`AS.110.653` : Stochastic Differential Equations

Notebook

James Guo

March 5, 2025

Contents

I Introduction to SDEs **1**
- I.1 Deterministic and Stochastic Differential Equations . . . 1
- I.2 Heuristics of SDEs . . . 3

II Probability Theory **5**
- II.1 Probability Space . . . 5
- II.2 Random Variable . . . 7
- II.3 Stochastic Process . . . 10
- II.4 Convergence of Probability Measure and Random Variables . . . 11
- II.5 Normal Random Variable . . . 14
- II.6 Brownian Motion . . . 17
- II.7 Conditional Expectation . . . 20
- II.8 Martingale . . . 22

III Stochastic Integration **24**
- III.1 Itô Integral . . . 24
- III.2 Measurability for Itô Integrals . . . 27
- III.3 Extensions of Itô Integral . . . 33

IV Itô Formula **34**
- IV.1 Itô Lemma . . . 34
- IV.2 Multidimensional Itô Formula . . . 37
- IV.3 Martingale Representation Theory . . . 40

Acknowledgements:

- This notebook records the course contents for `AS.110.653` Stochastic Differential Equationsinstructed by *Dr. Xiong Wang* at *Johns Hopkins University* in the Spring 2025 semester.
- It summaries lecture contents, notes, and adapts contents from the following text:
  - *Stochastic Differential Equations: An Introduction with Applications* by Bernt Øksendal.
- The notes is a summary of the lectures, and it might contain minor typos or errors. Please point out any notable error(s).

I Introduction to SDEs

I.1 Deterministic and Stochastic Differential Equations

Before getting into stochastic differential equations, we will see a more specific case, namely, ordinary differential equations.

Example I.1.1. Ordinary Differential Equation.

Consider an **ordinary differential equation** (ODE):

$$\begin{cases} \dot{x}(t) = b\big(x(t)\big), & t > 0, \\ x(0) = x_0. \end{cases}$$

Here, the $(\dot{-})$ is d/dt, $x_0 \in \mathbb{R}^n$ is the initial condition, and $b : \mathbb{R}^n \to \mathbb{R}^n$ is a given "good" vector field. Eventually, we have $x : [0, \infty) \to \mathbb{R}^n$ as the trajectory. ⌟

In applications, the ODE could be disturbed by a noise (potentially *Gaussian*), so we want to define a model to account for that. Hence, we formally define Stochastic differential equations.

Definition I.1.2. Stochastic Differential Equations.

A formal way to define **stochastic differential equations** (SDEs) is:

$$\begin{cases} \dot{x}(t) = b\big(x(t)\big) + \sigma\big(x(t)\big)\xi(t), & t > 0, \\ x(0) = x_0. \end{cases}$$

Here, the additional coefficients, respectively, are:

- b represents the **drift** coefficient,
- σ represents the **diffusion** coefficient, and
- ξ represents the m-dimensional **noise**, or the "white noise."

⌟

Remark I.1.3. In ODEs, we would enforce conditions on the vector field b to guarantee the existence of an unique solution. (c.f. Existence and Uniqueness theorem.) ⌟

Here, we can pose the following questions on SDEs:

1. What is ξ?
2. What is the solution to the SDE?
3. Are there existence and uniqueness on SDEs?
4. Are there asymptotic behaviors?

Then, we will introduce a few problems that concern SDEs.

Example I.1.4. Population Growth Model.

Let N be the population number and t is the time, we mode the population growth as:

$$\begin{cases} \dfrac{dN}{dt} = a(t)N(t), \\ N(0) = N_0, \end{cases}$$

where $a(t)$ can be interpreted as the control factor and N_0 is the initial population.
Note that we can model $a(t) = r(t) + \xi$, where $r(t)$ is the *growth rate* and ξ is the noise. ⌟

Example I.1.5. Filtering Problem.

Consider that Q is original function and Z is assorted with noise:

$$Z(s) = Q(s) + (\text{noise}).$$

We want to filter out the noise from observations over Z. ⌟

Example I.1.6. Dirichlet Problem (PDE).

Given a domain $U \subset \mathbb{R}^n$ and continuous function f on $\overline{U}$ such that:

$$\begin{cases} \Delta f = 0 & \text{in } U, \\ f = g & \text{on } \partial U. \end{cases}$$

Note that we need the boundary condition to make the PDE deterministic. (c.f. Laplace equation.) ⌟

Remark I.1.7. The solution to the above example could be complicated using the methods of PDEs. We can use SDEs or stopping time of SDEs to “solve” PDEs, namely through $\mathbb{E}[\tau_x^U]$. ⌟

Example I.1.8. Optimal Stopping Problem.

Let x_t model the price of asset or resource on the market and t represent the time. We can model through:

$$\frac{dx_t}{dt} = rx_t + \alpha x_t \cdot (\text{noise}).$$

We also acquire that the discount rate is known as ρ (Typically as the *bank rate*). The model aims to maximize the expected profit. ⌟

Furthermore, we have **Black-Sholes** option price formula for modeling the **Pricing of Option** problems.

I.2 Heuristics of SDEs

Recall the ODE as:
$$\frac{d}{dt}x(t) = b\big(x(t)\big),$$
and we let the noise be some random effects, *e.g.* measure errors or hidden parameters.
We assume that the discrete motion obeys:

$$x(t+\Delta t) - x(t) = F\big(t, x(t); \Delta t, \Gamma_{t,\Delta t}\big)$$

Here are some conditions with the discrete motion:

1. $F\big(t, x(t), 0, \Gamma_{t,0}\big) = 0$,
2. $\Gamma_{t,\Delta t} \sim \mathcal{N}(0, \Delta t)$,
3. $\Gamma_{t,\Delta t}$ is independent of $x(t)$. It only depends on the increment Δt.

In particular, We can have $\Gamma_{t,\Delta t}$ as $\Delta B_t \sim B_{t+\Delta t} - B_t$, where B is the Brownian motion.

When x is smooth we apply the Taylor expansion with respect to the third and forth variables (Δt and ΔB_t) centered at $\Delta t = 0$ and $\Delta B_t = 0$, yielding that:

$$\begin{aligned} F\big(t, x(t); \Delta t, \Delta B_t\big) - \underbrace{F\big(t, x(t), 0, 0\big)}_{0} &= \partial_4 F\big(t, x(t); \Delta t, \Delta B_t\big)\Delta B_t + \partial_3 F\big(t, x(t); \Delta t, \Delta B_t\big)\Delta t \\ &\quad + \frac{1}{2}\partial_4^2 F\big(t, x(t); \Delta t, \Delta B_t\big)(\Delta B_t)^2 + \frac{1}{2}\partial_3^2 F\big(t, x(t); \Delta t, \Delta B_t\big)(\Delta t)^2 \\ &\quad + \partial_3\partial_4 F\big(t, x(t); \Delta t, \Delta B_t\big)\Delta t \Delta B_t + R(\Delta t, B_t), \end{aligned}$$

where ∂_i means the partial derivative with respect to the i-th variable.

