\mathcal{F}_1 = \{\emptyset, \Omega, \{HT, HH\}, \{TT, TH\}\} \supset \mathcal{F}_0$ *could happen. At time 2, after 2 flips, we observe the final result of the experiment and know everything about the outcome. Hence we know these events:* $\mathcal{F}_2 = \{\emptyset, \Omega, \{HT, HH\}, \{TT, TH\}, \{TT\}, \{TH\}, \{HH\}, \{HT\}\} \supset \mathcal{F}_1$ *could happen.*

1 Martingales in Discrete Time

1.1 Conditional Expectation

Given probability space $(\Omega, \mathcal{A}, \mathcal{P})$*, and integrable random variable* X*. Let* $\mathcal{G}$ *be a sub* σ*-algebra of* $\mathcal{A}$*. Then* $E[X \mid \mathcal{G}]$ *is defined to be the unique* $\mathcal{G}$ *measurable random variable such that if* $A \in \mathcal{G}$*,*

$$\mathbb{E}[X 1_A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] 1_A] \tag{1}$$

Proposition 1.1. ***Properties of conditional expectation***

Suppose $X_1, X_2, ...$ *is a sequence of random variable and* $\mathcal{F}_n$ *be the corresponding filtration at time* n*. Then for a random variable* Y*:*

- *Give* Y *is* $\mathcal{F}_n$ *measurable, then* $E[Y \mid \mathcal{F}_n] = Y$
- *Given* A *is an* $\mathcal{F}_n$ *measurable event, then* $E[E[Y \mid \mathcal{F}_n] 1_A] = E[Y 1_A]$
- *Given* $\{X_i\}$ *is independent from* Y*, then* $E[Y \mid \mathcal{F}_n] = E[Y]$
- *Given random variable* Y, Z*, and constants* $a, b \in \mathbb{R}$*, then*

$$E[aY + bZ] = aE[Y] + bE[Z]$$

- *Given* $m, n \in \mathbb{N}$ *and* $m < n$*, then* $E[E[Y \mid \mathcal{F}_n] \mid \mathcal{F}_m] = E[Y \mid \mathcal{F}_m]$
- *Given* $\mathcal{F}_n$ *measurable random variable* Z*, then* $E[YZ \mid \mathcal{F}_n] = ZE[Y \mid \mathcal{F}_n]$

1.2 Martingales

Definition 1.1. *A sequence of random variables* $M_0, M_1, ...$ *is called a* ***martingale*** *with respect to the filtration* $\mathcal{F}_n$ *if:*

- $\forall N \in \mathbb{N}$, M_n *is* $\mathcal{F}_n$ *measurable with* $E[|M_n|] < \infty$
- *If* $m, n \in \mathbb{N}$ *and* $m < n$*, then*

$$E[M_n \mid \mathcal{F}_m] = M_m \text{ or } E[M_n - M_m \mid \mathcal{F}_m] = 0$$

1.3 Optional Sampling(Stopping) Theorem

This section focuses on a new concept, stopping time. Motivated by studying the behavior of a martingale up-to a certain time.

Definition 1.2. ***Stopping time***
A non-negative integer-valued random variable T *is a* ***stopping time*** *with respect to filtration* $\{\mathcal{F}_n\}$ *if* $\forall n \in \mathbb{N}$*, the event* $\{T = n\}$ *is* $\mathcal{F}_n$*-measurable.*

For convenience, the following notes will use a new notation, $M_{n \wedge T}$, to indicate

$$M_0 + \Sigma_{j=1}^{n} B_j[M_j - M_{j-1}]$$

where $n \wedge T$ means $min\{n, T\}$, and $B_j = 1$ for $j \leq T$ and $B_j = 0$ otherwise.
The following three theorems will yield the same result, yet the precondition will be less strict as we progress.

Theorem 1.1. ***Optional Sampling Theorem I***
(Named Option Stopping Lemma in STA447) *Suppose* T *is a stopping time and* M_n *is a martingale with respect to* $\{\mathcal{F}_n\}$*. Then* $Y_n = M_{n \wedge T}$ *is a martingale. If* T *is bounded, or if there exists a* $K \in \mathbb{R}, K \leq \infty$*, such that,* $\mathbb{P}\{T \leq K\} = 1$*, then*

$$E[M_T] = E[M_0] \tag{2}$$

First, we should note that even without the final precondition, as long as M_n is a martingale with respect to $\mathcal{F}_n$, then

$$E[M_{n \wedge T}] = E[M_0] \tag{3}$$

Proof. (3)
$\forall n \in \mathbb{N}$, WLOG, assume $n > T$

$$\begin{aligned} E[M_{n \wedge T}] &= E[M_0 + \Sigma_{j=1}^{n} B_j[M_j - M_{j-1}]] \\ &= E[M_0] + \Sigma_{j=1}^{n} B_j E[M_j - M_{j-1}] \\ &= E[M_0] + \Sigma_{j=1}^{T} 1 * E[M_j - M_{j-1}] + \Sigma_{j=T}^{n} 0 * E[M_j - M_{j-1}] \\ &= E[M_0] \leq \infty \end{aligned}$$

$\forall m, n \in \mathbb{N}$ and $m < n$

$$\begin{aligned} E[M_{n\wedge T} - M_{n-1\wedge T} \mid \mathcal{F}_m] &= E[B_n[M_n - M_{n-1}] \\ &= E[M_n - M_{n-1}] = 0 \text{ if } B_n = 1 \\ &= 0 \text{ if } B_n = 0 \end{aligned}$$

□

The proof of Theorem 1.1 is relatively straight forward: since T is bounded, $E[M_T] - E[M_0]$ can be separated into sums of finite steps. (i.e. Finitely many $E[M_i] - E[M_{i-1}]$) We have showed each step is equal to 0, therefore the sum is still 0.

However, if we were to change the last precondition of Theorem 1.1 to something less restrictive. Say instead of bounding T by K for some $K \in \mathbb{R}$, we only require $P\{T < \infty\} = 1$. Then (3) will still hold, and

$$E[M_0] = E[M_{n\wedge T}] = E[M_T] + E[M_{n\wedge T} - M_T].$$

If the latter term of the right hand equals to 0 for large n, then we will have (2). $M_{n\wedge T} - M_T$ is obviously 0 if $n \wedge T = T$. If $n > T$, then we have

$$M_{n\wedge T} - M_T = 1\{T > n\}[M_n - M_T].$$

Since $M_T 1\{T > n\}$ is a random variable converging to M_T and bounded by the random variable $|M_T| < \infty$, hence by the dominated convergence theorem, $\lim_{n\to\infty} E[M_T 1\{T > n\}] = 0$. Therefore, we just need the other term to behave nicely.

Theorem 1.2. ***Optional Sampling Theorem II***
(Named Option Stopping Theorem in STA447) *Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Suppose that $P\{T < \infty\} = 1$, and*

$$\lim_{n\to\infty} E[|M_n| 1\{T > n\}] = 0, \tag{4}$$

then,

$$E[M_T] = E[M_0]$$

Let us go a step further and examine (4). Start by separating (4) into two parts base on the value of each M_n, let $b \in \mathbb{R}$:

$$
\begin{aligned}
E[|M_n|1\{T>n\}] &= \boldsymbol{E[|M_n|1\{|M_n| \geq b, T>n\}]} + E[|M_n|1\{|M_n|<b, T>n\}] \\
&\leq \boldsymbol{\frac{1}{b}E[|M_n|^2 1\{|M_n| \geq b, T>n\}]} + E[|M_n|1\{|M_n|<b, T>n\}] \\
&\leq \boldsymbol{\frac{1}{b}(E[|M_n|^2 1\{|M_n| \geq b, T>n\}] + E[|M_n|^2 1\{|M_n|<b, T>n\}]} \\
&\quad \boldsymbol{+E[|M_T|^2 1\{T<n\}])} + E[|M_n|1\{|M_n|<b, T>n\}] \\
&\leq \boldsymbol{\frac{1}{b}(E[|M_n|^2 1\{T>n\}] + E[|M_T|^2 1\{T<n\}])} \\
&\quad + E[|M_n|1\{|M_n|<b, T>n\}] \\
&\leq \frac{E[|M_{n\wedge T}|^2]}{b} + bP\{T>n\}
\end{aligned}
$$

Now, let's bound $E[|M_{n\wedge T}|^2] < C$, for some $C \in \mathbb{R}$. Then we have

$$E[|M_n|1\{T>n\}] \leq \tfrac{C}{b} + bP\{T>n\}.$$

Continue with the inequality we just proved. First note that $E[|M_n|1\{T>n\}]$ and $\frac{C}{b} + bP\{T>n\}$ are sequences with respect to n. Moreover, $\frac{C}{b} + bP\{T>n\}$ is monotonically decreasing. Since $E[|M_n|1\{T>n\}]$ is bounded by a monotonically decreasing sequence, we have

$$
\begin{aligned}
\limsup_{n\to\infty} E[|M_n|1\{T>n\}] &\leq \limsup_{n\to\infty} \frac{C}{b} + P\{T>n\} \\
\limsup_{n\to\infty} E[|M_n|1\{T>n\}] &\leq \frac{C}{b} + \lim_{n\to\infty} P\{T>n\} \\
\limsup_{n\to\infty} E[|M_n|1\{T>n\}] &\leq \frac{C}{b}
\end{aligned}
$$

and,

$$0 \leq \lim_{n\to\infty} E[|M_n|1\{T>n\}] \leq \limsup_{n\to\infty} E[|M_n|1\{T>n\}] \leq \tfrac{C}{b}$$

Since the above inequality holds for all b, we have (4). This results in the final Optional Sampling Theorem.

Theorem 1.3. ***Optional Sampling Theorem III***

Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Suppose that $P\{T<\infty\} = 1$, and there exists $C<\infty$ such that for each n,

$$E[|M_{n\wedge T}|^2] \leq C \tag{5}$$

Then,

$$E[M_T] = E[M_0]$$

1.4 Martingale Convergence Theorem

Theorem 1.4. *Martingale Convergence Theorem*

Suppose M_n is a martingale with respect to $\{\mathcal{F}_n\}$ and there exists some $C \in \mathbb{R}$ such that $E[|M_n|] \leq C$ for all $n \in \mathbb{N}$. Then there exists a random variable M such that with probability one

$$\lim_{n\to\infty} M_n = M.$$

Proof. This proof of martingale convergence theorem will show that a bounded martingale will fluctuate finitely many times outside of any interval. i.e. for any $a, b \in \mathbb{R}$ and $a < b$, then there exist $K \in \mathbb{R}$ such that

$$|\{n : M_n \leq a, M_{n-1} > a\} \cup \{n : M_n \geq b, M_{n-1} < b\}| < K$$

Therefore, $\liminf M_n = \limsup M_n$ and hence the limit of $\lim M_n$ exists.

Start by define a sequence of stopping times: for any $a, b \in \mathbb{R}$ and $a < b$,

$$S_1 = \{n : M_n \leq a\}, \qquad T_1 = \{n : M_n \geq b, n > S_1\}$$

and for $i > 1$,

$$S_i = \{n : M_n \leq a, n > T_{i-1}\}, \qquad T_i = \{n : M_n \geq b, n > S_i\}$$

Simply speaking, S_1 is the first time M_n goes below a, T_1 is the first time M_n goes above b after S_1. Then S_2 is the first time M_n goes below a after T_1, and so on and so forth. Now define another martingale:

$$W_n = \Sigma_{i=1}^{n} B_i[M_{i+1} - M_i]$$

where,

$$B_i = 0 \ \textit{If } n < S_1$$
$$B_i = 1 \ \textit{If for some } j,\ S_j \leq i < T_j$$
$$B_i = 0 \ \textit{If for some } j,\ T_j \leq i < S_{j+1}$$

In other words, W_n records the change of M_n between each time M_n goes below a and the next time it goes above b.

It can be shown that W_n is also a martingale, which means $E[W_n] = E[W_0] = 0$. Now define U_n to be the count of the total number of "interval" recorded by W_n up to time n. i.e.

$$U_n = j, \ \textit{for } T_j \leq n < T_{j+1}$$

Then, WLOG assume $n > T_N$, where T_N represents the last time $M_n \geq b$,

$$W_n \geq U_n(b - a) + (M_n - a)$$

Using the property of margingale,

$$E[U_n](b - a) - E[a - M_n] \leq E[W_n] = 0$$
$$E[U_n](b - a) \leq E[a - M_n]$$
$$E[U_n](b - a) \leq |a| + E[|M_n|] = |a| + C$$
$$E[U_n] \leq \frac{|a| + C}{b - a}$$

Since this inequality holds for any $a, b \in \mathbb{R}$ we have $\liminf M_n = \limsup M_n$ and hence the limit of $\lim M_n = M$ exists. (Note, M can not be $\pm\infty$ with a positive probability, as if it is, then $E[|M_n|]$ cannot be bounded by C

□

1.5 Square Integrable Martingales

Definition 1.3. *Square Integrable Martingale A martingale M_n that is, for each n, $E[M_n^2] \leq \infty$*

Definition 1.4. *Orthogonality Two random variables are considered to be orthogonal if $E[XY] = E[X]E[Y]$.*

An important property is associated with martingales, that is the orthogonality between any two martingale increment.

Proposition 1.2. *Suppose that M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Then if $m < n$,*

$$E[(M_{n+1} - M_n)(M_{m+1} - M_m)] = 0 \tag{6}$$

Proof. Given that $m < n$, then $M_{m+1} - M_m$ is $\mathcal{F}_n$-measurable, and hence

$$\begin{aligned} &E[(M_{n+1} - M_n)(M_{m+1} - M_m)|\mathcal{F}_n] \\ &\qquad = (M_{m+1} - M_m)E[(M_{n+1} - M_n)|\mathcal{F}_n] = 0 \end{aligned}$$

Taking expectation of $\mathcal{F}_n$ again,

$$\begin{aligned} &E[(M_{n+1} - M_n)(M_{m+1} - M_m)] \\ &\qquad = E[(M_{m+1} - M_m)E[(M_{n+1} - M_n)|\mathcal{F}_n]] = 0 \end{aligned}$$

□

1.6 Integrals with respect to random walk

This section introduces discrete integral of martingales.

