STOCHASTIC CALCULUS ON BROWNIAN MOTION AND
STOCHASTIC INTEGRATION

LINGYUE YU

Abstract. In this paper, I will first introduce the basics of measure theoretic probability and give a proof of Central Limit Theorem using moment generating functions. This section will allow us to explore stochastic processes and Brownian motion in a more rigorous way. Finally, building upon Brownian motion, we can formally explain Itô’s Integral and Itô’s Formula on the stochastic calculus.

Contents

1. Introduction 1
2. Probability Measure, Random Variable, and Expectation 1
3. Stochastic Processes 5
4. Brownian Motion 6
5. Itô’s Formula 8
Acknowledgments 11
References 11

1. Introduction

Calculus is the study of continuous change and typically characterized by differentiable functions. However, a stochastic process is a collection of random variables and its derivative is not easily defined, which renders ordinary calculus ineffective. To define a stochastic integral, we need to deal with randomness in stochastic processes. Essential definitions and theorems in probability introduced in the first section allow our further discussions. In the remaining sections, we will delve into the stochastic integration on Brownian motion, a stochastic process modeling continuous random motion, and examine the construction of stochastic integrals under the framework of Itô’s Integral and Itô’s Formula.

2. Probability Measure, Random Variable, and Expectation

Definition 2.1 (Sample Space). A sample space $\Omega$ is a non-empty set of outcomes.

Definition 2.2 (Algebra of Sets). Let $X$ be a set (We usually work within the context of sample space $\Omega$ in probability). An algebra is a collection $\mathcal{A}$ of subsets of $X$ such that

1. $\emptyset \in \mathcal{A}$ and $X \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $A^c := X \setminus A \in \mathcal{A}$. 

(3) If \( A_1, \ldots, A_n \in \mathcal{A} \), then \( \bigcup_{k=1}^{n} A_k \in \mathcal{A} \) and \( \bigcap_{k=1}^{n} A_k \in \mathcal{A} \).

**Definition 2.3** (σ-Algebra of Sets). For an algebra of sets, we have in addition that

(4) If \( A_1, A_2, \ldots \in \mathcal{A} \), then \( \bigcup_{k=1}^{\infty} A_k \in \mathcal{A} \) and \( \bigcap_{k=1}^{\infty} A_k \in \mathcal{A} \).

then we call \( \mathcal{A} \) an \( \sigma \)-Algebra of Sets.

In (4), we only allow countable unions and intersections. Since \( \bigcap_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{\infty} A_i)^c \), the requirement that \( \bigcap_{k=1}^{\infty} A_k \in \mathcal{A} \) would be redundant. The pair \( (X, \mathcal{A}) \) is called a measurable space. A set \( A \) is measurable or \( A - \text{measurable} \) if \( A \in \mathcal{A} \).

**Definition 2.4** (Probability Measure). Let \( \Omega \) be a sample space and let \( \mathcal{F} \) be a \( \sigma \)-Algebra on \( \Omega \). A probability measure is a function \( P : \mathcal{F} \to [0, 1] \) such that

(1) \( P(\emptyset) = 0 \) and \( P(\Omega) = 1 \).

(2) If events \( E_1, E_2, \ldots \) are pairwise disjoint, then

\[
\sum_{i=1}^{\infty} P(E_i) = P\left( \bigcup_{i=1}^{\infty} E_i \right).
\]

**Definition 2.5** (Probability Space). Given a set \( \Omega \) and a \( \sigma \)-Algebra \( \mathcal{F} \) on \( \Omega \), a probability space is the triple \( (\Omega, \mathcal{F}, P) \), where \( P \) is the probability measure.

**Definition 2.6** (Borel \( \sigma \)-Algebra). If we have an arbitrary collection \( \mathcal{C} \) of subsets of \( X \), define

\[
\sigma(\mathcal{C}) := \bigcap_{A \in \mathcal{C}} A.
\]

We call \( \sigma(\mathcal{C}) \) the \( \sigma \)-Algebra generated by \( \mathcal{C} \). If \( \mathcal{G} \) is the collection of all open sets on \( X \), then we define \( \mathcal{B} = \sigma(\mathcal{G}) \) to be the Borel \( \sigma \)-Algebra on \( X \).

**Definition 2.7** (Borel \( \sigma \)-Algebra on \( \mathbb{R} \)). \( \mathbb{R} \) is the Borel \( \sigma \)-Algebra generated by \( \mathbb{R} \).