Remark I.2.1. Since we are dividing Δt on both sides, while $\Delta t \to 0$, all the terms with order greater than 1 of Δt could be omitted. ⌟

Hence, for the above Taylor approximation, we can get rid of the term $\frac{1}{2}\partial_3^2 F\big(t, x(t); \Delta t, \Delta B_t\big)(\Delta t)^2$ term since it involved $(\Delta t)^2$, while we can also omit the residue part $R(\Delta t, B_t)$.

Remark I.2.2. Properties of Gaussian Curve.

1. For random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, it is a normal distribution with center (mean) μ and variance σ^2. Hence, we have the following moments:
   - **First moment**: $\mathbb{E}[X] = \mu$,
   - **Second moment**: $\mathbb{E}[|X|^2] = \sigma^2$, and thus $\mathbb{E}[|X|] = |\sigma|$.
2. For a Gaussian curve, we can be *confident* around $[\mu - 3\sigma, \mu + 3\sigma]$ interval. ⌟

Recall that $\Delta B_t = B_{t+\Delta t} - B_t \sim \mathcal{N}(0, \Delta t)$, we can conclude with the moments that $\mathbb{E}[\Delta B_t] = 0$, $\mathbb{E}[|\Delta B_t|] = \sqrt{\Delta t}$, and $\mathbb{E}[|\Delta B_t|^2] = \Delta t$.

Thus, by substituting $\Delta t \Delta B_t \sim \Delta t \sqrt{\Delta t} = (\Delta t)^{3/2}$, so we can omit the term $\partial_3 \partial_4 F\big(t, x(t); \Delta t, \Delta B_t\big) \Delta t \Delta B_t$.

We can think of our Gaussian curve of $\Delta B_t \sim \mathcal{N}(0, \Delta t)$ as:

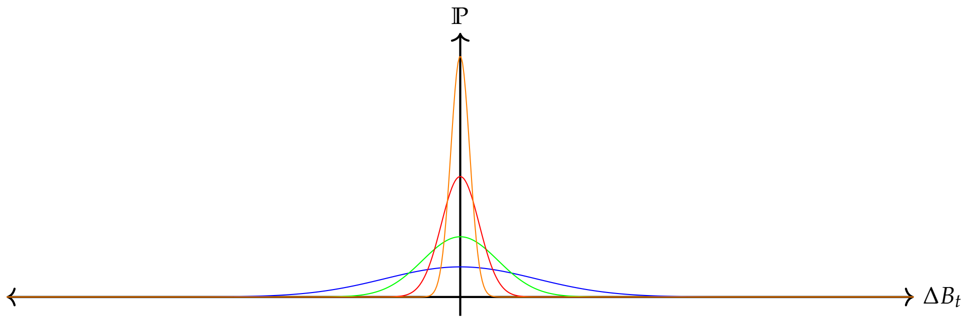

Figure I.1. Distribution of $\Delta B_t \sim \mathcal{N}(0, \Delta t)$ with $\Delta t = 1$ (blue), $\sqrt{2}/2$ (green), $1/2$ (red), $\sqrt{2}/4$ (orange).

Proposition I.2.3. Taylor Expansion of SDEs.

We consider the Taylor expansion of the discrete motion as:

$$x(t + \Delta t) - x(t) = \left(\partial_3 F\big(t, x(t); 0, 0\big) + \frac{1}{2}\partial_4^2 F\big(t, x(t); 0, 0\big)\right) \Delta t + \partial_4 F\big(t, x(t); 0, 0\big) \Delta B_t + \mathcal{O}(\Delta t)$$

$$dx(t) = b\big(t, x(t)\big) dt + \sigma\big(t, x(t)\big) dB_t, \qquad (fcn.1)$$

with $b\big(t, x(t)\big) = \partial_3 F + \frac{1}{2}\partial_4^2 F$ and $\sigma\big(t, x(t)\big) = \partial_4 F$.

Remark I.2.4. Here, we note that (*fcn*.1) is a "formal" derivation, since we approximately had $\sqrt{\Delta t}/\Delta t$, and it does not converge as $\Delta t \to 0$. Thus, the Brownian motion $B(t)$ is *not* differentiable everywhere. ⌟

It is notable that many functions are not "well-behaving," and we sometimes want to get around the derivatives by definition of integration (c.f. Functional analysis).

Example I.2.5. Formal Derivative of Characteristic Equation.

Consider the **characteristic equation** $\mathbb{1}_{[0,\infty)}$, which is defined as:

$$\mathbb{1}_{[0,\infty)}(x) = \begin{cases} 0 & \text{when } x < 0, \\ 1 & \text{when } x \geq 0. \end{cases}$$

We may have the formal derivative of the characteristic equation as:

$$\big(\mathbb{1}_{[0,\infty)}(x)\big)' = \delta_0(x) = \begin{cases} +\infty & \text{when } x = 0, \\ 0 & \text{when } x \neq 0. \end{cases}$$

In this way, we will get around the derivative of functions that are not "well-behaving." ⌟

II Probability Theory

II.1 Probability Space

Example II.1.1. Bertrand's Paradox.

Consider an equilateral triangle inscribed in a circle. Now, suppose that we are picking a chord, *randomly*, on the circle, what is the probability that the selected chord is longer than the side length of the equilateral triangle?
In general, there are three approaches, in which all of them give a different probability:

1. (Random Endpoints Method): Consider one endpoint of the chord fixed, the other endpoint free on the circle.

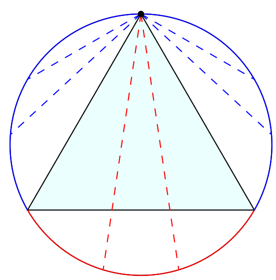

Figure II.1. Fixing an endpoint on the circle method.

Through this method, we can see that the chord is longer than the side length of the triangle at exactly 1/3 of the circumference. Hence, we have the probability as 1/3.

2. (Random Radial Point Method): Here, we fix a radius of the circle, and we look for the chords that are perpendicular to that radius.

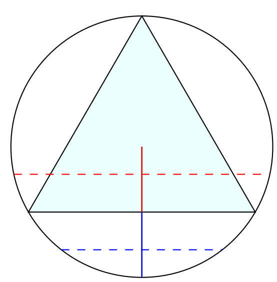

Figure II.2. Fixing a radius on the circle method.

Through this approach, it is not hard to observe that the chord is longer than the side length of the inscribed triangle on the top half and shorter on the bottom half. Thus, we have the probability as 1/2.

3. (Random Midpoint Method): Here, we note that the chord length is longer than the side length of the inscribed equilateral triangle if and only if it lies on the inscribed circle of the equilateral triangle.

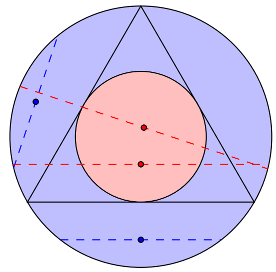

Figure II.3. Classifying the midpoint method.