Definition 1.5. *Predictable A sequence of random variables, X_n, is called predictable with respect to $\{\mathcal{F}_n\}$, if for each n, X_n is $\mathcal{F}_{n-1}$-measurable.*

Suppose that $\{X_n\}$ is a set of identical independently distributed random variable with mean zero and variance σ^2. Let $S_n = \Sigma_{i=1}^n X_i$, and $\{\mathcal{F}_n\}$ be the filtration generated by $\{X_n\}$. Now define J_n to be predictable sequence with $E[J_n^2] < \infty$ for each n. The integral of J_n with respect to S_n is defined by

$$Z_n = \Sigma_{i=1}^n J_i X_i = \Sigma_{i=1}^n J_i \Delta S_i \tag{7}$$

Three important properties immediately presents themselves.

1. **Martingale property**. The integral Z_n is a martingale with respect to $\{\mathcal{F}_n\}$

2. **Linearity**. If J_n, K_n are predictable sequences and $a, b \in \mathbb{R}$, then $aJ_n + bK_n$ is a predictable sequence and

$$\Sigma_{i=1}^n aJ_i + bK_i = a\Sigma_{i=1}^n J_i \Delta S_i + b\Sigma_{i=1}^n K_i \Delta S_i$$

3. **Variance Rule**.

$$Var[\Sigma_{i=1}^n J_i \Delta S_i] = E[(\Sigma_{i=1}^n J_i \Delta S_i)^2] = \sigma^2 \Sigma_{i=1}^n [J_i]^2$$

Proof. Properties

1. Given $\{J_n\}$ and $\{S_n\}$ are $\{\mathcal{F}_n\}$-measurable, Z_n, as a finite sum of products of $\{\mathcal{F}_n\}$-measurable random variable, is $\{\mathcal{F}_n\}$-measurable. Also,

$$E[Z_n - Z_{n-1}|\mathcal{F}_n] = E[J_n X_n|\mathcal{F}_n] = J_n E[X_n|\mathcal{F}_n] = 0$$

2. This property is immediate

3. First note Z_n has mean 0, so the first part of equality holds. Then due to the orthogonality of martingale increments

$$E[(\Sigma_{i=1}^n J_i \Delta S_i)^2] = \Sigma_{i=1}^n E[J_i^2 X_i^2]$$

Then using the double expectation property over $\mathcal{F}_{i-1}$ for each i

$$\begin{aligned}
\Sigma_{i=1}^n E[J_i^2 X_i^2] &= \Sigma_{i=1}^n E[E[J_i^2 X_i^2|\mathcal{F}_{i-1}]] \\
&= \Sigma_{i=1}^n E[J_i^2 E[X_i^2|\mathcal{F}_{i-1}]] \\
&= \Sigma_{i=1}^n E[J_i^2 E[X_i^2]] \\
&= \Sigma_{i=1}^n \sigma^2 E[J_i^2] \\
&= \sigma^2 \Sigma_{i=1}^n E[J_i^2]
\end{aligned}$$

□

1.7 A maximal inequality

I believe this is the *Doob's martingale inequality*?

Definition 1.6. ***Submartingale***
*A sequence of random variables $M_0, M_1, ...$ is called a **submartingale** with respect to the filtration $\mathcal{F}_n$ if:*

- $\forall N \in \mathbb{N}$, *M_n is $\mathcal{F}_N$ measurable with* $E[|M_n|] < \infty$
- *If $m, n \in \mathbb{N}$ and $m < n$, then*

$$E[M_n \mid \mathcal{F}_m] \geq M_m$$

Theorem 1.5. *Suppose M_n is a non-negative submartingale with respect to $\{\mathcal{F}_n\}$, and let*

$$\overline{M_n} = max(\{M_i\}_{i=0}^n)$$

Then for every $a \in \mathbb{R}$, $a > 0$,

$$P\{\overline{M_n} \geq a\} \leq \frac{1}{a}E[M_n]$$

Proof. First define τ_a to be $inf\{i \geq 1 : M_i \geq a\}$, then

$$P(\overline{M_n} \geq a) = \Sigma_{i=1}^n P(\tau_a = i).$$

Note $E[1\{M_i \geq a\}] \leq E[\frac{M_i}{a}]$, and $\{\tau_a = i\}$ is $\mathcal{F}_i$ measurable. Then for each i, such that $1 \leq i \leq n$,

$$\begin{aligned} P(\tau_a = i) &= E[1\{\tau_a = i\}] \\ &\leq E[\frac{M_i}{a}1\{\tau_a = i\}] \\ &\leq \frac{1}{a}E[M_i 1\{\tau_a = i\}] \\ &\leq \frac{1}{a}E[1\{\tau_a = i\}E[M_n|\mathcal{F}_i]]] \\ &= \frac{1}{a}E[1\{\tau_a = i\}M_n] \end{aligned}$$

Summing over $1 \leq i \leq n$ we have.

$$\begin{aligned} \Sigma_{i=1}^n P(\tau_a = i) &\leq \frac{1}{a}\Sigma_{i=1}^n E[1\{\tau_a = i\}M_n] \\ &\leq \frac{1}{a}E[1\{\overline{M_n} \geq a\}M_n] \\ &\leq \frac{1}{a}E[M_n] \end{aligned}$$

which is the desired statement. □

2 Brownian Motion

2.1 Limit of Sum of Independent Variables

2.2 Multivariate Normal

The first section of this chapter covers basic properties of limit of sums such as CLT and binomial converge to Poisson. The second part covers multivariate normal distribution properties such as the role of covariance matrix and independence between the sum and difference of two normal variables.

2.3 Limit of Random Walks

This section discusses the limit of a simple symmetric random walk and how it approaches something continuous (intuitively) as the length of each time interval decreases.

Suppose $X_1, X_2...$ are independent random variables with

$$\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = -1\} = \frac{1}{2}$$

Then define,

$$S_n = \Sigma_{i=0}^{n} X_i$$

be the corresponding SSRW. As in the discrete case, this SSRW have time increment $\Delta t = 1$ and space increment $\Delta x = 1$. Suppose define $\Delta t = 1/N$ for large natural number N, and observe the new process at times $\Delta t, 2\Delta t, 3\Delta t, ...$ Then with space increment being Δx, at time $1 = N\Delta t$, the value of the process is

$$W_1^{(N)} = \Delta x \Sigma_{i=0}^{N} X_i$$

In order to preserve the fluctuation/variance of the process to be 1, then

$$Var[\Delta x \Sigma_{i=0}^{N} X_i] = (\Delta x)^2 \Sigma_{i=0}^{N} Var[X_i]$$

consequentially, $\Delta x = \sqrt{\Delta t}$. Note by the central limit theorem

$$\frac{\Sigma_{i=0}^{N} X_i}{\sqrt{N}}$$

is approximately the standard normal distribution.

As we increase N, one can see that the process shifts from discrete to continuous space. The resulting process (the limit of random walk) is called Brownian motion or Wiener Process.

2.4 Brownian Motion

First let's introduce a few definition and theorems.

Definition 2.1. *Stochastic Process Let* $B_t = B(T)$ *be the value at a time* T*. For each* t*,* B_t *is a random variable. A collection of random variable indexed by time is called a stochastic process.*

There are three major assumptions about the random variable B_t

- **Stationary Increments.** If $s < t$, then the distribution of $B_t - B_s$ is the same as $B_{t-s} - B_0$
- **Independent Increments.** If $s < t$, then the random variable $B_t - B_s$ is independent of any value B_r for any $r < s$
- **Continuous Path.** The function that maps $t \to B_t$ is a continuous function of t.

Lemma 2.1. *Borel-Cantelli lemma Let* $\{E_i\}$ *be a sequence of events in some probability space* Ω*, then if*

$$\Sigma_{i=1}^{\infty} P(E_i) < \infty$$

then,

$$P(\limsup_{i\to\infty} E_i) = 0$$

The lim sup *denotes the limit supremum of the sequence of events, that is the set of outcomes that occur infinitely many times within the infinite sequence. Explicitly,*

$$\limsup_{i\to\infty} \bigcap_{i=1}^{\infty} \bigcup_{k\geq i}^{\infty} E_i$$

Proof. First note,

$$P(\Sigma_i^\infty \mathbb{1}[E_i] < \infty) = 1 \implies P(\limsup_{i\to\infty} E_i) = 0$$

Then

$$E[\Sigma_i^\infty \mathbb{1}[E_i]] = \Sigma_i^\infty E[\mathbb{1}[E_i]] = \Sigma_i^\infty P(E_i) < \infty$$

Then

$$P(\Sigma_i^\infty \mathbb{1}[E_i] < \infty) = 1$$

If not, then

$$E[\Sigma_i^\infty \mathbb{1}[E_i]] \geq \int_{\Sigma_i^\infty \mathbb{1}[E_i]=\infty} (\Sigma_i^\infty \mathbb{1}[E_i]) dP = \infty$$

□

Proposition 2.1. *Basic properties of Brownian motion For* $s < t$

- $E[B_t] = E[B_s] + E[B_{t-s}]$
- $Var[B_t] = Var[B_s] + Var[B_{t-s}]$

Definition 2.2. *Brownian Motion A stochastic process* B_t *or* $B(t)$ *is called Brownian Motion with drift* m*, variance* σ^2 *starting at the origin if it satisfies:*

- $B_0 = 0$.
- *For* $s < t$*, the distribution of* $B_t - B_s$ *is follows* $\mathcal{N}(m(t-s), \sigma^2(t-s))$.
- *If* $s < t$*, the random variable* $B_t - B_s$ *is independent of any* B_r *for* $r < s$.
- *With probability one, the function* $t \to B_t$ *is a continuous function of* t.

Proposition 2.2. *Scaling properties Suppose* B_t *is a standard Brownian motion (drift 0 and variance 1) and* $a > 0$*. Then* $Y_t = \frac{B_{at}}{\sqrt{a}}$ *is also a standard Brownian motion*

Proof. The properties of Brownian motion are still satisfied and is easy to see.
Expectation of Y_t is still 0 and the variance:

$$Var[Y_t] = Var[B_{at}/\sqrt{a}] = \frac{Var[B_{at}]}{a} = \frac{at}{a} = t$$

□

2.5 Existence of Brownian Motion

This section is a bit hard for me to understand fully. The flow of the proof is as follows:

- Proof Brownian Motion exists on discrete time.
- Proof Brownian Motion exists on countable infinite time.
- Proof Brownian Motion exists on a countable infinite time that is dense in real numbers
- Proof Brownian Motion exists on all real numbers

2.6 Understanding Brownian Motion

This section studies Brownian motion in depth, focuses on its differential, Hölder continuity, and Brownian motion as a martingale, Markov process, Gaussian process and self-similar process.

Theorem 2.1. *For any t, with probability one, the function $t \to B_t$ is not differentiable.*

Proof. First note, for any ϵ

$$B_{t+\epsilon} = B_t + \sqrt{\epsilon}N$$

where N is a standard normal random variable. Then

$$\lim_{\epsilon \to 0} \frac{B_{t+\epsilon} - B_t}{\epsilon} = \lim_{\epsilon \to 0} \frac{\sqrt{\epsilon}N}{\epsilon}$$

Therefore, with probability one, said limit goes to infinity as ϵ goes to 0. Hence the function is not differentiable for any t. □

In fact, a stronger statement is also true:

Theorem 2.2. *With probability one, the function $t \to B_t$ is nowhere differentiable.*

The logic of the proof is as follows:

- Assume the function is differentiable at some point t which falls in one of the 2^n intervals.
- Observe the behavior of the three intervals near t if the function is differentiable at t.
- Show that the probability of the intervals behaving that way has probability 0.
- Sum over all possible intervals that t can fall in, and show no matter where t is, the probability is still 0.
- Sum over all possible n, and show the sum of probability is finite.
- By Borel-Cantelli Lemma, the probability of the function being differentiable at some point has probability 0.

Proof. It is enough to show the function is not differentiable in [0,1].
Suppose B_t is differentiable at some point $t \in [0, 1]$, then its local rate of change is bounded by some finite constant M,

$$\sup_{\epsilon \in [0,1]} \frac{|B(t+\epsilon) - B(t)|}{\epsilon} < M$$

Fix M. Let $t \in [(k-1)/2^n, k/2^n]$ for some large n and $k \in [0, 2^n]$. If $B(t)$ is differentiable at t, then for all $j \in [1, 2^n - k]$:

$$\begin{aligned} &|B((k+j)/2^n) - B((k+j-1)/2^n)| \\ &\quad \le |B((k+j)/2^n) - B(t)| + |B(t) - B((k+j-1)/2^n)| \\ &\quad \le M(j/2^n) + M((j+1)/2^n) \\ &\quad = M(2j+1)/2^n \end{aligned}$$

Define a set of events:

$$A_{n,k} = \{|B((k+j)/2^n) - B((k+j-1)/2^n)| < M(2j+1)/2^n \text{ for j} = 1,2,3\}$$

Note $P(B_t \text{ is differentiable at t}) \leq P(A_{n,k})$. Hence,

$$\begin{aligned}
P(A_{n,k}) &\leq \Pi_{j=1}^{3} P(|B((k+j)/2^n) - B((k+j-1)/2^n)| < M(2j+1)/2^n) \\
&= \Pi_{j=1}^{3} P(|B(1/2^n)| < M(2j+1)/2^n) \\
&= \Pi_{j=1}^{3} P(\frac{|B(1)|}{\sqrt{2^n}} < M(2j+1)/2^n) \\
&= \Pi_{j=1}^{3} P(|B(1)| < M(2j+1)/\sqrt{2^n}) \\
&\leq P(|B(1)| < 7M/\sqrt{2^n})^3 \\
&\leq (7M/\sqrt{2^n})^3
\end{aligned}$$

The last inequality holds as standard normal is bounded by 0.5. As n increases, the probability goes to 0. Now summing over all possible intervals:

$$\begin{aligned}
P(B_t \text{ is differentiable somewhere}) &\leq P(\bigcup_{k=1}^{2^n} A_{n,k}) \\
&\leq 2^n (7M/\sqrt{2^n})^3 \\
&= (7M)^3/\sqrt{2^n}
\end{aligned}$$

The sequence goes to 0 drastically as we increase n, therefore the summation $\Sigma_n^{\infty} P(\bigcup_{k=1}^{2^n} A_{n,k})$ will be finite. Then by the Borel-Cantelli Lemma,

$$P(B_t \text{ is differentiable somewhere}) < P(\limsup_{n\to\infty} \bigcup_{k=1}^{2^n} A_{n,k}) = 0$$

which is the probability of function $B(t)$ having a point that is differentiable. □

2.6.1 Brownian Motion as Martingale

The martingale property is the consistency of expected value with respect to a filtration $\{\mathcal{F}_n\}$. i.e. for $s < t$

$$E[M_t|\mathcal{F}_s] = M_s$$

For a Brownian Motion B_t, let $\{\mathcal{F}_n\}$ be the martingale that B_t adapted to, then

$$\begin{aligned}
E[B_t|\mathcal{F}_s] &= E[B_s|\mathcal{F}_s] + E[B_t - B_s|\mathcal{F}_s]] \\
&= B_s + E[B_t - B_s] \\
&= B_s
\end{aligned}$$

To rigorously state a Brownian motion adapts some filtration $\{\mathcal{F}_s\}$, we often change the second condition for Brownian motion to:

If $s < t$, the random variable $B_t - B_s$ is independent of $\mathcal{F}_s$

The idea is even if we have more information at time s, they won't help us predicting the future increments. It is also worth noting that not all martingales that are defined on continuous time are continuous i.e. $f : t \to M_t$ is continuous. The most common example will be a Poisson Process which is a kind of *jumping process* and will be discussed later.