**Proposition 2.8** (Lebesgue Measure). Let \( X = [0, 1] \) and \( \mathcal{B} \) be the Borel \( \sigma \)-Algebra. There is a unique measure \( \lambda \) on \( (X, \mathcal{B}) \) such that for any interval \( J \subset X \), we have that \( \lambda(J) = \text{length}(J) \). We call the measure as Lebesgue Measure.

Proof. The construction of Lebesgue Measure involves Carathéodory Extension Theorem using the algebra of finite unions and intersections of intervals. This proposition also works with \( X = \mathbb{R} \) because the space is \( \sigma \)-finite. The whole construction of Lebesgue measure please see Chapter 4 of [7] from Page 24 to 41. \( \square \)

**Definition 2.9** (Random Variable). Let \( (\Omega, \mathcal{F}, P) \) be a probability space and let \( (\mathbb{R}, \mathcal{B}, \lambda) \) be the real line endowed with the Borel \( \sigma \)-Algebra and the Lebesgue measure. A real random variable is a measurable function \( X : \Omega \to \mathbb{R} \).

**Definition 2.10** (Expectation). The expectation for a discrete random variable is

\[
\mathbb{E}[X] = \sum_{i} i P(X = i).
\]

For a continuous random variable, the expectation is defined as the following integral:

\[
\mathbb{E}[X] = \int_{\Omega} X dP.
\]
Definition 2.11 (Variance). The variance of random variable $X$ is defined as

Definition 2.12 (Conditional Expectation). Let $X$ be a random variable with $E[X] < \infty$. Then there exists a unique $\mathcal{G}$-measurable random variable $E(X|\mathcal{G})$ such that, for very bounded $\mathcal{G}$-measurable random variable $Y$, we have
$$E(XY) = E(E(X|\mathcal{G})Y).$$

The unique random variable $E(X|\mathcal{G})$ is called the conditional expectation.

Definition 2.13 (Distribution Function). The distribution function $F : \mathbb{R} \to [0, 1]$ of a random variable $X$ on $\Omega$ is defined by
$$F_X(x) := \mathbb{P}(X \leq x).$$
$F_X$ is an increasing function whose corresponding Lebesgue measure is $\mathbb{P}_X$.

Definition 2.14 (Density and Distribution). Let $X$ be a continuous random variable. A function $F$ is the distribution function of $X$ if
$$F(x) = \int_{-\infty}^{x} f(y)dy.$$ 
Function $f$ is the density of $X$ if
$$\mathbb{P}_X(A) = \int_A f(x)dx.$$  

Definition 2.15 (Standard Normal Distribution). A standard distribution is defined to be
$$\Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}dx.$$  

Definition 2.16 (Independence). Two events $E_1, E_2$ are independent if
$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2).$$

Lemma 2.17 (Borel-Cantelli). Let $\{E_n\}$ be a sequence of events in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If the sum of the probabilities of the $E_n$ is finite
$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty,$$
then the probability of infinitely many of them to occur is 0, which is
$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m\right) = 0.$$  

Proof. By definition, for each $n$,
$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m\right) \leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} E_m\right) \leq \sum_{m=n}^{\infty} \mathbb{P}(E_m) < \infty.$$ 
When $n \to \infty$, $\sum_{m=n}^{\infty} \mathbb{P}(E_m) \to 0$. Therefore, we have proved the claim. \qed
Definition 2.18 (Moment Generating Function). The moment-generating function of a random variable \( X \) is given by

\[
M(t) := \mathbb{E}[e^{tX}], \quad t \in \mathbb{R}.
\]

If two random variables have the same moment-generating function, they are said to be identically distributed.

Lemma 2.19. Let \( Z_1, Z_2, \ldots \) be a sequence of random variables having distribution functions \( F_{Z_n} \) and moment generating function \( M_{Z_n}, n \geq 1 \). Furthermore, let \( Z \) be a random variable having distribution function \( F_Z \) and moment generating functions \( M_Z \). If \( M_{Z_n}(t) \to M_Z(t) \) for all \( t \), then we have \( F_{Z_n}(t) \to F_Z(t) \) for all \( t \) at which \( F_Z \) is continuous.