Observe that the radius of the inner circle is exactly 1/2 of the outer circle, so the area of the inner circle is exactly 1/4 of the outer circle. Thereby, the probability such that the chord is longer than the side length of the inscribed triangle is 1/4.

Here, the three methodologies give distinct results since the "*randomness*" are defined differently, *i.e.*, the distribution is not at random in each case with respect to the other ones. ⌟

To rigorously study the previous problem, we need to define the probability space, what comes first is the basic *measure*-based definitions.

Definition II.1.2. σ-Algebra.

Let Ω be a given set, then a σ-algebra $\mathcal{F}$ on Ω is a family of subsets of Ω such that:

1. $\emptyset \in \mathcal{F}$,
2. $F \in \mathcal{F}$ implies that $F^c \in \mathcal{F}$, where $F^c = \Omega \setminus F$, and
3. For any $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$, we have $A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

⌟

Definition II.1.3. Probability Measure Space.

The pair $(\Omega, \mathcal{F})$ of σ-algebra together with a probability measure $\mathbb{P} : \mathcal{F} \to [0,1]$ forms a **probability measure space**, while $\mathbb{P}$ satisfies that:

1. $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$,
2. (σ-additivity): For any $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ such that they are mutually disjoint, *i.e.*, $A_i \cap A_j = \emptyset$ for all $i \neq j$, we have $\mathbb{P}\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

⌟

Remark II.1.4. The pair $(\Omega, \mathcal{F}, \mathbb{P})$ defined as above forms a **probability space**. ⌟

Here, we enforced the σ-algebra $\mathcal{F}$ as the set of *measurable sets*. Without this enforcement, this would be an **outer measure**, where we can alternatively defined the **Carathéodary measurable sets** as the σ-algebra.

Definition II.1.5. Complete Probability Space.

If $\mathcal{F}$ contains all subsets $G \subset \Omega$ with $\mathbb{P}$-outer measure zero. ⌟

Remark II.1.6. Note that since all sets of outer measure 0 is **Carathéodary measurable**, it is always possible to form a σ-algebra including all sets with outer measure zero. ⌟

Definition II.1.7. Smallest σ-algebra.

Given any family $\mathcal{U}$ of subsets of Ω, there is a smallest σ-algebra $\mathcal{H}_{\mathcal{U}}$ containing $\mathcal{U}$, where:

$$\mathcal{H}_{\mathcal{U}} = \bigcap_{\mathcal{H}:\mathcal{H} \text{ is } \sigma\text{-algebra of } \Omega, \text{ and } \mathcal{U} \subset \mathcal{H}} \mathcal{H}.$$

⌟

For example, let $\mathcal{U}$ be the collection of all open subsets of an Euclidean space ($\mathbb{R}^n$), then $\mathcal{B} = \mathcal{H}_{\mathcal{U}}$ is called the **Borel σ-algebra** on Ω, and the elements $B \in \mathcal{B}$ is called the Borel sets.

Remark II.1.8. The Lebesgue measurable sets are the completion of Borel measurable sets. ⌟

II.2 Random Variable

Definition II.2.1. $\mathcal{F}$-measurable Function (Random Variable).

Given $(\Omega, \mathcal{F}, \mathbb{P})$, then a function $Y : \Omega \to \mathbb{R}^n$ is called $\mathcal{F}$-measurable of:

$$Y^{-1}(U) := \{\omega \in \Omega : Y(\omega) \in U\} \in \mathcal{F}$$

for all open sets $U \in \mathbb{R}^n$. Here, we say that $(\Omega, \mathcal{F}, \mathbb{P})$ is a **random variable**. ⌟

Definition II.2.2. σ-algebra Generated by a Function.

Let $X : \Omega \to \mathbb{R}^n$ be any function, then the σ-algebra generated by X is smallest σ-algebra on Ω containing all the sets $X^{-1}(U)$ where $U \subset \mathbb{R}^n$ is open. ⌟

Here, one can show that $\mathcal{H}_X = \{X^{-1}(B) : B \in \mathcal{B}\}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}^n$. Clearly, $\mathcal{H}_X$ is $\mathcal{H}_X$-measurable, and $\mathcal{H}_X$ smallest σ-algebra with such property.

Proposition II.2.3. Doob-Dynkin.

If $X, Y : \Omega \to \mathbb{R}^n$ are two random variables, then Y is $\mathcal{H}_X$-measurable if and only if there exists a Borel measurable function $g : \mathbb{R}^n \to \mathbb{R}^n$ such that $Y = g \circ X$.

Proof. ($\Longleftarrow$:) Composition of two measurable functions is measurable, so Y is trivially $\mathcal{H}_X$ measurable when g is $\mathcal{B}(\mathbb{R}^n)$-measurable and X is $\mathcal{H}_X$-measurable.
($\Longrightarrow$:) Here, we follow a similar procedure of defining Lebesgue integrals in *measure theory*, that is, starting

from simple functions, then extending to positive functions, and eventually extend to all function as a sum of positive and negative parts.

1. First, suppose that Y is a simple function, we have:
$$Y = Y_n := \sum_{i=1}^{n} y_i \cdot \mathbb{1}_{A_i} \text{ for disjoint } \{A_i\} \subset \mathcal{H}_X = X^{-1}\big(\mathcal{B}(\mathbb{R}^n)\big).$$
Let $B_i = X(A_i)$, we know that $B_i \in \mathcal{B}(\mathbb{R}^n)$ since A_i is in the preimage of a Borel set, so we can define the function:
$$g_n := \sum_{i=1}^{n} y_i \cdot \mathbb{1}_{B_i},$$
so that g_n suits the requirement for any simple function.

2. Then, assume that $Y \geq 0$. Recall that simple functions are dense, there exists a non-decreasing sequence of simple functions $\{Y_n\}_{n=1}^{\infty}$ such that $Y_n \nearrow Y$. By the first step, we have $Y_n = g_n \circ X$, and we may define:
$$g(x) = \sup_{n\geq 1} g_n(x),$$
which exists on $\mathbb{R}^n$ and is measurable by convergence of monotone subsets, hence $g_n(X) \to g(X)$ and g satisfies that $Y = g \circ X$.

3. Eventually, consider $Y = Y^+ - Y^-$, where Y^+ and Y^- are measurable and non-negative.
By the previous step, we have $Y^+ = g^+ \circ X$ and $Y^- = g^- \circ X$ with measurable functions g^+ and g^-, so $Y = g \circ X$ where $g = g^+ - g^-$.

Therefore, we finish the proof of the equivalent statement. □

Definition II.2.4. Distribution.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space with random variable X. Every X induces a probability measure on $\mathbb{R}^n$ defined by:
$$\mu(B) = \mathbb{P}\big(X^{-1}(B)\big),$$
where μ_X is called the distribution of X. ⌟

Example II.2.5. Normal Distribution.