2.6.2 Brownian Motion as a Markov Process

The Markov Process is about the memoryless of the random variable. i.e. let $s \geq t$ and let $A = \{X_i\}_0^t$

$$P(X_s < C|A) = P(X_s < C|X_t)$$

Brownian Motion satisfies such property as:

$$Y_s = B_{t+s} - B_t$$

is independent from $\{\mathcal{F}_t\}$.

2.6.3 Brownian Motion as a Gaussian Process

A process $\{X_t\}$ is called a *Gaussian Process* if each subset of sequence of random variables

$$(X_i, ...X_{i+n})$$

has a joint normal distribution which is defined by its mean and covariance matrix. Let B_t be a standard Brownian Motion, and $t_i < t_{i+1} < ... < t_{i+n}$, then the corresponding $(B_i, ...B_{i+n})$ can be expressed as a linear combinations of independent standard normal random variables:

$$Z_j = \frac{B_{t_j} - B_{t_{j-1}}}{\sqrt{t_j - t_{j-1}}}$$

for $j \in 1, .., n$. Then B_t is a Gaussian Process with mean zero, and if $s < t$

$$\begin{aligned} Cov(B_s, B_t) = E[B_s B_t] &= E[B_s(B_s + B_t - B_t)] \\ &= E[B_s^2] + E[B_s(B_t - B_s)] \\ &= s + 0 = s \end{aligned}$$

which gives us $Cov(B_s, B_t) = mins, t$

2.6.4 Brownian Motion as a self-similar process

The idea of self-similar process comes from the fact that if one were to (properly) scale up a small portion of Brownian Motion, then the small piece looks like another ordinary Brownian Motion.

Theorem 2.3. *Suppose B_t is a standard Brownian Motion and $a > 0$. Let $Y_t = \frac{B_{at}}{\sqrt{a}}$,*

then, Y_t is a standard Brownian Motion. The variance is preserved by the scaling factor $a^{1/2}$

2.7 Computations for Brownian Motion

This section introduces a few quantities about Brownian Motion and their calculations; as well as the Reflection Principle.

①

$$
\begin{aligned}
E[|B_t|] = E[t^{1/2}|B_1|] &= t^{1/2}(2\pi)^{-1/2}\int_{-\infty}^{\infty}|x|e^{-0.5x^2}dx \\
&= t^{1/2}(2\pi)^{-1/2} * 2\int_{0}^{\infty} xe^{-0.5x^2}dx \\
&= \sqrt{2t\pi^{-1}}
\end{aligned}
$$

The last equality uses the property of half-normal distribution

A random variable is said to follow half-normal distribution if its PDF takes form

$$
\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{8}
$$

for $x > 0$. It has expected value of $\frac{\sigma\sqrt{2}}{\sqrt{\pi}}$. In this case, the last integral is almost the expected value of a half normal random variable with $\sigma = 1$, hence the integral takes value 1

②

$$
\begin{aligned}
P(B_t \geq r) = P(\sqrt{t}B_1 \geq 1) &= P(B_1 \geq rt^{-0.5}) \\
&= 1 - \Phi(rt^{-0.5})
\end{aligned}
$$

Φ represents the distribution function of standard normal, and the last equality uses the fact that B_1 follows standard normal

③ For any $t > s$,

$$
\begin{aligned}
P(B_t > 0, B_s > 0) &= \int_0^{\infty} P(B_t > 0|B_s = x)P(B_s = x)dx \\
&= \int_0^{\infty} P(B_t - B_s > -x)\frac{1}{\sqrt{2(s)\pi}}e^{-\frac{x^2}{2(s))}}dx \\
&= \int_0^{\infty}\int_{-x}^{\infty}\frac{1}{\sqrt{2(t-s)\pi}}e^{-\frac{y^2}{2(t-s))}}\frac{1}{\sqrt{2(s)\pi}}e^{-\frac{x^2}{2(s))}}dydx
\end{aligned}
$$

In the case of $t = 2, s = 1$, one can use polar coordinates to compute the result to be $\frac{3}{8}$. Immediately, we have

$$
P(B_2 > 0|B_1 > 0) = \frac{3}{4}
$$

Theorem 2.4. *Strong Markov Property If T is a stopping time with $p(T < \infty) = 1$ and let*

$$
Y_t = B_{T+t} - B_T,
$$

then Y_t is a standard Brownian Motion. Also, Y is independent of

$$
\{B_t : 0 \leq t \leq T\}
$$

We will use this property to prove the famous Reflection Principle

Theorem 2.5. *Let B_t be a standard Brownian motion starting at the origin, then for any $a > 0$,*

$$P\left(\max_{0\le s\le t} B_s \ge a\right) = 2P(B_t > a) = 2[1 - \Phi(a/\sqrt{t})]$$

The intuition behind the proof is that: In order for the motion to be greater than a at time t, the motion needs to first reach a some time before t (no matter where) which is equivalent to say the max value of the motion before time t is greater or equal to a. Then after it touches t, it has a 50% to not drop below a. So the probability is twice as much.

Proof. First define $T_a = min\{s : B_s \ge a\} = min\{s : B_s = a\}$. Note T_a qualifies as a stopping time. Then,

$$P\left(\max_{0\le s\le t} B_s \ge a\right) = P(T_a \le t) = P(T_a < t)$$

The the second inequality holds automatically due to continuity of Brownian Motion. Now,

$$\begin{aligned} P(B_t > a) &= P(T_a < t, B_t > a) \\ &= P(T_a < t)P(B_t - B_{T_a} > 0|T_a < t) \end{aligned}$$

Since it is given that $t > T_a$, we can use the Strong Markov Property:

$$P(B_t - B_{T_a} > 0|T < t) = 1/2$$

due to independence. The numerical value is immediate. □

We will introduce one example as an application of the Reflection Principle. Let

$$q(r,t) = P(B_s = 0 : r \le s \le t).$$

The scaling property of Brownian Motion shows that $q(r,t)$ can be scaled to $q(1,t/r)$, which is equivalent to $q(1, 1+s)$ for some $s \in \mathbb{R}$. Then we redefine $q(s) = q(1, 1+s)$, and let A be the event that B_t touch 0 in $(1, 1+s)$:

$$q(s) = \int_{-\infty}^{\infty} P(A|B_1 = x)P(B_1 = x)dx$$

Note:

$$\begin{aligned} P(A|B_1 = x) &= P\left(\min_{1\le k\le 1+s} B_k \le 0|B_1 = t\right) \\ &= P\left(\max_{0\le k\le s} B_k \ge x\right) \\ &= 2P(B_s \ge x) \\ &= 2[1 - \Phi(x/\sqrt{s})] \end{aligned}$$

Then the integral becoms:

$$\int_{-\infty}^{\infty} 2[1 - \Phi(x/\sqrt{s})]P(B_1 = x)dx$$

Once again, using polar coordinates, we have:

$$q(s) = 1 - \frac{2}{\pi}\arctan\frac{1}{\sqrt{s}}$$

2.8 Quadratic Variation

This section studies the sum of the squares small increment changes in time. i.e

$$Q_n = \sum_{i=1}^{n} \left[B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right) \right]^2$$

Note we can rewrite Q_n as

$$\frac{1}{n} = \sum_{i=1}^{n} Y_i$$

where

$$Y_i = Y_{i,n} = \left[\frac{B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right)}{1/\sqrt{n}} \right]^2$$

which follows chi-square distribution. Consequentially,

$$E[Y_i] = E[Z^2] = 1, E[Y_i^2] = E[Z^4] = 3$$

Then we have:

$$E[Q_n] = \frac{1}{n} \sum_{i=1}^{n} E[Y_i] = 1, Var[Q_n] \frac{1}{n^2} \sum_{i=1}^{n} Var[Y_i] = \frac{2}{n}$$

As $n \to \infty$, the variance goes to 0 and the random variable goes to a constant random variable. Extending the concept:

$$Q_n(t) = \sum_{i \leq tn} \left[B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right) \right]^2$$

As $n \to \infty$, the random variable approaches a constant random variable with value t. The quadratic variation is the limit of the above expression.

Definition 2.3. *Let B_t be a process, the quadratic variation is*

$$\langle B \rangle_t = \lim_{n \to \infty} \sum_{j \leq tn} \left[B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right) \right]^2$$

As computed above, we see that $\langle B \rangle_t = t$.

Now let $W_t = \sigma B_t + mt$, then $\langle W \rangle_t$ is equal to

$$\sum_{i \leq tn} \left[B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right) \right]^2 + \frac{2\sigma m}{n} \sum_{i \leq tn} \left[B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right) \right] + \sum_{i \leq tn} \frac{m^2}{n^2}$$

Simplify we have,

$$\sigma^2 \langle B \rangle_t + \frac{2\sigma m}{n} B_t + \frac{tm^2}{n}$$

As n approaches infinity, it is just $\sigma^2 t$.

Theorem 2.6. *If W_t is a Brownian Motion with drift m and variance σ^2, then $\langle W \rangle_t = \sigma^2 t$*

Above computations have been based on 'nice' partitions, now we observe the behavior of the quadratic variation when the partitions are not as ordered.

For a partition $\Pi = \{t_i\}$, $0 = t_0 < t_1 < ... < t_n = t$, we define

$$||\Pi|| = \max_{1 \leq i \leq n} t_i - t_{i-1};$$

and the corresponding quadratic variation

$$Q(t;\Pi) = \sum_{i=1}^{n} [B(t_i) - B(t_{i-1})]^2$$

Recall each increment between B_{t_i} and $B_{t_{i-1}}$ follows $\mathcal{N}(0, t_i - t_{i-1})$

$$\begin{aligned} E[Q(t;\Pi)] &= \sum_{i=1}^{n} E[(B(t_i) - B(t_{i-1}))^2] \\ &= \sum_{i=1}^{n} t_i - t_{i-1} = t \end{aligned}$$

$$\begin{aligned} Var[Q(t;\Pi)] &= \sum_{i=1}^{n} Var[(B(t_i) - B(t_{i-1}))^2] \\ &= \sum_{i=1}^{n} (t_i - t_{i-1})^2 \\ &\leq 2||\Pi|| \sum_{i=1}^{n} (t_i - t_{i-1}) = 2||\Pi||t \end{aligned}$$

Theorem 2.7. *Suppose B_t is a standard Brownian Motion with $t > 0$ and Π_n is a sequence of partitions of the form*

$$0 = t_{0,n} < t_{1,n} < ... < t_{l_n,n} = t,$$

with $||\Pi_n|| \to 0$. Then $Q(t;\Pi_n) \to t$ in probability. Moreover, if

$$\sum_{n=1}^{\infty} ||\Pi_n|| < \infty \tag{9}$$

then with probability one $Q(t;\Pi_n) \to t$

Proof. Using Chebyshev's inequality, for any integer k:

$$P\left(|Q(t;\Pi_n) - t| > \frac{1}{k}\right) \leq \frac{Var[Q(t;\Pi_n)]}{(1/k)^2} \leq 2k^2||\Pi_n||t$$

As $n \to \infty$, the right hand side goes to 0, which gives convergence in probability. If (8) holds, then

$$\sum_{n=1}^{\infty} P\left(|Q(t;\Pi_n) - t| > \frac{1}{k}\right) \leq 2k^2 t \sum ||\Pi_n|| < \infty$$

By the Borel-Cantelli lemma, with probability one, for large enough n, we have

$$|Q(t;||\Pi_n||) - t| \leq \frac{1}{k}.$$

□

3 Stochastic Integration

3.1 Introduction

This section provides a very general idea and motivation for stochastic calculus. Normally, as we learned in Riemann integrals, we have ODE (ordinary differential equation) as

$$df(t) = C(t, f(t))dt$$

or

$$\frac{df}{dt} = f'(t) = C(t, f(t))$$

where function C represents the differentiation operator. For SDE(stochastic differential equation), we have

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB_t$$

where m and σ are the drift and variance of a process X_t. The first part of the right hand side is a normal ODE with respect to time with a random integrand $m(s, X_s)$; the second part is the tricky one. To solve it we will use *Itô integral*

3.2 Stochastic Integral

In this section we will introduce stochastic as follows:

1. Discuss the integration with respect to simple processes
2. Extend the idea to bounded continuous paths by using limit and sum
3. Extend the idea to all continuous paths with the help of stopping time

3.2.1 Integration on Simple Process

To start off let's think Z_t, which is defined as

$$Z_t = \int_0^t A_s dB_s$$

to be a Brownian motion that have variance A_s^2 at time s, which changes as time goes on. First, let A_t be a simple process, which means it has constant value over pre-defined intervals. (Similar to step function in Riemann integrals). Formally,

Definition 3.1. *Simple Process A process A_t is a simple process if there exist times*

$$0 = t_0 < t_1 < ... < t_n < \infty$$

and random variables Y_j $j = 0, 1, .., n$ that are $\mathcal{F}_{t_j}$-measurable such that

$$A_t = Y_j, \text{ for } t_j \le t < t_{j+1}$$

Now define

$$Z_t = \int_0^t A_s dB_s = \sum_{i=0}^{j-1} Y_i[B_{t_{i+1}} - B_{t_i}] + Y_j[B_t - B_{t_j}]$$

for $t_j \le t \le t_{j+1}$. Note

$$\int_r^t A_s dB_s = Z_t - Z_r$$

Proposition 3.1. *Let B_t be a standard Brownian Motion with respect to filtration $\{\mathcal{F}_t\}$, and A_t, C_t be simple processes.*

- ***Linearity*** *If a,b are constants, then* $aA_t + bC_t$ *is also a simple process and*

$$\int_0^t (aA_s + bC_s)\, dB_s = a\int_0^t A_s dB_s + b\int_0^t C_s dB_s$$

If 0<r<t,

$$\int_0^t A_s ds = \int_0^r A_s dB_s + \int_r^t A_s dB_s$$

- ***Martingale Property*** *The process*

$$Z_t = \int_0^t A_s dB_s$$

is a martingale with respect to $\{\mathcal{F}_t\}$

- ***Variance rule*** Z_t *is square integrable and*

$$\mathrm{Var}\left[Z_t\right] = \mathbb{E}\left[Z_t^2\right] = \int_0^t \mathbb{E}\left[A_s^2\right] ds$$

- ***Continuity*** *With probability one, the function* $t \to Z_t$ *is a continuous function.*

Proof. Linearity and continuity are immediate (from definition and the fact that Brownian motions are continuous).