This lemma is integral to the proof of the central limit theorem. As it is an advanced and technical proof, we will not prove this lemma in the paper. However, the whole proof can be seen in *Probability and Random Processes* [5].

Theorem 2.20 (Central Limit Theorem). Let \( X_1, X_2, \ldots, X_n \) be independent, identically distributed random variables with \( \mathbb{E}[X_i] = \mu \) and \( \text{Var}[X_i] = \sigma^2 < \infty \). Let

\[
Z_n = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma \sqrt{n}}.
\]

Then as \( n \to \infty \), the distribution of \( Z_n \) approaches a standard normal distribution. Precisely, this means: if \( a < b \)

\[
\lim_{n \to \infty} P(a \leq Z_n \leq b) = \Phi(b) - \Phi(a).
\]

Proof. We begin the proof with the assumption that \( \mu = 0, \sigma^2 = 1 \), and the moment generating function of \( X_i, M(t) \) exists and is finite. If not, we consider its standardized random variable \( X_i^* = (X_i - \mu) / \sigma^2 \) for the same. By definition, the moment generating function of \( \frac{X_i}{\sqrt{n}} \) is given as:

\[
M(\frac{t}{\sqrt{n}}) = \mathbb{E}[e^{t \frac{X_i}{\sqrt{n}}}].
\]

Thus, the moment generating function of \( \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}} \) is \([M(\frac{t}{\sqrt{n}})]^n\) since:

\[
\mathbb{E}\left[\exp\left(t \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}\right)\right] = \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} \frac{tX_i}{\sqrt{n}}\right)\right] = \prod_{i=1}^{n} \mathbb{E}[\exp(\frac{tX_i}{\sqrt{n}})] = \prod_{i=1}^{n} \mathbb{E}[e^{\frac{tX_i}{\sqrt{n}}}] = [M(\frac{t}{\sqrt{n}})]^n.
\]

Then, we define

\[
L(t) = \log M(t),
\]
and we have that
\[ L(0) = \log M(0) = \log \mathbb{E}(e^{0 \cdot X}) = 0 \]
\[ L'(0) = \frac{M'(0)}{M(0)} = \mathbb{E}(Xe^{0 \cdot X}) = \mathbb{E}[X] = \mu \]
\[ L''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2} = \frac{1 \cdot \mathbb{E}[X^2] - 0}{1^2} = \mathbb{E}[X^2] = 1 \]

Next, we need to show that \([M(t/\sqrt{n})]^n \to e^{t^2/2}\) as \(n \to \infty\). By taking log on both sides, we can see the moment generating function is equivalent to \(nL(t/\sqrt{n}) \to t^2/2\) as \(n \to \infty\). To show this result, we need to use L'Hospital's rule twice in (2.23) and (2.25):
\[ \lim_{n \to \infty} nL(t/\sqrt{n}) = \lim_{n \to \infty} \frac{L(t/\sqrt{n})}{n^{-1}} \]
\[ = \lim_{n \to \infty} \frac{-n^{-3/2}LL'(t/\sqrt{n})}{-2n^{-2}} \]
\[ = \lim_{n \to \infty} tL'(t/\sqrt{n}) \]
\[ = \lim_{n \to \infty} -\frac{n^{-3/2}LL''(t/\sqrt{n})}{-2n^{-4/2}} \]
\[ = \lim_{n \to \infty} -\frac{t^2n^{-3/2}LL''(t/\sqrt{n})}{-2n^{-4/2}} \]
\[ = \lim_{n \to \infty} t^2 \frac{LL''(t/\sqrt{n})}{2} \]
\[ = \lim_{n \to \infty} \frac{t^2}{2} L''(t/\sqrt{n}) \]
\[ = \frac{t^2}{2} L''(0) \]
\[ = \frac{t^2}{2}. \]

Since we know the moment generating function for a random variable under a normal distribution is \(M_Z(t) = e^{t^2/2}\). By the Lemma 2.20, we finish the proof the central limit theorem.

3. Stochastic Processes

Definition 3.1 (Stochastic Process). A stochastic process is a collection of random variables indexed by time \(t \in T \subset \mathbb{R}\). If \(T\) is an interval in \(\mathbb{R}\) then time is considered continuous; if \(T\) is a countable set in \(\mathbb{R}\) then time is considered discrete.