Consider X as a normal random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L}\big)$.
Graphically, we may distinguish the density function (ρ_X) and the cumulative density (μ_X): We can think of our Gaussian curve of $\Delta B_t \sim \mathcal{N}(0, \Delta t)$ as:

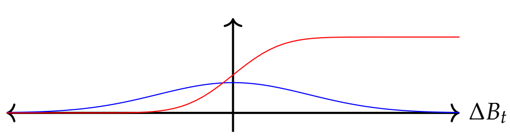

Figure II.4. Probability density function (blue) and cumulative density function (red) of $\mathcal{N}(0,1)$.

Here, we consider the density function as $\rho_X(x)$ as the density, the distribution would be induced over $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ that is:

$$\mu_X\big((-\infty, x)\big) = \int_{-\infty}^{x} \rho_X(y)dy,$$

and for any Borel set $B \in \mathcal{B}(X)$, we have $\mu_X(B) = \int_B \rho_X(x)dx$. ⌟

With these basics about probability, we may define more concepts related to probability.

Definition II.2.6. Expectation.

If $\int_\Omega \big|X(\omega)\big| d\mathbb{P}(\omega) < \infty$ (integrable), then:

$$\mathbb{E}[X] := \int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} x d\mu_X(x) = \int_{\mathbb{R}^n} x\rho_X(x)dx.$$

This is called the expectation of X with respect to $\mathbb{P}$. ⌟

More generally, if $f : \mathbb{R}^n \to \mathbb{R}$ is Borel measurable and $\int_\Omega \big|f\big(x(\omega)\big)\big| d\mathbb{P}(\omega) < \infty$, then:

$$\mathbb{E}\big[f(x)\big] = \int_\Omega f\big(X(\omega)\big)d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} f(x)d\mu_X(x).$$

Definition II.2.7. L^p-norm and L^p-space.

If $X : \Omega \to \mathbb{R}^n$ is a random variable and $p \in [1, \infty)$, we defined the ***L^p*-norm** of X (denoted $\|X\|_p$) as:

$$\|X\|_p = \|X\|_{L^p(\mathbb{P})} = \left(\int_\Omega |X(\omega)|^p d\mathbb{P}(\omega)\right)^{1/p}.$$

The corresponding ***L^p*-space** are defined by:

$$L^p(\mathbb{P}) = L^p(\Omega) = \{X : \Omega \to \mathbb{R}^n \mid \|X\|_p < \infty\}.$$

⌟

Other than some definition differences, the Lebesgue measure and probability measure differs in the definition of **independence**.

Definition II.2.8. Independence.

Two subsets $A, B \in \mathcal{F}$ are called independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
A collection of $\mathcal{A} := \{\mathcal{H}_i : i \in I\}$ of families $\mathcal{H}_i$ of measurable sets is independent if:

$$\mathbb{P}(H_{i_1} \cap \cdots H_{i_k}) = \mathbb{P}(H_{i_1}) \cdots \mathbb{P}(H_{i_k})$$

for all choices $H_{i_1} \in \mathcal{H}_{i_1}, \cdots, H_{i_k} \in \mathcal{H}_{i_k}$ with different indices $i_1, \cdots, i_k$.
A collection of random variables $\{X_i\}_{i \in I}$ is independent if the collection of $\mathcal{H}_{X_i}$ is independent. ⌟

Remark II.2.9. If $X, Y : \Omega \to \mathbb{R}$ are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ provided that $\|X\|_1 < \infty$ and $\|Y\|_1 < \infty$. ⌟

Remark II.2.10. With independence, suppose that $\mathbb{P}(B) > 0$, then we have:

$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A \mid B),$$

which is the conditional probability. Hence, any information about B gives no clue on what A is. ⌟

II.3 Stochastic Process

Definition II.3.1. Stochastic Process.

A **stochastic process** is a parametrized collection of **random variables**:

$$\{X_t\}_{t \in \mathcal{T}}.$$

⌟

Note that we can have $\mathcal{T} = \mathbb{Z}^+$, then we have $X_1, X_2, \cdots$.
We can also have $\mathcal{T} = [0, 1]$, which is over a uncountable set of indices.

Remark II.3.2. The parametric space $\mathcal{T}$ is usually the **half-line** $[0, \infty)$. We sometimes may also use $[a, b]$ or $\mathbb{Z}^+$. Then, for each fixed $t \in T$, we have a random variables:

$$\omega \mapsto X_t(\omega), \text{ for any } \omega \in \Omega.$$

For each fixed $\omega \in \Omega$, we can consider the function:

$$t \mapsto X_t(\omega), \text{ for any } t \in \mathcal{T}.$$

Also, when nothing is fixed, we can consider the multivariable function:

$$(t, \omega) \mapsto X_t(\omega) =: X(t, \omega), \text{ for any } (t, \omega) \in \mathcal{T} \times \Omega.$$

⌟

Remark II.3.3. Cylinderical Sets.

The σ-algebra $\mathcal{F}$ will contain the σ-algebra $\mathcal{B}$ generated by sets of the form:

$$\{\omega : \omega(t_i) \in F_i, \text{ where } i \in \mathcal{I} \text{ and } F_i \in \mathbb{R}^n \text{ are Borel sets}\}.$$

⌟

Consider the Brownian motions, say:

$$\widetilde{\Omega} = \mathbb{R}^T = \mathbb{R}^{[0,1]}.$$

We note that $[0, 1]$ is an uncountable set, so we want to have some $\mathcal{I} = \{1, 2, \cdots\}$, which is countable, or even finite.

Remark II.3.4. Note that it is hard to observe a uncountably infinite set for Brownian motion. The common strategy to use is to consider a countable (or finite) subset of the domain and observe if the Brownian motion falls into the designated area for each value in the observed subset of the domain. In particular, we enforce the designated area to be a Borel set. ⌟

Definition II.3.5. Finite Dimensional Distribution.

The **finite dimensional distribution** of the process $X = \{X_t\}_{t \in \mathcal{T}}$ are the $\mu_{t_1,\cdots,t_k}$ defined on $(\mathbb{R}^n)^k$, for $k = 1, 2, \cdots$ by:

$$\mu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \mathbb{P}\left[X_{t_1} \in F_1, \cdots, X_{t_n} \in F_k\right]$$

for $t_i \in \mathcal{T}$, and $F_1, \cdots, F_k \in \mathcal{B}(\mathbb{R}^n)$. ⌟

Theorem II.3.6. Kolmogorov's Extension Theorem.

For all $t_1, \cdots, t_k \in \mathcal{T}$, where $k \in \mathbb{N}$, let $V_{t_1,\cdots,t_k}$ be the probability measure on $(\mathbb{R}^n)^k$ such that:

(K1) $V_{t_{\sigma(1)},\cdots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = V_{t_1,\cdots,t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)})$, and

(K2) $V_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = V_{t_1,\cdots,t_k,t_{k+1},\cdots,t_{k+m}}(F_1 \times \cdots \times F_k \times \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{m})$.