- **Martingale Property** We need to show that $E\left(Z_t|\mathcal{F}_s\right) = Z_s$, $\forall s < t$. This proof will show the case when $r = t_j, s = t_k$ for some $j > k$, the other cases are similar. (Just add a few terms here and there) By definition

$$Z_s = \sum_{i=0}^{k-1} Y_i \left[B_{t_{i+1}} - B_{t_i}\right]$$

$$Z_r = Z_s + \sum_{i=k}^{j-1} Y_i \left[B_{t_{i+1}} - B_{t_i}\right]$$

We know that $E(Z_s|\mathcal{F}_s) = Z_s$, then

$$E\left(Z_r|\mathcal{F}_s\right) = Z_s + \sum_{i=k}^{j-1} E\left[Y_i\left[B_{t_{i+1}} - B_{t_i}\right]|\mathcal{F}_s\right]$$

For $i \in \{k, j-1\}$, we have $t_i \geq s$, since $\mathcal{F}_s \subset \mathcal{F}_{t_i}$then

$$E\left[Y_i\left[B_{t_{i+1}} - B_{t_i}\right]|\mathcal{F}_s\right] = E\left[E\left(Y_i\left[B_{t_{i+1}} - B_{t_i}\right]|\mathcal{F}_{t_i}\right)|\mathcal{F}_s\right]$$

Since Y_i is $\mathcal{F}_{t_i}$-measurable and $B_{t_{i+1}} - B_{t_i}$ is independent of $\mathcal{F}_{t_i}$, we have

$$E\left(Y_i\left[B_{t_{i+1}} - B_{t_i}\right]|\mathcal{F}_{t_i}\right) = Y_i E\left(B_{t_{i+1}} - B_{t_i}|\mathcal{F}_{t_i}\right) = Y_i \mathbb{E}\left[B_{t_{i+1}} - B_{t_i}\right] = 0$$

If t, s are not chosen to be one of the increments, i.e.

$$r \neq t_j \text{ and } s \neq t_k \quad \forall j, k$$

Then then $\exists j, k$ such that

$$k = \min\{i|t_i > s\} \quad j = \max\{i|t_i < r\}$$

Hence we have

$$Z_s = \sum_{i=0}^{k-1} Y_i \left[B_{t_{i+1}} - B_{t_i}\right] + Y_{k+1}[B_s - B_{t_k}]$$

$$Z_r = Z_s + Y_{k+1}[B_{t_{k+1}} - B_s] + \sum_{i=k}^{j-1} Y_i \left[B_{t_{i+1}} - B_{t_i}\right] + Y_{j+1}[B_r - B_{t_j}]$$

The expected value of the additional terms are all 0, the above proof's logic still works

- **Variance Rule** We will show for $s = t_j$, the other case is similar as well (just add one extra term), then

$$Z_s^2 = \sum_{i=0}^{j-1}\sum_{k=0}^{j-1} Y_i \left[B_{t_{i+1}} - B_{t_i}\right] Y_k \left[B_{t_{k+1}} - B_{t_k}\right]$$

For any $i \neq k$ (assume $i < k$), then $(B_{t_{i+1}} - B_{t_i}), Y_i, Y_k$ are $\mathcal{F}_{t_k}$ measurable and $(B_{t_{k+1}} - B_{k_i})$ is not.

$$\begin{aligned}
&E\left[Y_i \left[B_{t_{i+1}} - B_{t_i}\right] Y_k \left[B_{t_{k+1}} - B_{t_k}\right]\right] \\
=&E\left[E\left(Y_i \left[B_{t_i+1} - B_{t_i}\right] Y_k \left[B_{t_{k+1}} - B_{t_k}\right] |\mathcal{F}_{t_k}\right)\right] \\
=&Y_i[B_{t_{i+1}} - B_{t_i}]Y_k E\left[B_{t_{k+1}} - B_{t_k}|\mathcal{F}_{t_k}\right] \\
=&0
\end{aligned}$$

Therefore any two term with different increment will have expectation 0. For the rest,

$$\begin{aligned}
E[Z_s^2] &= \sum_{i=0}^{j-1} E[Y_i^2(B_{t_{i+1}} - B_{t_i})^2] \\
&= \sum_{i=0}^{j-1} E[E(Y_i^2(B_{t_{i+1}} - B_{t_i})^2|\mathcal{F}_{t_i})] \\
&= \sum_{i=0}^{j-1} E[Y_i^2 E((B_{t_{i+1}} - B_{t_i})^2|\mathcal{F}_{t_i})] \\
&= \sum_{i=0}^{j-1} E[Y_i^2(t_{i+1} - t_i)] \\
&= \sum_{i=0}^{j-1}(t_{i+1} - t_i)E[Y_i^2]
\end{aligned}$$

By definition, A_s is a step function with values from Y_i, therefore we have

$$E\left[Z_s^2\right] = \sum_{i=0}^{j-1} E\left[Y_i^2\right](t_{i+1} - t_i) = \int_0^s E\left[A_r^2\right] dr$$

Similar to the **Martingale Property** proof, if s are not incremental points, then Z_s have an additional term, $Y_{j+1}[B_s - B_{t_j}]$. The interaction terms with these additional terms all have expectation 0, therefore the only additional terms remaining in $E[Z_s^2]$ is $(Y_{j+1}[B_s - B_{t_j}])^2$, which equals to $\int_{t_j}^s E[A_r^2]dr$

□

3.2.2 Integration on Continuous Processes

In this section we discuss the integration on continuous processes, A_t.

Lemma 3.1. *Suppose A_t is a process with continuous paths, adapted to the filtration $\{\mathcal{F}_t\}$. Suppose also that there exists $C < \infty$ such that with probability one $|A_t| \leq C$ for all t. then there exists a sequence of simple processes $A_t^{(n)}$ such that for all t,*

$$\lim_{n\to\infty} \int_0^t \mathbb{E}\left[\left|A_s - A_s^{(n)}\right|^2\right] ds = 0 \tag{10}$$

Moreover, for all n, t, $|A_t^{(n)}| \leq C$.

Proof. The proof is rather simple. We will show for $t = 1$. Define the sequence of simple processes as

,

$$A_t^{(n)} = A_{\frac{i}{n}} \quad \textit{where} \quad \frac{i}{n} \leq t < \frac{i+1}{n}.$$

One can easily see that $A_t^{(n)}$ converges point-wise to A_t, and is bounded by C as well. Therefore by the dominated convergence theorem,

$$\lim_{n\to\infty} \int_0^1 \left[A_t^{(n)} - A_t\right]^2 dt = 0$$

Since the integral is a bounded random variable, the expectation of the integral is also 0, i.e.

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^1 \left[A_t^{(n)} - A_t\right]^2 dt\right] = 0$$

□

Given this lemma, we can define integration on a bounded, continuous paths as a limit of integral on the simple paths that satisfies (9). i.e.

$$Z_t = \int_a^b A_s ds = \lim_{n\to\infty} \int_a^b A_s^{(n)} dB_s = \lim_{n\to\infty} \sum_{i=m}^{k-1} A_{\frac{i}{n}} \left[B_{\frac{i+1}{n}} - B_{\frac{i}{n}}\right]$$

where $\frac{k}{n} = b$, and $\frac{m}{n} = a$ Lemma 3.1 gives us the tool to approximate an integration with respect to a continuous Stochastic process. So for an **bounded** continuous process A_t, we can find sequence of simple processes $A_t^{(n)}$ described in lemma 3.1. Then for any given t, we can define

$$\int_0^t A_s dB_s = \lim_{n\to\infty} \int_0^t A_s^{(n)} dB_s. \tag{11}$$

We can call the integral, which is a random variable, Z_t. Immediately, Z_t presents four nice properties:

Proposition 3.2. *Let B_t be a standard Brownian Motion respect to filtration $\{\mathcal{F}_t\}$, A_t and C_t be bounded, adapted process with continuous paths, then*

- ***Linearity***. *If a, b are constants, then*

$$\int_0^t (aA_s + bC_s)\, dB_s = a \int_0^t A_s dB_s + b \int_0^t C_s dB_s$$

In addition, if r<t, then

$$\int_0^t (aA_s)\, dB_s = a \int_0^t A_s dB_s + b \int_r^t A_s dB_s$$

- ***Martingale property***. *The random variable/process*

$$Z_t \int_0^t A_s dB_s$$

is a martingale with respect to the filtration $\{\mathcal{F}_t\}$.

- ***Variance rule***. Z_t *is square integrable and*

$$\mathrm{Var}\left[Z_t\right] = \mathbb{E}\left[Z_t^2\right] = \int_0^t \mathbb{E}\left[A_s^2\right] ds$$

- ***Continuity***. *The function* $Z : t \to Z_t$ *is a continuous function with probability one.*

The above proposition and lemma deals with bounded continuous processes, for unbounded processes, we can approximate them using bounded ones incremented by natural numbers. Let A_t be a continuous process, not necessarily bounded, we define $T_n = min\{t : |A_n| = n\}$ (i.e. the first time A_t hits n, and redefine $A_t^n = A_{s \wedge T_n}$. Then each

$$Z_t^{(n)} = \int_0^t A_s^{(n)} dB_s$$

is well defined as A_t^n are bounded for every n. Then define

$$Z_t = \lim_{n \to \infty} Z_t^{(n)}$$

The implied assumption here is that A_t is not bounded when t can take all real value, but bounded when t is finite.

Under this construction, the Z_t will saticify linearity and continuity. If A_t is square-integralable, then Z_t will also saticify the variance rule, if A_t is not, then $\mathrm{Var}\left[Z_t\right] = \mathbb{E}\left[Z_t^2\right] = \int_0^t \mathbb{E}\left[A_s^2\right] ds = \infty$. The martingale property, which we will study more in depth later in the report, may not be satisfied.

Also, because we are dealing with a probability space here, the requirement of the paths can be relaxed to piece-wise continuous except a set of points of measure 0.

To incorporate stopping time into the integration (with respect to the same $\{\mathcal{F}_t\}$, we can add the stopping time restriction into the integrand.

Let A_t be a continuous process and we wish to integrate A_t from 0 to some stopping time T, then

$$Z_{t \wedge T} = \int_0^{t \wedge T} A_s dB_s = \int_0^t A_{s,T} dB_s$$

In other words, stopping the integral is equal to adjusting A_t to 0 after a certain time.

From our definition of stochastic integral, we can now attempt to define stochastic differential equations (SDE).

Let X_t be a process that satisfies

$$X_t = X_0 + \int_0^t A_s dB_s$$

where A_t is a continuous process. We can intemperate the equations as describing X_t to be a process that has a shift of X_0 and variation of A_t^2 then we can define dX_t to be

$$dX_t = \phi\left(X_t\right) dBt$$

where ϕ represents the differentiation operator. Our goal will be to derive the expression for said ϕ. However, note that stochastic calculus differs from the classic calculus in many ways. Take the most simple example, integrating a Brownian motion against a Brownian motin.

$$Z_t = \int_0^t B_s dB_s.$$

Note Z_t have finite expectation. The traditional approach would be to apply integrating rules and assume that

$$Z_t = \frac{1}{2}[B_t^2 - B_0^2] = \frac{B_t^2}{2}$$

However, from our previous calculations, we know that $E[Z_t] = 0$ and $E[B_t^2/2] = t/2$. The two values clearly do not agree, hence we need to explore for another method.

3.3 Itô's formula

Before we start to derive the Itô's formula, recall quadratic variation of a process.

$$\langle B \rangle_t = \lim_{n\to\infty} \sum_{j \le tn} \left[B\left(\frac{i}{n}\right) - B\left(\frac{i-1}{n}\right)\right]^2 = t \tag{12}$$

Now we extend the idea to any stochastic process, let Z_t be defined as

$$Z_t = \int_0^t A_s dB_s = t$$

Then

$$\begin{aligned}\langle Z \rangle_t &= \lim_{n\to\infty} \sum_{j \le tn} \left[Z\left(\frac{i}{n}\right) - Z\left(\frac{i-1}{n}\right)\right]^2 \\ &= \int_0^t A_s^2 ds\end{aligned}$$

Now suppose f is a C^1 function, then we may expand the function f using Taylor approximation

$$f(t+s) = f(t) + s * f'(t) + o(s)$$

where $o(s)$ approaches 0 as s^2 approaches 0. Itô's formula is derived using similar ideology.

Theorem 3.1. *(Itô's formula I). Suppose f is a C^2 function and B_t is a standard Brownian motion, then for every t,*

$$f(B_t) = f(B_s) + \int_s^t f'(B_s)\, dB_s + \frac{1}{2}\int_s^t f''(B_s)\, ds$$

which yields

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

We can interpolate this result as the process $X_t = f(B_t)$ at a certain time, t, behaves like a Brownian motion with drift $f''(B_t)/2$ and variance $f(B_t)^2$.