Definition 3.2 (Filtration). A discrete filtration of a set \(\Omega\) is a collection \(\{F_n\}\) of \(\sigma\)-algebras of subsets of \(\Omega\) such that \(F_n \subset F_{n+1}\) for all \(n \in \mathbb{N}\). A continuous filtration of a set \(\Omega\) is a collection \(\{F_n\}\) of \(\sigma\)-algebras of subsets of \(\Omega\) such that \(F_s \subset F_t\) for all \(s < t\).

Definition 3.3 (Natural Filtration). In particular, for a discrete stochastic process \(X_n\), the natural filtration \(\{F_n\}\) is a filtration such that each \(F_n\) is the \(\sigma\)-algebra generated by \(X_1, \ldots, X_n\). Namely, \(F_n\) contains all the information in \(X_1, \ldots, X_n\).

Definition 3.4 (Martingale). A sequence of random variables \(M_0, M_1, \ldots\) is called a martingale with respect to the filtration \(\{F_n\}\) if:

1. For each \(n\), \(M_n\) is an \(F_n\)-measurable random variable with \(\mathbb{E}[|M_n|] < \infty\)
(2) If \( m < n \), then
\[
\mathbb{E}[M_n | \mathcal{F}_m] = M_m.
\]

Or we can write it in another form:
\[
\mathbb{E}[M_n - M_m | \mathcal{F}_m] = 0.
\]

Moreover, a martingale \( M_t \) is called \textit{continuous martingale} if with probability one the function \( t \mapsto M_t \) is a continuous function.

4. Brownian Motion

Definition 4.1 (Brownian Motion). A stochastic process \( B_t \) is a \textit{Brownian motion} with respect to filtration \( \{ \mathcal{F}_t \} \) if it has the following three properties:

1. For \( s < t \), the distribution of \( B_t - B_s \) is normal with mean 0 and variance \( t - s \). We denote this by \( B_t - B_s \sim N(0, t - s) \).
2. If \( s < t \), the random variable \( B_t - B_s \) is independent of \( \mathcal{F}_s \).
3. With probability one, the function \( f : t \mapsto B_t \) is a continuous function of \( t \).

Theorem 4.2. Suppose \( B_t \) is a standard Brownian motion and \( a > 0 \). Then
\[
W_t := \frac{B_{at}}{\sqrt{a}}
\]
is a a standard Brownian motion.

Proof. We can simply check three properties of Brownian motion for \( W_t \). □

Theorem 4.3. A standard Brownian motion \( B_t \) is a continuous martingale with respect to filtration \( \{ \mathcal{F}_t \} \)

Proof. Let \( s < t \). Using properties of conditional expectation, we have that
\[
\mathbb{E}[B_t | \mathcal{F}_s] = \mathbb{E}[B_s | \mathcal{F}_s] + \mathbb{E}[B_t - B_s | \mathcal{F}_s] = B_s + \mathbb{E}[B_t - B_s] = B_s.
\]
By the third property of Brownian motion, we have proved \( B_t \) is a continuous martingale. □

Definition 4.4 (Quadratic Variation). Let \( B_t \) be a standard Brownian motion. We define the quadratic variation \( Q_t \) to be:
\[
Q_n = \sum_{j=1}^{n} \left[ B_t \left( \frac{j}{n} \right) - B_t \left( \frac{j-1}{n} \right) \right]^2.
\]

Definition 4.5 (Quadratic Variation). If \( X_t \) is a stochastic process, the quadratic variation is defined by:
\[
\langle X \rangle_t = \lim_{n \to \infty} \sum_{j \leq n \cdot t} \left[ X_t \left( \frac{j}{n} \right) - X_t \left( \frac{j-1}{n} \right) \right]^2.
\]

Theorem 4.6. Let \( B_t \) be a standard Brownian motion. We have \( \langle B \rangle_t = t \) for all \( t > 0 \)

Proof. We define function \( Q_n \)
\[
Q_n := \sum_{j=1}^{n} \left[ B_t \left( \frac{j}{n} \right) - B_t \left( \frac{j-1}{n} \right) \right]^2.
\]
where \( B_t \) is a standard Brownian motion, and then \( \langle B \rangle_t = \lim_{n \to \infty} Q_t \).
Equivalently, we can express $Q_t$ by defining independent random variables $Y_1, Y_2, \ldots, Y_n$:

$$Q_t = \frac{1}{n} \sum_{j=1}^{n} \left[ B_t \left( \frac{j}{n} \right) - B_t \left( \frac{j-1}{n} \right) \right]^2 := \frac{1}{n} \sum_{j=1}^{n} Y_j.$$  