Then there exists a probability measure $(\Omega, \mathcal{F}, \mathbb{P})$ and stochastic process $\{X_t\}_{t \in \mathcal{T}}$ on Ω, where $X_t : \Omega \to \mathbb{R}^n$ such that:

$$V_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \mathbb{P}(X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k) \text{ for } t_1, \cdots, t_k \in \mathcal{T} \text{ and } F_1, \cdots, F_k \in \mathcal{B}(\mathbb{R}^n).$$

This theorem makes sure that a finite distribution would coincide with the probability distribution, so it is an important remark on SDEs. The proof of the theorem is omitted due to its high complexity.

II.4 Convergence of Probability Measure and Random Variables

Setup II.4.1. For this section, we set down a measure space $\big(E, \mathcal{B}(E)\big)$, where E is a topology and $\mathcal{B}(E)$ is the σ-algebra over E. ⌟

Definition II.4.2. Weak Convergence.

Let $\{\mu_n\}_{n \in \mathbb{N}^+}$ be a sequence of finite measures on $\big(E, \mathcal{B}(E)\big)$, it **converges weakly** to μ if for every continuous bounded function $f : E \to \mathbb{R}$:

$$\lim_{n \to \infty} \int f d\mu_n = \int f d\mu.$$

⌟

Setup II.4.3. Let $\{X_n\}_{n\in\mathbb{N}^+}$ be a sequence, where X_n are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in $\big(E, \mathcal{B}(E)\big)$. ⌟

Definition II.4.4. Almost Surely Convergence.

Consider $\{X_n\}_{n\in\mathbb{N}^+}$, X_n converges to X **almost surely**, denoted by $X_n \xrightarrow{\text{a.s.}} X$ if there exists a negligible event $N \in \mathcal{F}$ such that $\mathbb{P}(N) = 0$ in which :

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \text{ for every } \omega \in \Omega \setminus \mathbb{N}.$$

⌟

Definition II.4.5. Convergence in Probability.

Consider $\{X_n\}_{n\in\mathbb{N}^+}$, it **converges** to X **in probability**, denoted by $X_n \xrightarrow{\mathbb{P}} X$ if for all $\delta > 0$:

$$\lim_{n\to\infty} \mathbb{P}\big(d(X_n, X) > \delta\big) = 0.$$

⌟

Note that convergence **almost surely** is a stronger conclusion than convergence **in probability**, since we have $\delta > 0$ fixed for convergence in probability and that is not free over convergence almost surely.

Definition II.4.6. L^p-Convergence.

Consider $\{X_n\}_{n\in\mathbb{N}^+}$, and $E = \mathbb{R}^n$, it converges to X in L^p, denoted by $X_n \xrightarrow{L^p} X$ if $X \in L^p$ and:

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

⌟

Definition II.4.7. Convregence in Law.

Consider $\{X_n\}_{n\in\mathbb{N}^+}$, it converges to X in Law, denoted $X_n \xrightarrow{\mathscr{L}} X$ as $\mu_n \xrightarrow{w} \mu$, where μ_n is a distribution of X_n and μ is the distribution of X. ⌟

Proposition II.4.8. Relationship of Convergences.

$$\begin{tikzcd}
L^p \arrow[dr, "\text{implies}"'] & & \text{A.S.} \arrow[dl, shift right, "\text{implies}"'] \\
& \mathbb{P} \arrow[ur, shift right, "\exists\text{ subset}"'] \arrow[d, shift right, "\text{implies}"'] & \\
& \mathscr{L} \arrow[u, shift right, "\text{When } \mathscr{L} \text{ converges to constant}"'] &
\end{tikzcd}$$

The deduction of the above relationships are omitted, while some of them are parallel to convergence of sequences of functions.

Example II.4.9. Construction of Stochastic Process.

Consider $X_n = \{Z, -Z, Z, -Z, \cdots\}$ where $Z \sim \mathcal{N}(0,1)$, then:

- $X \xrightarrow{\mathscr{L}} X \sim \mathcal{N}(0,1)$ since we have μ_n having the distribution $\mathcal{N}(0,1)$.
- $X_n \xrightarrow{\mathbb{P}} n$ is **not true**. Suppose for all δ that $\mathbb{P}\big(d(X_n, X) > \delta\big) = 0$, then $\{X_n\}$ must be Cauchy, then we must have:

  $$\mathbb{P}(|X_{2k+1} - X_{2k}| > \delta) = \mathbb{P}(|Z| > \delta/2) > 0,$$

  which is a contradiction.

⌟

Proposition II.4.10. Borel-Cantelli Lemma.

Let $\{A_n\}_{n\in\mathbb{N}^+}$ be a sequence of sets, and:

$$A = \limsup_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcup_{k\geq n} A_k,$$

then:

1. Suppose $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, if $\mathbb{P}(A) = 0$, then we
2. (0-1 Law) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$, and $\{A_n\}_n$ are independent, then $\mathbb{P}(A) = 1$.

Then, we will recall the three fundamental convergence theorems in Real Analysis.

Theorem II.4.11. Convergence Theorems in Real Analysis.

The following convergence theorems holds over $(\Omega, \mathcal{F}, \mathbb{P})$:

- (Fatous's Lemma). If $X_n \geq 0$, then $\mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.
- (Monotone Convergence Theorem, MCT). If $X_n \nearrow X$, then $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[\lim_{n\to\infty} X_n]$.
- (Lebesgue's Dominant Convergence Theorem, DCT). If $X_n \xrightarrow{\mathbb{P}} X$, $|X_m| \leq Y$, and $\mathbb{E}[|Y|] < \infty$, then $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[\lim_{n\to\infty} X_n] = \mathbb{E}[X]$.

These proofs aligns with the proof of the convergences in Real Analysis, please refer to any measure theory textbook for a parallel proof.

Remark II.4.12. Discrete and Continuous time Stochastic Process.

A discrete time stochastic process is $\{X_n\}_{n\in\mathbb{Z}^+}$, and a continuous time stochastic process is $\{X_t\}_{t\in[0,\infty]}$. ⌟

After the construction of a countable (or finite) number of observation points, we would want to develop a finite dimensional distribution:

$$\mu_{t_1,\cdots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = \mathbb{P}[X_{t_1}, \cdots, X_{t_k}].$$

II.5 Normal Random Variable

One goal of normal random variable is towards the **Brownian motion**, which was developed in 1827 from *R. Brown* of the "rapid oscillatory motion."

Remark II.5.1. Sketch on Brownian Motion.