Proof. The logic of the proof goes as follows (roughly):

1. We will prove the case for $t = 1$ and $s = 0$, the general case is easily salable.

2. Separate the whole interval into finite sub-intervals, and label them using natural numbers.

3. Approximate each sub-interval using Taylor approximation.

4. Study each component of the approximation.

5. Let the number of sub-intervals go to infinity and observe the final limit.

6. The differential form can be attained by taking t infinitely close to s.

First separate the interval into n sub-intervals,

$$f(B_1) - f(B_0) = \sum_{i=1}^{n} \left[f\left(B_{j/n}\right) - f\left(B_{(j-1)/n}\right) \right]$$

Now expand the sub-intervals using Taylor approximation

$$\begin{aligned} f\left(B_{j/n}\right) - f\left(B_{(j-1)/n}\right) &= f'\left(B_{(j-1)/n}\right) * [B_{j/n} - B_{(j-1)/n}] \\ &+ \frac{1}{2} f''\left(B_{(j-1)/n}\right) * [B_{j/n} - B_{(j-1)/n}]^2 \\ &+ o\left([B_{j/n} - B_{(j-1)/n}]^2\right) \end{aligned}$$

Now let the number of sub-intervals go to infinity. Then we can see that difference between $f(B_1)$ and $f(B_0)$ contains three components:

$$\lim_{n\to\infty} \sum_{j=1}^{n} f'\left(B_{(j-1)/n}\right) \left[B_{j/n} - B_{(j-1)/n}\right] \tag{13}$$

$$\lim_{n\to\infty} \frac{1}{2} \sum_{j=1}^{n} f''\left(B_{(j-1)/n}\right) \left[B_{j/n} - B_{(j-1)/n}\right]^2 \tag{14}$$

$$\lim_{n\to\infty} \sum_{j=1}^{n} o\left(\left[B_{j/n} - B_{(j-1)/n}\right]^2\right) \tag{15}$$

By equation (14) we see that $[B_{j/n} - B_{(j-1)/n}] = 1/n$, which makes the last term goes to 0 quickly as we let n go to infinity. In addition, equation (12) is the approximation of the stochastic integration of $f'(B_t)$ through simple processes. Therefore equals to

$$\int_0^1 f'(B_t) dB_t$$

For the second term, equation (13), we can see that a part of the limit is the quadratic variation of B_t, if we can extract that part, then we can reduce the limit to something simple. Since we assumed f to be C^2, then we can define $h(t) = f''(B_t)$ and h is continuous. Hence we can find step functions to approximate h. i.e. for every ϵ given, we can find h_ϵ such that $|h(t) - h_\epsilon(t)| < \epsilon$, then we have

$$\left| \sum_{j=1}^{n} [h(t) - h_\epsilon(t)] \left[B_{j/n} - B_{(j-1)/n}\right]^2 \right| < \epsilon \sum_{j=1}^{n} \left[B_{j/n} - B_{(j-1)/n}\right]^2 \to \epsilon$$

Now take the limit of the term replacing h with h_ϵ

$$\lim_{n\to\infty} \sum_{j=1}^{n} h_\epsilon(t) \left[B_{j/n} - B_{(j-1)/n}\right]^2 = \int_0^1 h_\epsilon(t) dt$$

Note here h_ϵ is a function of real number t rather than a Brownian motion. Since ϵ is arbitrary, we have the following

$$\lim_{\epsilon \to 0} \frac{1}{2} \int_0^1 h_\epsilon(t) dt = \frac{1}{2} \int_0^1 h(t) dt = \frac{1}{2} \int_0^1 f''(B_t)\, dt$$

Hence,

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\, dB_s + \frac{1}{2} \int_0^t f''(B_s)\, ds$$

□

The proof given is somewhat a simplification of full the rigorous proof, difference being the application Taylor approximation to stochastic processes. As we know, Taylor approximation relies on approximating the value of a function by using the rate of change, or derivative, near that point. However, the intuition is slightly different when we replace the input variable with a Brownian motion.

We will outline the proof for the validity of Taylor approximation on Brownian motions

1. Choose a sequence of partition between (in this case) 0 and 1 with restrictions on their limiting sum of max norm being finite.
2. Proof for each partition in the sequence, given bounded second derivative, the upper and lower second order term of the approximation agrees as the partition becomes infinitely fine.
3. Conclude that we can approximate the Taylor approximation of f by simple processes base on the chosen partition.
4. Let the partition go to infinite.

3.4 More versions of Itô's formula

We have studied the integration of a process solely depending on time, next we look at a more general case involving the position (Brownian motion) as well.

Theorem 3.2. *(Itô's Formula II). Suppose $f(t,x)$ is a function that is C^1 in t and C^2 in x. If B_t is a standard Brownian motion, then*

$$f(t, B_t) = f(0, B_0) + \int_0^t \partial_x f(s, B_s)\, dB_s + \int_0^t \left[\partial_s f(s, B_s) + \frac{1}{2}\partial_{xx} f(s, B_s)\right] ds$$

Or equivalently

$$df(t, B_t) = \partial_x f(t, B_t)\, dB_t + \left[\partial_t f(t, B_t) + \frac{1}{2}\partial_{xx} f(t, B_t)\right] dt$$

The logic of deriving this formula is similar to the one dimensional case. The expansion of Taylor approximation around (t, x) is

$$f(t + \Delta t, B_t + \Delta B_t) - f(t, B_t) =$$
$$\partial_t f(t, B_t)\Delta t + o(\Delta t) + \partial_{B_t} f(t, B_t)\Delta x + \tfrac{1}{2}\partial_{B_t B_t} f(t, B_t)(\Delta B_t)^2 + o\left((\Delta B_t)^2\right)$$

The second order term for B_t survives because of quadratic variation, while the rest of the term quickly vanishes as the increment becomes smaller.

3.4.1 Geometric Brownian motion

Definition 3.2. *A process X_t is a geometric Brownian motion with drift m and volatility σ if it satisfies the SDE*

$$dX_t = mX_t dt + \sigma X_t dB_t = X_t \left[m dt + \sigma dB_t \right]$$

where B_t is the standard Brownian motion

Example 3.1. *Let $f(t,x) = e^{at+bx}$, where $a, b \in \mathbb{R}$, then*

$$\partial_t f(t,x) = af(t,x), \quad \partial_x f(t,x) = bf(t,x), \quad \partial_{xx} f(t,x) = b^2 f(t,x)$$

and we have

$$\begin{aligned} dX_t &= \left[\partial_t f\left(t, B_t\right) + \frac{1}{2} \partial_{xx} f\left(t, B_t\right) \right] dt + \partial_x f\left(t, B_t\right) dB_t \\ &= \left(a + \frac{b^2}{2} \right) X_t dt + bX_t dB_t \end{aligned}$$

The format the above example has is worth exploring.

Definition 3.3. *A process X_t is a geometric Brownian motion with drift m and volatility σ if it satisfies the SDE*

$$dX_t = mX_t dt + \sigma X_t dB_t = X_t \left[m dt + \sigma dB_t \right]$$

where B_t is the standard Brownian motion

The solution to geometric Brownian motions is

$$X_t = X_0 \exp \left\{ \left(m - \frac{\sigma^2}{2} \right) t + \sigma B_t \right\}$$

The intuition behind geometric Brownian motions is that the change between each increment is no longer normally distributed, but the change in percentage between each increment is. The above solution to the geometric Brownian motion SDE is called a 'strong' solution.

Now suppose that X_t satisfies the following form:

$$dX_t = R_t dt + A_t dB_t \tag{16}$$

or equivalently,

$$X_t = X_0 + \int_0^t R_s ds + \int_0^t A_s dB_s$$

Then the quadratic variation of X_t only depends on the drift term.

$$\begin{aligned}
\langle X\rangle_t &= \lim_{n\to\infty} \sum_{j<t_n} \left(X_{\frac{j}{n}} - X_{\frac{j-1}{n}}\right)^2 \\
&= \lim_{n\to\infty} \sum_{j<tn} \left(\int_0^{\frac{j}{n}} R_s ds + \int_0^{\frac{j}{n}} A_s dB_s - \int_0^{\frac{j-1}{n}} R_s ds - \int_0^{\frac{j-1}{n}} A_s dB_s\right)^2 \\
&= \lim_{n\to\infty} \sum_{j<tn} \left(\left[\sum_1^j R^m_{\frac{i}{n}} \cdot \frac{1}{n} - \sum_1^{j-1} R^m_{\frac{i}{n}} \cdot \frac{1}{n}\right] + \left[\sum_1^j A^m_{\frac{i}{n}} \Delta B_{t_i} - \sum_1^{j-1} A^m_{\frac{i}{n}} \Delta B_{t_i}\right]\right)^2 \\
&= \lim_{n\to\infty} \sum_{j<tn} \left(\left[R^m_{\frac{j}{n}} \cdot \frac{1}{n} + A^m_{\frac{j}{n}} \cdot \Delta B_{t_j})^2\right]\right)^2 \\
&= \lim_{n\to\infty} \sum_{j<tn} \left(\left[\frac{1}{m^2} \cdot \left(R^m_{\frac{j}{n}}\right)^2\right] + \left[\Delta B^2_{t_j} \cdot \left(A^m_{\frac{j}{n}}\right)^2\right] + 2 \cdot \frac{1}{n} \cdot \Delta B_{t_j} \cdot A^m_{\frac{j}{n}} \cdot R^m_{\frac{j}{n}}\right)
\end{aligned}$$

where A^m_t and R^m_t are simple processes used to approximate A_t and R_t. In the end, if we take m to go to infinity and n to go to infinity, the first and third term of the inner summation vanishes, and we are left with

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{j<tn} \left[\left(A^m_{\frac{j}{n}}\right)^2 \cdot \Delta B^2_{t_j}\right] \\
&= \sum_{j<tn} \left[\left(A_{\frac{j}{n}}\right)^2 \cdot dt\right] \\
&= \int_0^t A_s dt^2
\end{aligned}$$

If for another adapted process H_t, we can define the integration of H_t with respect to X_t as the following

$$\int_0^t H_s dX_s = \int_0^t H_s R_s ds + \int_0^t H_s A_s dB_s$$

To approximate the integral, one can simulate using the following discrete form

$$H_t \Delta X_t = H_t \left[X_{t+\Delta t} - X_t\right] = H_t \left[R_t \Delta t + A_t \sqrt{\Delta t} N\right]$$

Proceeding from this example, we have our final form of Itô's formula

Theorem 3.3. *Suppose X_t satisfies equation (15), and $f(t,x)$ is C^1 in t and C^2 in x, then*

$$\begin{aligned}
df\left(t, X_t\right) =& \partial_t f\left(t, X_t\right) dt + \partial_x f\left(t, X_t\right) dX_t + \frac{1}{2} \partial_{xx} f\left(t, X_t\right) d\langle X\rangle_t \\
=& \left[\partial_t f\left(t, X_t\right) + R_t \partial_x f\left(t, X_t\right) + \frac{A_t^2}{2} \partial_{xx} f\left(t, X_t\right)\right] dt \\
&+ A_t \partial_x f\left(t, X_t\right) dB_t
\end{aligned}$$

Notice the function involved in theorem is a map for time and a stochastic process (not a standard Brownian motion).

Example 3.2. *Let X_t be an SDE satisfying*

$$dX_t = A_t X_t dB_t, \quad X_0 = x_0$$

Then X_t is an exponential SDE, and the solution is:

$$X_t = x_0 \exp\left\{\int_0^t A_s dB_s - \frac{1}{2}\int_0^t A_s^2 ds\right\}$$

To verify, observe the exponential term first,

$$Y_t = \int_0^t A_s dB_s - \frac{1}{2}\int_0^t A_s^2 ds$$

By Itô's lemma, we have

$$dY_t = -\frac{A_t^2}{2}dt + A^t dB_t,$$

and $d\langle Y\rangle_t = A_t^2 dt$.

Then $X_t = x_0 exp(Y_t)$, and by chain rule, we obtain the following,

$$dX_t = X_t dY_t + \frac{1}{2}X_t d\langle Y_t\rangle = X_t(-\frac{A_t^2}{2}dt + A)tdB_t + \frac{1}{2}X_t d\langle Y_t\rangle = A_t X_t dB_t$$

The requirement regarding the smoothness of Itô's formula is sometime too strict. To allow ourselves to work with a partial smooth function, we have the following local form of the formula. This form just restricts a the time (and hence the location) in the desirable range before it escapes to bad behaved section of the function.

Theorem 3.4. *Suppose X_t satisfies (15) with $a < X_0 < b$, and $f(t,x)$ is C^1 in t and C^2 in $x \in (a,b)$. Then define $T = inf\{t : X_t = aorX_t = b\}$, then for t<T*

$$\begin{gathered} f\left(t, X_t\right) = f\left(0, X_0\right) + \int_0^t A_s \partial_x f\left(s, X_s\right) dB_s \\ + \int_0^t \left[\partial_s f\left(s, X_s\right) + R_s \partial_x f\left(s, X_s\right) + \frac{A_s^2}{2}\partial_{xx} f\left(s, X_s\right)\right] ds \end{gathered}$$

The theorem is a simple extension of the previous ones, and the proof is not as enlightening. The idea is to approximate T by restricting it to a slightly tighter interval, and then using the denseness of smooth functions to approximate the original function, and finally take the limit.

3.5 Diffusion

Definition 3.4. *A process is a diffusion process if it is a solution to an SDE of the form*

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB_t \tag{17}$$

where m, σ are functions. When the two functions does not depend on t, the solution is considered to be time-homogeneous. An example we have already encountered is the geometric Brownian motion.

Diffusion processes are Markov processes. To recap, the special property of Markov processes is the only valuable information needed to evaluate X_s is X_t for any $s > t$.

In this section, we will study the concept of generator of a Markov process.

Definition 3.5. *The generator $L = L_0$ of a Markov process X_t is*

$$Lf(x) = \lim_{t \to 0^+} \frac{E[f(X_t)] - f(x)}{t}$$

We may rewrite it as

$$E[f(X_t)] = t * Lf(x) + f(x)$$

Intuitively, the generator contains the information of the behavior of the process X_t on an infinitesimal small interval.