As $Y_1, Y_2, \ldots, Y_n$ follow the distribution of $Z^2$ where $Z$ is a standard normal distribution, we have

$$E[Y_j] = E[Z^2] = 1, \quad E[Y_j^2] = E[Z^4] = (2 \cdot 2 - 1)! \cdot \sigma^2 = 3.$$  

by using the lemma $E(Z^{2n}) = (2n - 1)! \cdot \sigma^{2n}$, which can be proved from moment generating function or integration by parts. Then, we have


$$E[Q_n] = \frac{1}{n} \sum_{j=1}^{n} E[Y_j] = 1.$$  

$$\text{Var}[Q_n] = \frac{1}{n^2} \sum_{j=1}^{n} \text{Var}[Y_j] = \frac{2}{n}.$$  

Therefore, as $n \to \infty$, $\text{Var}[Q_n] \to 0$ and $E[Q_n] = 1$.

Similarly, for any time $t$, we define

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

As $n \to \infty$, $\text{Var}[Q_n(t)] \to 0$ and $E[Q_n] = t$. By definition, this calculation shows $\langle B \rangle_t = t$.

\[ \square \]

Theorem 4.7. With probability one, the function $f : t \mapsto B_t$ is nowhere differentiable.

Proof. We will consider $B_t$, $0 \leq t \leq 1$, and then take a countable collection of such Brownian motions to obtain $B_t$, $0 \leq t \leq \infty$. Suppose $B_t$ were differentiable at some $0 \leq t \leq 1$ with derivative $r$. Then there would exist $\delta$ such that for $|t - s| \leq \delta$ for continuous differentiable functions have the Hölder continuous of order 1,

$$|B_t - B_s| \leq 2 \leq r(t - s).$$

Theorems about the Hölder continuous of order $\alpha$ can be seen in Page 52 of [1]. We can find a positive integer $M < \infty$ such that for all sufficiently large integers $n$, there exists $k \leq n$ such that $Y_{k,n} \leq M/n$, where $Y_{k,n}$ is

$$\max\{|B \left( \frac{k+1}{n} \right) - B \left( \frac{k}{n} \right)|, |B \left( \frac{k+2}{n} \right) - B \left( \frac{k+1}{n} \right)|, |B \left( \frac{k+3}{n} \right) - B \left( \frac{k+2}{n} \right)|\}$$

max \{ |B \left( \frac{k+1}{n} \right) - B \left( \frac{k}{n} \right)|, |B \left( \frac{k+2}{n} \right) - B \left( \frac{k+1}{n} \right)|, |B \left( \frac{k+3}{n} \right) - B \left( \frac{k+2}{n} \right)| \}.$$
Let \( Y_n = \min\{Y_{k,n} : k = 0, 1, \ldots, n - 1\} \) and let \( A_M \) be the event that for all \( n \) sufficiently large, \( Y_n \leq M/n \). For each positive integer \( M \),

\[
\mathbb{P}\{Y_{k,n} \leq M/n\} = \left[ \mathbb{P}\{|B(1/n)| \leq M/n\}\right]^3 \\
= \left[ \mathbb{P}\{|n^{-1/2}|B_1| \leq M/n\}\right]^3 \\
= \left[ \int_{|x| \leq M/\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx \right]^3 \\
\leq \frac{2M}{\sqrt{n}} \left[ \frac{1}{\sqrt{2\pi}} \right]^3 \\
\leq \frac{M^3}{n^{3/2}}.
\]

Therefore,

\[
\mathbb{P}\{Y_n \leq M/n\} \leq \sum_{k=0}^{n-1} \mathbb{P}\{Y_{k,n} \leq M/n\} \leq \frac{M^3}{n^{3/2}} \to 0.
\]

This shows that \( \mathbb{P}(A_M) = 0 \) for each \( M \), and hence

\[
\mathbb{P}\left[ \bigcup_{M=1}^{\infty} A_M \right] = 0.
\]

However, this result contradicts that the event that \( B_t \) is differentiable at some point is contained in \( \bigcup M A_M \). Hence, \( B_t \) is no where differentiable on \( 0 \leq t \leq 1 \) and then on \( 0 \leq t \leq \infty \). \( \square \)