Let $F_1, \cdots, F_k$ be Borel sets in $\mathbb{R}^n$, we have the **Brownian motion** measured by:

$$\mu_{t_1,\cdots,t_k}(F_1, \cdots, F_k) = \mathbb{P}[B_{t_1} \in F_1, \cdots, B_{t_k} \in F_k].$$

Here, in particular, let $t_1 = 0$ and $t_2 = t$, we have:

$$\mu_{0,t} = \mu_t = \mathbb{P}(b_t \in F_1),$$

and when $t_1 = 0$, $t_2 = s$, and $t_3 = t$, we have:

$$\mu_{0,s,t} = \mu_{s,t} = \mathbb{P}(B_s \in F_1, B_t \in F_2) = \mathbb{P}(B_s \in F_2) \cdot \mathbb{P}(B_t \in F_1 \mid B_s \in F_2),$$

by the Markov property. ⌟

In 1900, there are motions used to detect stock price fluctuations.
In 1905, Einstein derived the transition density for:

$$\mathbb{P}[B_t \in F] \sim \mathcal{N}.$$

In 1923, Wiener rigorously defined the math over $(C[0,1], \mathcal{B}(C[0,1]), \mathbb{P})$, *i.e.*, infinite dimensional space.
In 1933, Kolmogrov developed the extension theory.
In 1960s, L. Gross defined the **Abstract Wiener Space** of $(\mathbb{H}, \mathbb{B}, \mathbb{P})$, which is over the a Hilbert space.

Definition II.5.2. 1-dimensional Normal Random Variable.

Let the probability space be $(\Omega, \mathcal{F}, \mathbb{P})$, $X : \Omega \to \mathbb{R}$ is normal if the distribution of X has density:

$$\rho_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right),$$

where m is the mean and σ^2 is the variance. Meanwhile, the probability is:

$$\mathbb{P}(X \in G) = \int_G \rho_x(x)dx \text{ for all Borel sets } G \in \mathbb{R}.$$

⌟

It is noted that this is a distribution since $\int_{\mathbb{R}} \rho_x(x)dx = 1$.

Definition II.5.3. *n*-dimensional Normal Random Variable.

Let the probability space be $(\Omega, \mathcal{F}, \mathbb{P})$, with $X : \Omega \to \mathbb{R}^n$, it is **multi-normal** $\mathcal{N}(m, C)$ if the distribution of X has density of the form:

$$\rho_X(x_1, \cdots, x_n) = \frac{\sqrt{|A|}}{(2\pi)^{n/2}} \exp\bigg(-\frac{1}{2} \sum_{j,k} (x_j - m_j) a_{j,k} (x_k - m_k) \bigg),$$

where $m = (m_1, \cdots, m_n) \in \mathbb{R}^n$ and $C^{-1} = A = \left[a_{j,k} \right] \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. ⌟

Definition II.5.4. Characteristic Function.

Consider the random variable $X : \Omega \to \mathbb{R}^n$, we let the **characteristic function** $\phi_X : \mathbb{R}^n \to \mathbb{C}$ be defined as:

$$\phi_X(u_1, u_2, \cdots, u_n) = \mathbb{E}\big[\exp\{ \mathrm{i}(u_1 x_1 + \cdots + u_n x_n) \} \big] = \int_{\mathbb{R}^n} e^{\mathrm{i}\langle u, x \rangle} \underbrace{\mathbb{P}(x \in dX)}_{\rho_X(x)dx \text{ if the density exists}} .$$ ⌟

Remark II.5.5. The characteristic function is the **Fourier transformation** of X with measure $\mathbb{P}[X \in dx]$. ⌟

Then, we will give a few properties of the normal distributions and characteristic functions.

Theorem II.5.6. Unique Determination of Distribution.

ϕ_X determine the distribution of X uniquely.

Theorem II.5.7. Characteristic Function for Normal Distribution.

If $X : \Omega \to \mathbb{R}^n$ is normal $\mathcal{N}(m, C)$, then:

$$\phi_X(u_1, \cdots, u_n) = \exp\bigg(-\frac{1}{2} \sum_{j,k} (x_j - m_j) a_{j,k} (x_k - m_k) \bigg) \text{ for all } u_1, \cdots, u_n \in \mathbb{R}.$$

Theorem II.5.8. Equivalence under Sequence of Random Variables.

Let $X_j : \Omega \to \mathbb{R}$ be random variables for $1 \leq j \leq n$, then $X = (X_1, \cdots, X_n)$ is normal if and only if $Y = \lambda_1 X_1 + \cdots + \lambda_n X_n$ for all $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$.

Proof. ($\Longrightarrow$:) Suppose X_j is normal for all $1 \leq j \leq n$, then:

$$\mathbb{E}\left[\exp\left(\mathrm{i}u \sum_{j=1}^n \lambda_j X_j \right) \right] = \exp\left[-\frac{1}{2} \sum_{j,k} u\lambda_j c_{j,k} u\lambda_k + \mathrm{i} \sum_j u\lambda_j m_j \right] = \exp\left[-\frac{u^2}{2} \sum_{j,k} \lambda_j c_{j,k} \lambda_k + \mathrm{i}u \sum_j \lambda_j m_j \right].$$

Therefore, Y is normal with $\mathbb{E}[Y] = \sum_j \lambda_j m_j$ and $\mathrm{Var}[Y] = \sum_{j,k} \lambda_j c_{j,k} \lambda_k$.

($\Longleftarrow$:) If $Y = \sum_{j=1}^{n} \lambda_j m_j$ is normal with $\mathbb{E}[Y] = m$ and $\mathrm{Var}[Y] = \sigma^2$, then:

$$\mathbb{E}\left[\exp\left(\mathrm{i}u \sum_{j=1}^{n} \lambda_j x_j\right)\right] = \exp\left(-\frac{1}{2}u^2\sigma^2 + \mathrm{i}\sum\right),$$

where $m = \sum_j \lambda_j m_j$ for $m_j = \mathbb{E}[X_j]$ and $\sigma^2 = \mathbb{E}\left[\left(\sum_j \lambda_j X_j - \sum_j \lambda_j m_j\right)^2\right] = \sum_{j,k} \lambda_j \lambda_k \mathbb{E}[(x_j - m_j)(X_k - m_k)]$. Since m_j's are arbitrary, then X is normal. □

Theorem II.5.9. Uncorrelated $\Longrightarrow$ Independent for Normal Distributions.

Let $Y_0, Y_1, \cdots, Y_n$ be real random variables on Ω. Assume $X = (Y_0, \cdots, Y_n)$ is normal and Y_0 and Y_j are uncorrelated for all $j \geq 1$, *i.e.*:

$$\mathbb{E}[(Y_0 - \mathbb{E}[Y_0])(Y_j - \mathbb{E}[Y_j])] = 0 \text{ for } 1 \leq j \leq n.$$

Then Y_0 is independent of $\{Y_1, \cdots, Y_n\}$.

The idea to prove the above theorem is by using the characteristic function, and obtain that:

$$\phi_X(u_1, u_2, \cdots, u_n) = \phi_X(u_1) \cdot \phi_X(u_2) \cdots \phi_X(u_n),$$

which is the definition of independence.

Remark II.5.10. Note that independence implies uncorrelated for all random variable, so we have them equivalent with normal distributions. ⌟

Theorem II.5.11. Convergent Sequence of Normal Distribution Converges to Normal Distribution.

Suppose $X_k : \Omega \to \mathbb{R}^n$ is normal for all k and that $X_k \to X$ in $L^2(\Omega)$, *i.e.*:

$$\mathbb{E}[|X_k - X|^2] \to 0 \text{ as } k \to \infty.$$

Then X is normal.