To study it more precisely, we will use Itô's formula to understand the generator of the diffusion X_t. For now, assume function m, σ are bounded smooth functions. Then by Itô's formula we have

$$\begin{aligned} df(X_t) =& f'(X_t)\,dX_t + \frac{1}{2}f''(X_t)\,d\langle X\rangle_t \\ =& \left[m(t, X_t)\,f'(X_t) + \frac{\sigma^2(t, X_t)}{2}f''(X_t)\right]dt \\ &+ f'(X_t)\,\sigma(t, X_t)\,dB_t \end{aligned}$$

in other words,

$$f(X_t) - f(X_0) = \int_0^t \left[m(s, X_s)\,f'(X_s) + \frac{\sigma^2(s, X_s)}{2}f''(X_s)\right]ds + \int_0^t f'(X_s)\,\sigma(s, X_s)\,dB_s$$

Now take expectation on both sides, realize the second term on the right side is a martingale as the intergand is bounded, and it can be expressed as a sum of Brownian motion increments. Then let tY_t be the first integral on the righthand side, we have

$$\frac{E[f(X_t) - f(X_0)]}{t} = E[Y_t]$$

Rewriting it slightly,

$$\lim_{t\to 0^+} Y_t = m(0, X_0)f'(X_0) + \frac{\sigma^2(0, X_0)}{2}f''(X_0) \tag{18}$$

Since the integrand is bounded, we can apply Monotone Convergence Theorem (in Lebesgue or Riemann) to take limit of expectation

$$Lf(x) = \lim_{t\to 0^+}\frac{E[f(X_t)] - f(x)}{t} = m(0, x)f'(x) + \frac{\sigma^2(0, x)}{2}f''(x)$$

We can extend the idea to other time intervals by replacing t and 0 with $t+s$ and t, given $X_t = x$ and we can obtain

$$m(t, x)f'(x) + \frac{\sigma^2(t, x)}{2}f''(x)$$

The above computations shows how the generator is computed for a diffusion process and how the generator can help generate the process on small intervals. So far, we have assumed the diffusion process does in fact have a solution. The proof is not as trivial as one might think.

Theorem 3.5. *Itô's existence and uniqueness theorem*

The outline of the proof is as follows:

1. Prove the general existence of a solution to a function with Lipschitz derivative.

   (a) Iteratively approximate the function by taking integral over small region.

(b) Show each step integral is bounded, and the sum of integral is bounded as well

(c) Conclude the existence of solution

2. Consider process X_t as a function of time

3. Validate the process still satisfies boundedness when taking step integrals/expectations

4. Conclude the convergence of approximate

5. Conclude the existence of solution

Consider equation

$$y'(t) = F(t, y(t)), y(0) = y_0 \tag{19}$$

We will assume F is uniform Lipschitz, meaning, there exists $L < \infty$ such that for all s, t, x, y

$$|F(s, x) - F(t, y)| \leq L|(s-t) + (x-y)| \tag{20}$$

We will now leverage the Picard iteration and construct a solution to (18) upto some point t_0, therefore all t below satisfies $t \leq t_0$. Start with the initial function

$$y_0(t) = y_0$$

then

$$y_k(t) = y_0 + \int_0^t F(s, y_{k-1}(s))ds$$

Let, $K = \max_{s\in[0,t_0]} |F(s, y_0)|$By construction we have

$$|y_k(t) - y_0(t)| \leq \int_0^t |F(s, y_0)|ds \leq K * t$$

For $k \geq 1$ we have

$$\begin{aligned} |y_{k+1}(t) - y_k(t)| &\leq \int_0^t |F(s, y_k(s)) - F(s, y_{k-1}(s))|\, ds \\ &\leq L \int_0^t |y_k(s) - y_{k-1}(s)|\, ds \end{aligned}$$

Applying induction we have

$$|y_{k+1}(t) - y_k(t)| \leq \frac{L^k C t^{k+1}}{(k+1)!} \tag{21}$$

Since each $y_k(t)$ is essentially small steps to approximate $y(t)$, the limit of $y_k(t)$ agrees with $y(t)$, and hence exists, then

$$|y_{k+1}(t) - y_k(t)| \leq K \sum_{i=k}^{\infty} \frac{L^i t^{i+1}}{(i+1)!}$$

Then consequence, $y_k(t)$ approaches

$$y_0 + \int_0^t F(s, y(s))ds$$

which satisfies the original ODE, it is easy to check $y(t)$ agrees with this expression. The involvement of t_0 is to eliminate cases where time is near 0 and x, y are close to y_0. We will not delve into the details on this topic.

Now lets relate what we just proved to the diffusion process, suppose m, σ both satisfy (18). For ease, choose $t_0 = 1$ and define the iteration for $t \in [0, 1]$

$$X_t^0 = y_0$$
$$X_t^{k+1} = y_0 + \int_0^t m(s, X_s^k)ds + \int_0^t \sigma(s, X_s^k)dB_s$$

Take expectation on both sides

$$\mathbb{E}\left[|X_t^{(k+1)} - X_t^k|^2\right] \leq 2\mathbb{E}\left[\left(\int_0^t L\left|X_s^{(k)} - X_s^{(k-1)}\right| ds\right)^2\right]$$
$$+ 2\mathbb{E}\left[\left(\int_0^t \left[\sigma\left(s, X_s^{(k)}\right) - \sigma\left(s, X_s^{(k-1)}\right)\right] dB_s\right)^2\right]$$

Applying Hölder inequality on the first integral we have

$$E\left[\left(\int_0^t L|X_s^k - X_s^{k-1}|ds\right)^2\right] \leq \mathbb{E}\left[L^2 t \int_0^t |X_s^k - X_s^{k-1}|^2 ds\right]$$
$$\leq L^2 \int_0^t \mathbb{E}\left[|X_s^k - X_s^{k-1}|^2\right] ds$$

The second integral can be bounded by applying the variance rule

$$E\left[\left(\int_0^t \left[\sigma\left(s, X_s^k\right) - \sigma\left(s, X_s^{k-1}\right)\right] dB_s\right)^2\right] = \int_0^t E\left(\left[\sigma\left(s, X_s^k\right) - \sigma\left(s, X_s^{k-1}\right)\right]^2\right) ds$$
$$\leq \beta^2 \int_0^t E\left[|X_s^k - X_s^{k-1}|^2\right] ds$$

Since $E[|X_t^k - X_t^{k+1}|]$ is bounded, then (20) suggests the existence of λ satisfying

$$\left|X_t^{k+1}(t) - X_t^k(t)\right| \leq \frac{\lambda^k t^{k+1}}{(k+1)!} \tag{22}$$

The above proof can be extended to all rational numbers due to countability, then leveraging the denseness of rationals, the result can be extended to t continuous cases. It is also worth noting the Lipschitz is stronger than we need. We can still bound or restrict our case on a locally Lipschitz region like we did for Theorem 3.4.

3.6 Covariation and the product rule

Let X_t, Y_t saticify

$$dX_t = H_t dt + A_t dB_t, dY_t = K_t dt + C_t dB_t$$

Then the covariation process is defined by

$$\langle X, Y\rangle_t = \lim_{n\to\infty} \sum_{j\leq tn} \left[X_{\frac{j}{n}}) - X_{\frac{j-1}{n}})\right] \left[Y_{\frac{j}{n}}) - Y_{\frac{j-1}{n}})\right]$$

Note X_t, Y_t are independent, then

$$[dX_t][dY_t] = [H_t dt + A_t dB_t][K_t dt + D_t dB_t]$$
$$= A_t C_t dt + O(dt^2) + O(dt dB_t)$$
$$= \int_0^t A_s C_s ds$$

Equivalently

$$d\langle X, Y\rangle_t = A_t C_t dt$$

If we look at the traditional product rule in calculus, for functions f, g

$$\begin{aligned} d(fg) &= f(x+dx)g(x+dx) - f(x)g(x) \\ &= [f(x+dx) - t]g(x+dx) + f(x)[g(x+dx) - g(x)] \\ &= (df)g + (dg)f + (df)(dg) \\ &= gf'dt + fg'dt + f'g'dt^2 \end{aligned}$$

and the last term vanishes in traditional calculus. In stochastic calculus, if we replace f, g with X_t, Y_t as functions of t, then the last term, $(dX_t)(dY_t)$, does not vanish, instead becomes $d\langle X, Y\rangle_t$. Combining all of the above we have

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + d\langle X, Y\rangle_t$$

in other words.

Theorem 3.6. *Let X_t, Y_t be defined as above, then*

$$\begin{aligned} X_t Y_t &= X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \int_0^t d\langle XY\rangle s \\ &= X_0 Y_0 + \int_0^t [X_s K_s + Y_s H_s + A_s C_s] ds + \int_0^t [X_s C_s + Y_s A_s] dB_s \end{aligned}$$

4 More on Stochastic Calculus

4.1 Martingales and local martingales

Recall a square integrable process $Z_t = \int_0^t A_t dt$ satisfies

$$\int_0^t E[A_s^2] ds < \infty$$

This section will introduce the optional sampling theorem by first showing an example where the the process is not square integrable and not a martingale. The example is a continuous extension of previously introduced martingale betting strategy where an individual bets twice of what was lost until victory. The game allowed the player to have infinite amount of money and was allowed to bet infinitely large amounts, the following example carries the same idea but in a limited time interval which forced the frequency of betting to increase dramatically.

Example 4.1. *Let Z_t be the outcome of a continuous betting strategy A_t, i.e.*

$$Z_t = \int_0^t A_s dB_s$$

where A_s takes constant value across $[t_i, t_{i+1}]$, and

$$t_i = 1 - 2^{-i}$$

the value of A_t is defined as the following.
Let $A_t = 1$ for $t \in [0, 1/2]$, then

$$P(Z_{1/2} > 1) = P(B_{1/2} > 1) > 0$$

The process stops if the outcome exceeds 1, which means $A_t = 0$ for $t \in [0.5, 1]$ if $B_{1/2} > 1$.
If $B_{1/2} < 1$, then define $A_t = \alpha$ for $t \in [0.5, 0.75]$ where α satisfies

$$P(\alpha[B_{3/4} - B_{1/2}] \geq 1 - Z_{1/2}) = P(Z_{1/2} > 1)$$

Realize α is $\mathcal{F}_P 1/2$ measurable. Then observe $Z_{3/4}$

$$Z_{3/4} = \int_0^{3/4} A_s dB_s = Z_{1/2} + \alpha[B_{3/4} - B_{1/2}] \geq 1$$

So by construction

$$P(Z_{3/4} > 1 | Z_{1/2 < 1}) = 1$$

Now repeat the process for $t_n = \sum_{i=1}^{n} 2^{-i}$. Then by simple induction one can check

$$P(Z_{t_n} > 1 | Z_{t_{n-1}} < 1) = q$$

and hence,

$$P(Z_{t_n} < 1) < (1 - q)^n$$

which goes to 0 as n tends to infinity. So $E[Z_t] \geq 1$ almost surely. Yet $Z_0 = 0$, so Z_t fails to be a martingale. By the denseness of piece-wise continuous processes, there exists a continuous processes that is equal to A_t (almost) everywhere with the same result.

Even though this betting strategy fails to be a martingale, if one were to restrict the allowed betting amount, it will become a martingale.

Definition 4.1. *Local Martingale A continuous process M_t adapted to the filtration $\{\mathcal{F}_t\}$ is a local martingale on $[0, T)$ if there exists an increasing sequence of stopping times $\{\tau_n\}$ such $\{\tau_n\} \to T$ almost surely, and $M_{t \wedge \tau_i}$ is a martingale for each i.*

For stochastic integrals, we can define $\{\tau_n\}$ to be

$$\tau_i = inf \left\{ t : \langle Z \rangle_t = \int_0^t A_s dB_s = i \right\}$$

Then $Z_{t \wedge \tau_i}$ is square integrable for each i and hence on $[0, T)$. T is consequently defined as

$$\tau_i = inf \left\{ t : \langle Z \rangle_t = \int_0^t A_s dB_s = \infty \right\}$$

Note for general Z_t satisfying

$$dZ_t = R_t dt + A_t dB_t$$

to be a martingale, R_t needs to be 0. However, as we just saw, stronger conditions are needed to guarantee martingale property.

Theorem 4.1. *Optional Sampling Theorem Suppose Z_t is a continuous martingale and T is a stopping time, with respect to same filtration $\{\mathcal{F}_t\}$. Then*

1. *If $Z_{t\wedge T}$ is a continuous martingale with respect to the filtration. Also, $E[Z_{t\wedge T}] = E[Z_0]$*
2. *If there exists $C < \infty$ such that for all t, $Z_{t\wedge T}^2 \leq C$. Then if $P[T < \infty] = 1$, $E[Z_T] = E[Z_0]$*

Proof. The proofs are analogous to the discrete version of optional sampling theorem stated earlier in the notes.
For 1, $\forall t \in \mathbb{N}$, WLOG, assume $t > T$

$$\begin{aligned} E[Z_{t\wedge T}] &= E[A_0 + \int_0^t A_s dB_s \\ &= A_0 + E[\int_1^t A_s dB_s] \\ &= E[Z_0] + E[\int_1^T A_s dB_s] + E[\int_T^t 0 dB_s] \\ &= E[Z_0] \end{aligned}$$

The second term in the second last equality has value 0, this can be calculated by approximating A_s upto finite time T using simple processes and take expectation of each increment.
The proof for 2 is nearly identical with the discrete case. We start by observing

$$\lim_{t\to\infty} E[|M_t| 1\{T > t\}]$$

$$\begin{aligned} E[|Z_t| 1\{T > t\}] &= \boldsymbol{E[|Z_t| 1\{|Z_t| \geq b, T > t\}]} + E[|Z_t| 1\{|Z_t| < b, T > t\}] \\ &\leq \boldsymbol{\frac{1}{b} E[|Z_t|^2 1\{|Z_t| \geq b, T > t\}]} + E[|Z_t| 1\{|Z_t| < b, T > t\}] \\ &\leq \boldsymbol{\frac{1}{b} (E[|Z_t|^2 1\{|Z_t| \geq b, T > t\}] + E[|Z_t|^2 1\{|Z_t| < b, T > t\}]} \\ &\quad \boldsymbol{+ E[|Z_T|^2 1\{T < t\}])} + E[|Z_t| 1\{|Z_t| < b, T > t\}] \\ &\leq \boldsymbol{\frac{1}{b} (E[|Z_t|^2 1\{T > t\}] + E[|Z_T|^2 1\{T < t\}])} \\ &\quad + E[|Z_t| 1\{|Z_t| < b, T > t\}] \\ &\leq \frac{E[|Z_{t\wedge T}|^2]}{b} + bP\{T > t\} \end{aligned}$$

Then we have

$$E[|Z_t| 1\{T > t\}] \leq \tfrac{C}{b} + bP\{T > t\}.$$

Observe for each n,

$$\begin{aligned} \limsup_{n\to\infty} E[|Z_t| 1\{T > t\}] &\leq \limsup_{n\to\infty} \frac{C}{b} + P\{T > t\} \\ \limsup_{n\to\infty} E[|Z_t| 1\{T > t\}] &\leq \frac{C}{b} + \lim_{n\to\infty} P\{T > t\} \\ \limsup_{n\to\infty} E[|Z_t| 1\{T > t\}] &\leq \frac{C}{b} \end{aligned}$$

and hence

$$0 \leq \lim_{t\to\infty} E[|Z_t|1\{T > n\}] \leq \limsup_{t\to\infty} E[|Z_t|1\{T > t\}] \leq \frac{C}{b}$$

holds for all b, take b to infinity we have

$$\lim_{t\to\infty} E[|M_t|1\{T > t\}] = 0,$$

Combining above we have

$$\begin{aligned} E[M_0] &= E[M_{t\wedge T}] \\ &= E[M_T] + E[M_{t\wedge T} - M_T] \\ &= E[M_T] + E[1\{T > t\}[M_t - M_T]] \\ &= E[M_T] + E[1\{T > t\}[M_t]] - E[1\{T > t\}M_T]] \end{aligned}$$

As n approaches infinity, the second term goes to 0 as $1\{T > t\}$ goes to 0, and the third term, as we just proved, goes to 0 as well. Hence we have the result. □

After obtaining the tools needed, we will be looking at a few examples of stochastic processes.