5. Itô’s Formula

**Definition 5.1** (Simple Adapted Process). The analogue of step functions for stochastic integral is simple processes. A process \( A_t \) is a simple process if there exist times \( \{t_0, t_1, \ldots, t_n\} \) satisfying

\[
0 = t_0 < t_1 < \cdots < t_n < \infty,
\]

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

\[
A_t = Y_t, \quad t_j \leq t < t_{j+1}.
\]

If \( A_t \) is a simple process we define

\[
Z_t = \int_0^t A_s dB_s,
\]

by

\[
Z_{t_j} = \sum_{i=0}^{j-1} Y_i [B_{t_{i+1}} - B_{t_i}],
\]

and more generally,

\[
Z_t = Z_{t_j} + Y_j [B_t - B - s] \quad t_j \leq t \leq t_{j+1},
\]

\[
\int_r^t A_s dB_s = Z_t - Z_r.
\]
Theorem 5.2 (Properties of Itô Integral). Suppose $B_t$ is a standard Brownian motion with respect to a filtration $\{F_t\}$, and $A_t, C_t$ are adapted processes with continuous paths.

1. **Linearity.** If $a, b$ are constants and $r < t$, then
   \[ \int_0^t (aA_s + bC_s)dB_s = a \int_0^t A_s dB_s + b \int_0^t C_s dB_s. \]
   \[ \int_0^t A_s dB_s = \int_0^r A_s dB_s + \int_r^t A_s dB_s. \]

2. **Variance rule.** The variance of $Z_t$ satisfies
   \[ \text{var}[Z_t] = \mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2]ds. \]

3. **Continuity.** With probability one, $t \mapsto Z_t$ is a continuous function.

The proofs can be seen in Section 3.2.3 Integration of Continuous Processes on Page 89 of [1].

Theorem 5.3 (Itô’s Formula). If $f$ is a function with two continuous derivatives, and $W_t$ is a standard Brownian motion, then

\[ f(W_t) - f(W_0) = \int_0^t f'(W_s)dW_s + \frac{1}{2} \int_0^t f''(W_s)d(W_s). \]

The formula in differential form is:

\[ d(W_t) = df(W_t)dt + \frac{1}{2} f''(W_t)dt. \]

Proof. We start by writing $f(W_t)$ as a telescoping sum,

\[ f(W_t) = f(W_0) + \sum_{j=0}^{n-1} [f(W_{(j+1)\frac{t}{n}}) - f(W_{j\frac{t}{n}})]. \]

Then we use Taylor’s formula about $W_{j\frac{t}{n}}$ to write

\[ f(W_{(j+1)\frac{t}{n}}) = f(W_{j\frac{t}{n}}) + f'(W_{j\frac{t}{n}})[W_{(j+1)\frac{t}{n}} - W_{j\frac{t}{n}}] + \frac{1}{2} f''(W_{j\frac{t}{n}})[W_{(j+1)\frac{t}{n}} - W_{j\frac{t}{n}}]^2 + o\left(\frac{t}{n}\right), \]

and

\[ f(W_t) - f(W_0) = \sum_{j=0}^{n-1} f'(W_{j\frac{t}{n}})[W_{(j+1)\frac{t}{n}} - W_{j\frac{t}{n}}] + \frac{1}{2} \sum_{j=0}^{n-1} f''(W_{j\frac{t}{n}})[W_{(j+1)\frac{t}{n}} - W_{j\frac{t}{n}}]^2 + \frac{n-1}{2} o\left(\frac{t}{n}\right). \]

As $n \to \infty$, the third term on the right tends to 0. Since $f'$ is continuous, the first term will approach

\[ \int_0^t f'(W_s)dW_s. \]

For the second term, we consider the general limit of

\[ \sum_{j=0}^{n-1} g(W_{j\frac{t}{n}})[W_{(j+1)\frac{t}{n}} - W_{j\frac{t}{n}}]^2, \]
where \( g \) is a continuous function. First consider the case where \( g \) is a constant function equal to 1. We let

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

The limit

\[
Q_t = \lim_{n \to \infty} Q_t^{(n)}.
\]

is defined as the quadratic variation of \( W_t \). \([W_{\frac{j+1}{n}} - W_{\frac{j}{n}}]^2\) has the same distribution as \( \left( \frac{1}{n} Z^2 \right) \), where \( Z \) is a standard normal random variable with mean 0 and variance 1. Therefore

\[
\mathbb{E}(Z^2) = 1. \quad \text{Var}(Z^2) = \mathbb{E}(Z^4) - [\mathbb{E}(Z^2)]^2 = 2.
\]