Proof. First, note that $\left|e^{\mathrm{i}\langle u,x\rangle} - e^{\mathrm{i}\langle u,y\rangle}\right| < |u| \cdot |x - y|$, we have:

$$\mathbb{E}[\left|e^{\mathrm{i}\langle u,x\rangle} - e^{\mathrm{i}\langle u,y\rangle}\right|^2] \leq |u|^2 \cdot \mathbb{E}[|X_k - X|^2] \to 0 \text{ as } k \to \infty.$$

Thus, we have:

$$\mathbb{E}[e^{\mathrm{i}\langle u,x\rangle}] \to \mathbb{E}[e^{\mathrm{i}\langle u,y\rangle}] \text{ as } k \to \infty.$$

Therefore, X is normal with mean $\mathbb{E}[X] = \lim_{k\to\infty} \mathbb{E}[X_k]$ and covariance $C = \left[x_{j,n}\right] = \lim_{k\to\infty} C_k$. □

Remark II.5.12. To develop the Brownian motion, we consider the independence, we will have:

$$\begin{aligned}\nu_{t_1,\cdots,t_k}(F_1,\cdots,F_k) &= \int_{F_1\times\cdots\times F_k} \rho_X(x_1,\cdots,x_k)dx_1dx_2\cdots dx_k \\ &= \int_{F_1\times\cdots\times F_k} \rho_{t_1}(x_1)\rho_{t_2-t_1}(x_2-x_1)\cdots\rho_{t_k-t_{k-1}}(x_k-x_{k-1})dx_1dx_2\cdots dx_k,\end{aligned}$$

where we interpret the distributions are all normal distributions. ⌟

II.6 Brownian Motion

For simplicity, we first reduce the Brownian Motion to 1-dimensional case.

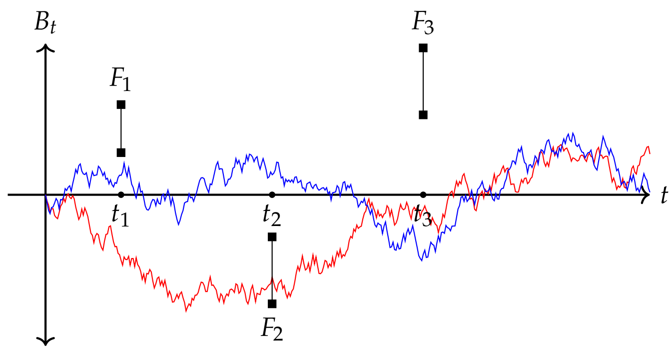

Figure II.5. Illustration of Brownian Motion in 1D.

Now, consider for $0 \le t_1 \le t_2 \le \cdots \le t_k$, we define:

$$\nu_{t_1,\cdots,t_k}(F_1\times\cdots\times F_k) = \int_{F_1\times\cdots\times F_k} \rho(t_1,x_0,x_1)\rho(t_2-t_1,x_1,x_2)\cdots\rho(t_k-t_{k-1},x_{k-1},x_k)dx_1\cdots dx_j.$$

Here, the transition density is for all $x, y \in \mathbb{R}^n$, $t > 0$ that:

$$\rho(t,x,y) = \rho(t,x-y) = (2\pi t)^{-n/2}\exp\left(-\frac{|x-y|^2}{2t}\right),$$

and for example $n = 1$, we have:

$$\rho(t_2-t_1,x_1,x_2) = \frac{1}{\sqrt{2\pi(t_2-t_1)}}\exp\left[-\frac{|x_1-x_2|^2}{2(t_2-t_1)}\right].$$

Note that this definition is based of Theorem II.3.6 Kolmogrov's extension theorem so we make a finite dimensional probability distribution into a continuous distribution.

Definition II.6.1. Brownian Motion.

The above processes is called (a version of) **Brownian motion** starting at x. ⌟

Proposition II.6.2. Properties of Brownian Motion.

Here are some basic properties of Brownian motion:

1. B_t is a Gaussian process, *i.e.*, for all $0 \leq t_1 \leq \cdots \leq t_k$, the random variable $Z = (B_{t_1}, \cdots, B_{t_k}) \in \mathbb{R}^{nk}$ is a multi-normal distribution.

2. B_t has independent increments, *i.e.*:

$$B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_k} - B_{t_{k-1}} \text{ are independent, } \textit{i.e.}, \mathbb{P}_{X,Y}(x,y) = \mathbb{P}_X(x)\mathbb{P}_Y(y).$$

3. $t \mapsto B_t(\omega)$ is continuous for almost all $\omega \in \Omega$.

Remark II.6.3. We only consider continuous versions of Brownian motion. ⌟

Theorem II.6.4. Kolmogrov's Continuity Theorem.

Suppose that the process $X = \{X_t\}_{t \geq 0}$ satisfies that for all $T > 0$, there exists α, β, D such that:

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq D \cdot |t-s|^{1+\beta} \text{ for } 0 \leq s, t \leq T.$$

Then there exists a continuous version of X.

For example, with Brownian motion, we have:

$$\mathbb{E}[|B_t - B_s|^4] = n(n+2)|t-s|^2,$$

then we have $\alpha = 4$, $\beta = 1$, and $D = n(n+2)$, so Brownian motion has a continuous version.

Remark II.6.5. Here, we have the Brownian motion continuous almost everywhere, *i.e.*, except for a set of probability zero, but the Kolmogrov's Continuity theorem ensures that there exists a continuous version everywhere. ⌟

Remark II.6.6. Gaussian/Markov Definition of Brownian Motion.

A real-valued stochastic process $\omega(\cdot)$ is called 1-dimensional standard **Brownian motion** if:

1. $B_0 = 0$,

2. $B_t - B_s \sim \mathcal{N}(0, t-s)$, *i.e.*, $\mathbb{P}(t-s, x)$ is normal, and

3. For any $0 < t_1 < \cdots < t_k$, we have:

$$B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_k} - B_{t_{k-1}} \text{ are independent, } \textit{i.e.}, \mathbb{P}_{X,Y}(x,y) = \mathbb{P}_X(x)\mathbb{P}_Y(y).$$

There is another definition using Martingale definition. ⌟

Then, we will talk about **filtration**.

Definition II.6.7. Filtration.

Let $B_t(\omega)$ be n-dimensional Brownian motion, then we define $\mathcal{F}_t = \mathcal{F}_t^{(n)}$ to be the σ-algebra generated by the random variables $\{B_i(s)\}_{\substack{1\le i\le n \\ 0\le s\le t}}$. ⌟

Namely, $\mathcal{F}_t$ is the smallest σ-algebra containing all the sets of the form:

$$\{\omega : B_{t_1}(\omega) \in F_1, \cdots, B_{t_k}(\omega) \in F_k\},$$

where $t_j \le t$ and all $F_i \subset \mathbb{R}^n$ are Borel sets.

Remark II.6.8.