4.2 Bessel Process

Definition 4.2. *Bessel Process A Bessel process with parameter α is the solution to the SDE*

$$dX_t = \frac{\alpha}{X_t}dt + dB_t, X_0 = x_0 > 0$$

The process constantly swings away and back towards the axis due to the inverse drift parameter. The main motivation of this section is to observe how α influences the processes' behavior near the x-axis.

The deriving process using the following ideology,

1. Realize X_t has solution on any interval away from 0.
2. Observe the behavior of the process with in (r, R)
3. Reduce the problem to an ODE
4. Take limit of both r and R to get final result

First note the process starts above the axis, therefore for any ϵ, and $T_\epsilon = inf\{t : X_t \leq \epsilon\}$, the process is well defined. Moreover, the process is Lipschitz on the interval $[\epsilon, \infty)$, so by the Itô's existence and uniqueness theorem, the equation has a unique solution. Then only time we need to worry about the well-definiteness of the process is at time T, where

$$T = inf\{t : X_t = 0\}$$

Now suppose $0 < r < x < R < \infty$, and $\phi(x)$ be the probability of the process, starting at x, to reach R before r. Now consider

$$M_t = E[\chi_{\{X_\tau = R\}}|\mathcal{F}_t]$$

So M_t is the expectation of the process reaching R before r. Due to the Markov property of diffusion processes:

$$M_t = \phi(X_{t\wedge\tau})$$

and by the tower property, M_t satisfies a martingale:

$$E\left[M_t|\mathcal{F}_s\right] = E\left[E\left(J|\mathcal{F}_t\right)|\mathcal{F}_s\right] = E\left[J|\mathcal{F}_s\right] = M_s$$

Then by Itô's formula we have

$$\begin{aligned} d\phi\left(X_t\right) &= \phi'\left(X_t\right)dX_t + \frac{1}{2}\phi''\left(X_t\right)d\langle X\rangle_t \\ &= \left[\frac{\alpha\phi'\left(X_t\right)}{X_t} + \frac{\phi''\left(X_t\right)}{2}\right]dt + \phi'\left(X_t\right)dB_t \end{aligned}$$

Since we have showed the process is a martingale, then the dt term must vanish. So we have obtained the standard one dimensional differential equation

$$x\phi''(x) + 2\alpha\phi'(x) = 0$$

with solutions

$$\begin{aligned} \phi(x) &= c_1 + c_2 x^{1-2a}, \quad a \neq \frac{1}{2} \\ \phi(x) &= c_1 + c_2\log x, \quad a = \frac{1}{2} \end{aligned}$$

Applying boundary conditions of $\phi(r) = 0$, $\phi(R) = 1$

$$\begin{aligned} \phi(x) &= \frac{x^{1-2a} - r^{1-2a}}{R^{1-2a} - r^{1-2a}}, \quad a \neq \frac{1}{2} \\ \phi(x) &= \frac{\log x - \log r}{\log R - \log r}, \quad a = \frac{1}{2} \end{aligned}$$

Now fix any $R > x$, consider all $r \in (0, x)$ and observe $\phi(x)$. Recall $\phi(x)$ is the probability of x reaching R before r, then $P(X_\tau = r) = 1 - \phi(x)$. Now take r to be infinitely small

$$\lim_{r\to 0}\mathbb{P}\left\{X_\tau = r\right\} = \begin{cases} 0 & \text{if } a \geq 1/2 \\ 1 - (x/R)^{1-2a} & \text{if } a < 1/2 \end{cases}$$

This result is proposition 4.2.1 in the book.

4.3 Feynman-Kac Formula

This section we take a look at at popular model for evaluating an option price. Suppose the price of a stock follows a geometric Brownian motion

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB + t$$

An option is an arrangement between two parties that will be executed if the price of the stock at time T, X_T is above a certain threshold, S. Then the option's present value of the option is

$$F(X_T) = max\{X_T - S, 0\}$$

Normally, we assume there is a inflation rate r (typically the saving interest rate at the banks). Then the present value of the option will be $e^{rt}F(X_T)$, now we define the function for the expected value of the option value at a certain point in time,

$$\phi(t,x) = E[e^{-(T-t)}F(X_T)|X_t = x] \tag{23}$$

and the function for inflation/interest rate

$$dR_t = r(t, X_t)R_t dt$$

Note R_t denotes the value at time t_0.

$$R_t = R_0 exp\{\int_0^t r(s, X_s)ds\}$$

The Feynman-Kac formula provides some insight to this value. Putting this into ϕ,

$$\phi(t,x) = \mathbb{E}\left[\exp\left\{-\int_t^T r(s, X_s)\, ds\right\} F(X_T)\,|X_t = x\right]$$

We will assume ϕ is twice differentiable in x, and differntiable in t. Now let

$$M_t = E[R_T^{-1}F(X_T)|\mathcal{F}_t]$$

Note M_t is a martingale, take any $s < t$,

$$\begin{aligned}
E\left[M_t|\mathcal{F}_s\right] =& E\left[E\left(E[R_T^{-1}F(X_T)|\mathcal{F}_T]|\mathcal{F}_t\right)|\mathcal{F}_s\right] \\
=& E\left[E\left(M_T|\mathcal{F}_t\right)|\mathcal{F}_s\right] \\
=& E\left[M_T|\mathcal{F}_s\right] \\
=& E\left[R_T^{-1}F(X_T)|\mathcal{F}_s\right] \\
=& M_s
\end{aligned}$$

putting R_t into M_t we have

$$M_t = R_t^{-1}E\left[\exp\left\{-\int_t^T r(s, X_s)\, ds\right\} F(X_T)\,|\mathcal{F}_t\right]$$

Since X_t is a Markov process, we have the result

$$M_t = R_t^{-1}\phi(t, X)t$$

Then apply Itô's formula we have

$$\begin{aligned}
d\phi(t, X_t) &= \partial_t\phi(t, X_t)\, dt + \partial_x\phi(t, X_t)\, dX_t + \frac{1}{2}\partial_{xx}\phi(t, X_t)\, d\langle X\rangle_t \\
&= \left(\partial_t\phi(t, X_t) + m(t, X_t)\,\partial_x\phi(t, X_t) + \frac{1}{2}\sigma(t, X_t)^2\,\partial_{xx}\phi(t, X_t)\right) dt + \sigma(t, X_t)\partial_x\phi(t, X_t)
\end{aligned}$$

Now return to M_t, since $\langle R\rangle_t = 0$, we can apply the product rule,

$$d\left[R_t^{-1}\phi(t, X_t)\right] = R_t^{-1}d\phi(t, X_t) + \phi(t, X_t)\, d\left[R_t^{-1}\right]$$

Then we have the drift term to be

$$R_t^{-1}(-r\left(t,X_t\right)\phi\left(t,X_t\right)+\partial_t\phi\left(t,X_t\right)+m\left(t,X_t\right)\partial_x\phi\left(t,X_t\right)$$
$$+\frac{1}{2}\sigma\left(t,X_t\right)^2\partial_{xx}\phi\left(t,X_t\right))$$

As M_t is a martingale, the above must equal to 0, and hence we have,

$$-r\left(t,X_t\right)\phi\left(t,X_t\right)+\partial_t\phi\left(t,X_t\right)+m\left(t,X_t\right)\partial_x\phi\left(t,X_t\right)+\frac{1}{2}\sigma\left(t,X_t\right)^2\partial_{xx}\phi\left(t,X_t\right)=0$$

We have the following theorem.

Theorem 4.2. *Feynman-Kac Formula Suppose X_t is a geometric Brownian motion with drift $m(t,X_t)$, variance $\sigma(t,X_t)$, $r(t,x)\geq 0$ is a discounting rate. Then a payoff F_T with $E[|F(X_T)|]<\infty$ for an option with strike price S, if $\phi(t,x)$ for $t<T$ is C^1 in t, and C^2 in x, then $\phi(t,x)$ satisfies the PDE*

$$\phi_t(t,x)=-m(t,x)\partial_x\phi(t,x)-\frac{1}{2}\sigma(t,x)^2\partial_{xx}\phi(t,x)+r(t,x)\phi(t,x)$$

with terminal condition $\phi(T,x)=F(x)$

4.4 Binomial Approximations

So far, we have been approximating SDEs using the Euler method. This section wishes to introduce sampling methods where each $X(t+\Delta t)$ takes one of two values. Let X_t be a Brownian motion with zero drift and constant variance σ^2. Then binomial scheme is approximation by a random walk where

$$P(X_{t+\Delta t}-X_t=\pm\sigma\sqrt{\Delta t})=1/2$$

Then the value of $X_{k\Delta t}$ takes value in the lattice of points

$$\left\{....-\sigma\sqrt{\Delta t},0,\sigma\sqrt{\Delta t}....\right\}$$

Suppose the variance is constant, but the drift depends on time and location, then we have two methods to approximate. One is to use Euler's method

$$P(X_{t+\Delta t}-X_t=m(t,X_t)\Delta t\pm\sigma\sqrt{\Delta t})=1/2$$

The other method is to adjust the probability base on the drift,

$$\mathbb{P}\{X_{t+\Delta t}-X_t=\pm\sigma\sqrt{\Delta t}|X_t\}=\frac{1}{2}\left[1\pm\frac{m\left(t,X_t\right)}{\sigma}\sqrt{\Delta t}\right]$$

Note the expectation of difference between X_t and $X_{t+\Delta t}$ is still $m(t,X_t)\Delta t$, same as the first method.

Example 4.2. *We will use the second rule to simulate a Brownian motion with constant nonnegative drift and constant variance of 1.*

Suppose $\Delta t=1/N$ for large N, and we are interested in X_1. If we denote each upward/downward moment by $a_i=\pm1$, then the behavior of the motion is dictated by

$$\omega=(a_1,a_2...a_N)$$

Note if we let $J = J(\omega)$ be the number of $+1$'s and define $r = \frac{1}{\sqrt{N}}(J - (N/2))$, then

$$\begin{aligned} X_1 &= \sqrt{\Delta t}\,[a_1 + \cdots + a_N] \\ &= J\sqrt{\Delta t} - (N-J)\sqrt{\Delta t} \\ &= 2r\sqrt{N}\sqrt{\Delta t} = 2r \end{aligned}$$

And for each ω, the corresponding probability is

$$\begin{aligned} q(\omega) &= \left(\frac{1}{2}\right)^N [1 + m\sqrt{\Delta t}]^J [1 - m\sqrt{\Delta t}]^{N-J} \\ &= \left[1 + \frac{m}{\sqrt{N}}\right]^J \left[1 - \frac{m}{\sqrt{N}}\right]^{N-J} \\ &= \left[1 - \frac{m^2}{N}\right]^{N/2} \left[1 + \frac{m}{\sqrt{N}}\right]^{r\sqrt{N}} \left[1 - \frac{m}{\sqrt{N}}\right]^{-r\sqrt{N}} \end{aligned}$$

Using the approximation for e, we have

$$e^{-m^2/2}e^{2rm} = e^{mX_1}e^{-m^2/2}$$

This result essentially shows we can simulate this motion by a Brownian motion without drift and scale it proportionally. This leads to our final theorem.

Theorem 4.3. *Suppose*

$$dX_t = m(X_t)dt + \sigma dB_t$$

where m is continuously differntiable, let $p(t,x)$ denote the density of X_t, then

$$\partial_t p(t,x) = L_x^* p(t,x)$$

where

$$\begin{aligned} L^* f(x) =& [m(x)f(x)]' + \frac{\sigma^2}{2} f''(x) \\ =& - m'(x)f(x) - m(x)f'(x) + \frac{\sigma^2}{2} f''(x) \end{aligned}$$

Note if m is constant, it resort to the expression we saw for generators earlier. For non constant m, we will derive the expression by using the second binomial approximation,

$$\mathbb{P}\{X(t+\Delta t) - X(t) = \pm\sigma\sqrt{\Delta t} | X(t)\} = \frac{1}{2}\left[1 \pm \frac{m\left(X_t\right)}{\sigma}\sqrt{\Delta t}\right]$$

Then for the motion to be at position $x = k\sqrt{\Delta t}$ at time $t + \Delta t$, it must be at $x \pm \sigma x = k\sqrt{\Delta t}$ at time t, then

$$\begin{aligned} p\left(t+\epsilon^2, x\right) =& p(t, x-\sigma\epsilon)\frac{1}{2}\left[1 + \frac{m(x-\sigma\epsilon)}{\sigma}\epsilon\right] \\ &+ p(t, x+\sigma\epsilon)\frac{1}{2}\left[1 - \frac{m(x+\sigma\epsilon)}{\sigma}\epsilon\right] \end{aligned} \tag{24}$$

We also know

$$p\left(t + \Delta t, x\right) = p(t,x) + \partial_t p(t,x)\Delta t + o\left(\Delta t\right)$$

and

$$\begin{aligned}
p(t,x+\sigma\epsilon)+p(t,x-\sigma\epsilon) =& p(t,x)+\frac{\sigma^2\epsilon^2}{2}\partial_{xx}p(t,x)+o\left(\epsilon^2\right)\\
p(t,x\pm\sigma\epsilon) =& p(t,x)\pm\partial_x p(t,x)\sigma\epsilon+o(\epsilon)\\
m(x\pm\sigma\epsilon) =& m(x)\pm m'(x)\sigma\epsilon+o(\epsilon)
\end{aligned}$$

Plugging the above into (23), we have the result.