Hence, since the increments of \( W \) are independent,

\[
\mathbb{E}(Q_t^{(n)}) = \sum_{j=0}^{n-1} \mathbb{E}([W_{\frac{j+1}{n}} - W_{\frac{j}{n}}]^2) = t.
\]

\[
\text{Var}(Q_t^{(n)}) = \sum_{j=0}^{n-1} \text{Var}([W_{\frac{j+1}{n}} - W_{\frac{j}{n}}]^2) = n \mathbb{E}(\frac{1}{n} Z^2) = 2t^2/n.
\]

As \( n \to \infty \), the expectation of \( Q_t^{(n)} \) stays constant but the variance goes to 1. For any \( g \) let

\[
Q_t^{(n)}(g) = \sum_{j=0}^{n-1} g(t)[W_{\frac{j+1}{n}} - W_{\frac{j}{n}}]^2.
\]

The limit

\[
Q_t(g) = \lim_{n \to \infty} Q_t^{(n)}(g).
\]

If \( g \) is a step function of the form

\[
g(s) := u(W_{\frac{j}{m}}, t), \quad \text{where} \quad \frac{j}{m} t \leq s < \frac{j+1}{m} t,
\]

then

\[
Q_t(g) = \lim_{n \to \infty} Q_t^{(n)}(g)
\]

\[
= \lim_{k \to \infty} Q_t^{(km)}(g)
\]

\[
= \lim_{k \to \infty} \sum_{j=0}^{m-1} u(W_{\frac{j}{m}}, t) \sum_{i=0}^{k-1} [W_{\frac{j+i+1}{km}} - W_{\frac{j+i}{km}}]^2
\]

\[
= \sum_{j=0}^{m-1} u(W_{\frac{j}{m}}, t) \lim_{k \to \infty} \sum_{i=0}^{k-1} [W_{\frac{j+i+1}{km}} - W_{\frac{j+i}{km}}]^2.
\]

The result about quadratic variation in **Theorem 4.5** leads to

\[
\lim_{k \to \infty} \sum_{i=0}^{k-1} [W_{\frac{j+i+1}{km}} - W_{\frac{j+i}{km}}]^2 = \frac{t}{m}.
\]
Therefore,\[Q_t(g) = \sum_{j=0}^{m-1} u(W_{\frac{jt}{m}} t) \frac{t}{m}.\]

Now assume \( g \) is continuous. For each \( n \), let \( g_n \) be the step function.
\[g_n(s) = g(t), \quad \frac{j}{n} t \leq s < \frac{j+1}{n} t.\]

Note that \(|Q_t(g) - Q_t(g_n)| \leq \sup_{0 \leq s \leq t} |g(s) - g_n(s)| Q_t = t \sup_{0 \leq s \leq t} |g(s) - g_n(s)|.\]

The continuity of \( g \) implies that \( \sup_{0 \leq s \leq t} \to 0 \) as \( n \to \infty \). Hence
\[Q_t(g) = \lim_{n \to \infty} Q_t(g_n) = \lim_{n \to \infty} \sum_{j=0}^{n-1} g(\frac{j+1}{n} t) \frac{t}{n}.\]

The last expression is the representation of the integral of \( g \) as a limit of Riemman sums. Therefore, if \( g \) is continuous,
\[Q_t(g) = \int_0^t g(s)ds.\]

We plug this result into (5.4) to conclude the proof. There are several more presentations of Itô formula. However, they are beyond the scope of this short paper. For interested readers, please read Section 3.4 starting Page 105 from [1]. □

Acknowledgments

I would like to thank my mentor, Nixia Chen, for all her help and guidance throughout the REU with this paper. I would also like to thank Prof. Peter May for running the program. If there were not his and numerous instructor’s efforts, the program would have not been possible and enlightening during the pandemic. I have also learned a lot from Daniil Rudenko’s lectures and problem sets on combinatorics as well as Victor Pineda’s lectures on rigorous treatment of measure theoretic probability. Finally, I want to thank peers in the morning class for forming a wonderful study group to discuss mathematical ideas and problems.

References