4.5 Continuous martingales

In this section, we will prove that Brownian motion is the only type of continuous martingale.

Proposition 4.1. *Suppose M_t is a continuous martingale with respect to a filtration $\{\mathcal{F}\}$ with $M_0 = 0$, and suppose that the quadratic variation of M_t is the same as that of standard Brownian motion,*

$$\langle M\rangle_t = \lim_{n\to\infty}\sum_{j<2n_t}\left[M\left(\frac{j+1}{2^n}\right)-M\left(\frac{j}{2^n}\right)\right]^2 = t$$

Then for all $\lambda \in \mathcal{R}$

$$E[exp\{i\lambda M_t\}] = e^{\lambda^2 t/2}$$

This proposition shows the form of the characteristic function of any continuous martingale is in the above form, hence the distribution is is normal. Recall the first term of the Itô Integral is a martingale, then,

$$f\left(M_t\right)-f\left(M_0\right)=N_t+\frac{1}{2}\int_0^t f''\left(M_s\right)ds = N_t-\frac{\lambda^2}{2}\int_0^t f\left(M_s\right)ds$$

where N_t is a martingale. Then for $r < t$ we have

$$\mathbb{E}\left[f\left(M_t\right)-f\left(M_r\right)\right]=\frac{1}{2}\mathbb{E}\left[\int_r^t f''\left(M_s\right)ds\right]=-\frac{\lambda^2}{2}\int_r^t\mathbb{E}\left[f\left(M_s\right)\right]ds$$

Take $G(t) = E[F(M_t)]$, we have

$$G'(t) = -\frac{\lambda^2}{2}G(t)$$

and the solution of $G(t)$ is the result.

Theorem 4.4. *Let M_t saticify the above proposition, then M_t is a standard Brownian motion.*

All that is left to show is the independent, normal increment. Since the process is adapted to filtration $\mathcal{F}_t$, then the independence is obvious, and the normality follows from the characteristic function.

5 Change of Measure and Girsanov Theorem

5.1 Absolutely continuous measures

This section we will be introducing measures into the play, as well as measure spaces.

Definition 5.1. *Suppose μ, ν are measures on space Ω with sigma-algebra $\mathcal{F}$, then*

- *ν is absolutely continuous with respect to μ, $\nu << \mu$, if for all $E \in \mathcal{F}$, $\mu(E) = 0 \to \nu(E) = 0$*

- *μ and ν are mutually absolutely continuous if $\nu << \mu$, $\mu << \nu$*

- *μ and ν are singular measures, $\mu \perp \nu$, if $\Omega = E \cup F$, and $\mu(E) = 0$, $\nu(F) = 0$*

The fundamental theorem we will be using is the Radon-Nikodym Theorem, we will now prove the theorem rigorously. First we introduce a lemma we will be using,

Lemma 5.1. *Suppose that ν and μ are finite measures on $(X, \mathbb{M})$. Then either $\nu \perp \mu$ or there exists $\epsilon > 0$ and $E \in \mathbb{M}$ such that $\mu(E) > 0$ and $\nu \geq \epsilon\mu$ on E.*

Theorem 5.1. *Radon-Nikodym Theorem Let ν be a σ-finite signed measure and μ a σ-finite positive measure on a measure space $(X, \mathbb{M})$ with corresponding sigma algebra. Then there exists a unique σ-finite signed measures λ, ρ such that*

$$\lambda \perp \mu, \rho << \mu, and \nu = \lambda + \rho$$

More over, there exists an extended μ-integrable function f such that $d\rho = f d\mu$.

Since we are only dealing with probability spaces, we can assume the two measures on the space are finite, positive measures. We also ignore the λ measure and assume $\nu << \mu$

Proof. Define set

$$\mathfrak{F} = \left\{ f : X \to [0, \infty] : \int_E f d\mu \leq \nu(E) \text{for all} E \in \mathbb{M}] \right\}$$

Then $\mathfrak{F}$ is nonempty as the zero function is in it. If, $f, g \in \mathfrak{F}$, then $hmax(f, g) \in \mathfrak{F}$.

$$\int_E h d\mu = \int_{E \cap A} f d\mu + \int_{E \backslash A} g d\mu \leq \nu(E \cap A) + \nu(E \backslash A) = \nu(E)$$

Let $a = sup\{\int f d\mu : f \in \mathfrak{F}\}$, then by above, $a < \nu(X) < \infty$. Then we can find a sequence $\{f_n\} in\, \mathfrak{F}$ that increasingly converge to a, by monotone convergence theorem, let $f = sup\{f_n\}$ we have $\int f_n = \int f$, so $f \in \mathfrak{F}$ Now we check f satisfies the requirement in the theorem. Observe $d\lambda = d\nu + f d\mu$ is singular with respect to $d\mu$, assume not, by lemma, there exists E and $\epsilon > 0$ such that $\mu(E) > 0$ and $\lambda \geq \epsilon\mu$ on E. Then $\epsilon\chi_E d\mu < d\lambda$, then we found a new function $f + \epsilon\chi_E$ that has integral value greater than a. So we reached a contradiction. ☐

On a rough sense, the above theorem gives us a 'derivative' of measures. Later in the chapter, when we want to switch the base measure from one to another, we can do so using this method, and take expectation of a random variable with respect to another setting.

Earlier in the book, we touched on the notion of conditional probability as the following. Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{G} \in \mathcal{F}$ is a sub σ-algebra. Then $E[X|\mathcal{G}]$ is the conditional probability given $\mathcal{G}$. More precisely, $Q(A) = E[1_A X]$, for $A \in \mathcal{G}$ defines a measure that satisfies $Q << P$. There is a $\mathcal{G}$ measurable random variable Y such that $Q(A) = E[1_A Y]$, for $A \in \mathcal{G}$, and Y is the conditional expectation of X given $\mathcal{G}$.

Example 5.1. *Let Ω be the set of continuous function from $[0, 1] \to \mathbb{R}$. Let B_t be the standard Brownian motion with 0 drift and σ variance, then there is a measure P_σ as the distribution of the* *"function-valued" random variable $t \to B_t$. If V is a subset of Ω, then $P_\sigma(V)$ is the probability that the Brownian motion lies in V. Furthermore, if $\sigma \neq \lambda$, then $P_\sigma \perp P_\lambda$. Certainly, define*

$$E_r = \left\{ f : \lim_{n\to\infty} \sum_{j=1}^{2^n} \left[f\left(\frac{j}{2^n}\right) - f\left(\frac{j-1}{2^n}\right)\right]^2 = r^2 \right\}$$

Then $P_\sigma(E_\sigma) = 1$, and $P_\lambda(E_\lambda) = 1$, and $E_\sigma \cap E_\lambda = \emptyset$

5.2 Give drift to a Brownian motion

This section we will be studying the different behavior of Brownian motion under different measures. Keep in mind expectation in probability is essentially just a integral. Suppose B_t is defined on the probability space $(\Omega, \mathcal{F}, P)$, now consider

$$M_t = exp\{mB_t - \frac{m^2 t}{2}\} \tag{25}$$

Then M_t is a martingale, by Itô's formula we have

$$dM_t = mM_t dB_t$$

Now define $Q_t(V) = e[1_V M_t]$ for $\mathcal{F}$ measurable event V. Equivalently for each t we have

$$dP = M_t dQ_t \tag{26}$$

For $s < t$, by the towering property, we have $Q_s(V) = E[1_V E(M_t|\mathcal{F}_S)] = Q_t(V)$. Now we write Q for the measure. and we claim

- For standard B_t in the P measure is a Brownian motion with drift m and variance 1 under the Q measure.

We can, of course, alter the variance as well, but that will be the topic for another day. The continuity of path is immediate, so we need to show the increments over period t are independent and normal with mean mt, variance t. To do so we will show the moment generating function is of the normal form, i.e.

$$E_Q\left(\exp\left\{\lambda\left(B_{t+s} - B_s\right)\right\} | \mathcal{F}_s\right) = e^{\lambda mt} e^{\lambda^2 t/2} \tag{27}$$

Since we are now dealing with more than one measure, the subscript Q denotes which measure we are taking expctation over. To establish the above, by the definition of conditional probability, we need to show for every $\mathcal{F}_s$ measurable set V

$$\begin{aligned} \mathbb{E}_Q\left[1_V \exp\left\{\lambda\left(B_{t+s} - B_s\right)\right\}\right] &= \mathbb{E}_Q\left[1_V e^{\lambda mt} e^{\lambda^2 t/2}\right] \\ &= e^{\lambda mt} e^{\lambda^2 t/2} Q(V) \end{aligned}$$

Equivalently,

$$\mathbb{E}\left[1_V \exp\left\{\lambda\left(B_{t+s} - B_s\right)\right\} M_{t+s}\right] = e^{\lambda mt} e^{\lambda^2 t/2} \mathbb{E}\left[1_V M_s\right]$$

Since $Y = B_{t+s} - B_s$ is independent of $\mathcal{F}_s$, we have

$$\begin{aligned}
E_Q\left(1_V \exp\left\{\lambda\left(B_{t+s} - B_s\right)\right\}|\mathcal{F}_s\right) &= E(1_V E\left(e^{\lambda Y} M_{t+s}|\mathcal{F}_s\right)) \\
&= E(1_V E\left(e^{\lambda Y} e^{mB_{t+s}} e^{-m^2(t+s)/2}|\mathcal{F}_s\right)) \\
&= E(1_V e^{-m^2 t/2} E\left(e^{\lambda Y} e^{mY} e^{mB_s} e^{-m^2(s)/2}|\mathcal{F}_s\right)) \\
&= E(1_V M_s e^{-m^2 t/2} E\left(e^{\lambda Y} e^{mY}|\mathcal{F}_s\right)) \\
&= E(1_V M_s e^{-m^2 t/2} \mathbb{E}\left[e^{(\lambda+m)Y}\right]) \\
&= E(1_V M_s) e^{-m^2 t/2} e^{(\lambda+m)^2 t/2} \\
&= E(1_V M_s) e^{\lambda^2 t/2} e^{\lambda m t}
\end{aligned}$$

The proof is completed. As a result of the above theorem, suppose X_t is a geometric Brownian motion with drift m, and variance σ, then we can find a new probability measure q such that

$$dB_t = rdt + dW_t$$

where W_t is a Brownian motion with respect to Q, hence X_t is

$$dX_t = X_t[(m + \sigma r]dt + \sigma dW_t$$

With respect to Q, X_t is a Brownian motion with same variance but new drift.

Example 5.2. *Suppose we have B_t the standard Brownian motion, and M_t, the martingale defined in (25). Then for $a > 0$, let $T_a = \inf\{t : B_t = a\}$. Then under measure Q as defined in (26), B_t is a Brownian motion with drift m.*
First note since $P\{T_a < \infty\} = 1$, we have

$$Q\left\{T_a < \infty\right\} = \mathbb{E}\left[M_{T_a} 1\left\{T_a < \infty\right\}\right] = \mathbb{E}\left[M_{T_a}\right]$$

also,

$$\begin{aligned}
Q\left\{T_a < \infty\right\} &= \mathbb{E}\left[\exp\left\{mB_{T_a} - \frac{m^2 T_a}{2}\right\}\right] \\
&= e^{am} \mathbb{E}\left[\exp\left\{-\frac{m^2 T_a}{2}\right\}\right]
\end{aligned}$$

Since we know $P\{T_a < \infty\} = 1$, we have

$$Q\{T_a < \infty\} = \int_{T_a < \infty} M_t dP = 1$$

Then,

$$\mathbb{E}\left[\exp\left\{-\frac{m^2 T_a}{2}\right\}\right] = e^{-am}$$

5.3 Girsanov Theorem

Girsanov Theorem establishes a way to observe a Brownian motion from the prospective of another measure. The M_t defined in the previous section, is one example, in this section, we generalize it to a family of (local) martingales. Suppose B_t is the standard Brownian motion, and M_t satisfies

$$dM_t = A_t M_t dB_t \tag{28}$$

Then we have seen before this is a exponential SDE, and has solution

$$M_t = exp\left\{\int_0^t A_s dB_s - \frac{1}{2}A_s^2 ds\right\}$$

Now we define probability measure $P*$ to be

$$P^*(V) = E[1_V M_t] \tag{29}$$

equivalently,

$$\frac{dP^*}{dP} = M_t \tag{30}$$

All the findings from the previous section still holds. One thing to keep in mind is for all $\mathcal{F}_t$ measurable X.

$$E^*[X] = E[XM_t]$$

Theorem 5.2. *(Girsanov Theorem) Suppose M_t is a nonnegative martingale satisfying () and let P^* be the probability measure defined in.If*

$$W_t = B_t - \int_0^t A_s ds$$

then with respect to the measure P^*, W_t is a standard Brownian motion. In other words

$$dB_t = A_t dt + dW_t$$

where W_t is a P^* Brownian motion.

Proof. The proof will use notations B_t and $B(t)$ interchangeably.
Here we will provide a derivation using binomial approximation. Suppose δt is give, by binomial approximation,

$$\mathbb{P}\{B(t+\Delta t) - B(t) = \pm\sqrt{\Delta t}|B(t)\} = \frac{1}{2}$$

Then the approximation for (28) is

$$\mathbb{P}\{M(t+\Delta t) = M(t)[1 \pm A(t)\sqrt{\Delta t}]|B(t)\} = \frac{1}{2}$$

In other words, the probability to jump one increment for $\mathbb{P}*$ is scaled by $M(t)$, then

$$\mathbb{P}^*\{B(t+\Delta t) - B(t) = \pm\sqrt{\Delta t}|B(t)\} = \frac{1}{2}[1 \pm A(t)\sqrt{\Delta t}]$$

As we showed in section 4.4, this implies,

$$E^*[B(t+\Delta t) - B(t)|B(t)] = A(t)\Delta t$$

In other words, in $P*$, the process obtained a drift of $A(t)$.

□