- The **filtration** only concerns the behavior of the Brownian motion before time t, which can be interpreted as the "history of Brownian motion up to time t."
- A random function h is $\mathcal{F}_t$-measurable if and only if h can be written as the almost surely limit of sums of functions of the form $g_1(B_{t_1}), \cdots, g_k(B_{t_k})$.
- Hence, we have $h_1(\omega) = B_{t/2}(\omega)$ $\mathcal{F}_t$-measurable but $h_2(\omega) = B_{2t}(\omega)$ being not $\mathcal{F}_t$-measurable. ⌟

Definition II.6.9. Adapated Models.

Let $\{N_t\}_{t\ge 0}$ be an increasing family of σ-algebras. A process $g(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}^n$ is $\mathcal{N}_t$-adapted if for all $t > 0$, the function $\omega \mapsto g(t,\omega)$ is $\mathcal{N}_t$-measurable. ⌟

Example II.6.10. Discrete Stochastic Process in Stock Market.

Consider the model for trading in stock market, $t = 1, 2, \cdots$. At each time, the price can go up by factor u or go down by factor d.
Hence, the sample space is:

$$\Omega = \{\omega_1 = (u,u), \omega_2 = (u,d), \omega_3 = (d,u), \omega_4 = (d,d)\}.$$

Take an event $A = \{\omega_1, \omega_2\}$ means the stock goes up at $t = 1$. There, the σ-algebra generated is:

$$\mathcal{F}_1 = \{\emptyset, A, A^c, \Omega\}.$$

Note that the biggest σ-algebra is the power set, namely $\mathcal{F}_2 = \mathcal{P}(\Omega)$.
Now, consider two functions:

$$X(\omega_1) = X(\omega_2) = 1.5 \text{ and } X(\omega_3) = X(\omega_4) = 0.5, \qquad \text{with}$$
$$Y(\omega_1) = 2, Y(\omega_2) = Y(\omega_3) = 0.75, \text{ and } Y(\omega_4) = 0.25.$$

Then X is $\mathcal{F}_1$-measurable, since have the preimage of a (at most) countable image has each discrete preimage measurable.

Y is not $\mathcal{F}_1$-measurable, but it is $\mathcal{F}_2$-measurable. ⌟

Then, we consider some path properties of Brownian motion.

Proposition II.6.11. Path Properties of Brownian motion.

Let $\{B_t\}$ be a sequence of Brownian motion.

1. $\{B_t\}$ has a continuous version, so it is C^0.
2. $\{B_t\}$ is nowhere differentiable, that is, $\frac{dB_t}{dt} = \infty$ a.s., so it is not C^1.
3. $\{B_t\}$ is C^γ, where $\gamma \leq \frac{1}{2} - \epsilon$ for all $\epsilon > 0$, that is, $\mathbb{E}[|dB_t|^2] = dt$.

By Proposition II.6.11, we may consider B_t having Hölder index of $1/2$.

II.7 Conditional Expectation

First, we shall consider the conditional probability.

Definition II.7.1. Conditional Probability.

Given the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $A, B \in \mathcal{F}$ we have the probability of A given B defined as:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ for } \mathbb{P}(B) \neq 0.$$

⌟

Remark II.7.2. We say A and B are independent of $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$, and a direct consequence is:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A).$$

⌟

Then, our goal is to define the conditional expectations on two random variables.

Example II.7.3. A Case with Random Variable.

We consider a random variable such that $Y = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ (Step function), which means:

$$Y = \begin{cases} a_1 & \text{on } A_1, \\ a_2 & \text{on } A_2, \\ \vdots & \\ a_m & \text{on } A_m. \end{cases}$$

In particular, a_i's are distinct and A_i's are mutually disjoint.

Then, for any X, we may define the conditional expectation as:

$$\mathbb{E}[X \mid Y] = \begin{cases} \frac{1}{\mathbb{P}(A_1)} \int_{A_1} X d\mathbb{P} & \text{on } A_1, \\ \frac{1}{\mathbb{P}(A_2)} \int_{A_2} X d\mathbb{P} & \text{on } A_2, \\ \vdots \\ \frac{1}{\mathbb{P}(A_n)} \int_{A_n} X d\mathbb{P} & \text{on } A_n. \end{cases}$$

In fact, we have $\mathbb{E}[X \mid Y]$ is a random variable on Y, *i.e.*, it is $\mathcal{H}_Y$-measurable, meaning that there exists a measure h such that $h(Y) = \mathbb{E}[X \mid Y]$.
Now, consider any measurable $A \in \mathcal{H}_Y$ (while it can intersect any A_i's), then we have:

$$\int_A X d\mathbb{P} = \int_A \mathbb{E}[X \mid Y] d\mathbb{P}.$$

⌟

Then, we formally define the conditional expectation.

Definition II.7.4. Conditional Expectation.

The conditional expectation of X given Y is any $\mathcal{H}_{\mathcal{Y}}$-measurable random variable Z such that:

$$\int_A X d\mathbb{P} = \int_A Z d\mathbb{P} \text{ for all } A \in \mathcal{H}_Y,$$

and we denote $Z = \mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \mathcal{H}_Y]$. ⌟

Theorem II.7.5. Existence and Uniqueness of Conditional Expectation.

Let X be integrable random variable, then for each σ-algebra $\mathcal{H} \subset \mathcal{F}$, the conditional expectation $\mathbb{E}[X \mid \mathcal{H}]$ exists and is unique *up to probability zero*.

Now, we consider certain properties with conditional expectation.

Proposition II.7.6. Properties of Conditional Expectation.

Let X, Y be random variable and λ be a constant.

1. **Linearity**. $\mathbb{E}[\lambda \cdot X + Y] = \lambda \mathbb{E}[X] + \mathbb{E}[Y]$.
2. **Order**. $\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{H}]\big] = \mathbb{E}[X]$.
3. **Homogeneity**. $\mathbb{E}[YX \mid \mathcal{H}] = Y\mathbb{E}[X \mid \mathcal{H}]$ if Y is $\mathcal{H}$-measurable.
4. **Independence**. $\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[X]$ if X is independent of $\mathcal{H}$.
5. **Towering**. $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}] \mid \mathcal{G}]$ if $\mathcal{G} \subset \mathcal{H}$.

Another important property is:

Theorem II.7.7. Jensen's Inequality.

If $\Phi : \mathbb{R} \to \mathbb{R}$ is convex, and $\mathbb{E}[|\Phi(X)|] < \infty$, then:

$$\Phi\big(\mathbb{E}[X \mid \mathcal{H}]\big) \leq \mathbb{E}[\Phi(X) \mid \mathcal{H}].$$

This leads to the following consequences from the above theorem:

Corollary II.7.8. Consequences of Jensen's Inequality.

- (Cauchy Schwartz). $|E[X \mid \mathcal{H}]| \leq \mathbb{E}[|X| \mid \mathcal{H}]$ and $|\mathbb{E}[X \mid \mathcal{H}]|^2 \leq \mathbb{E}[|X|^2 \mid \mathcal{H}]$.
- (L^2 Convergence). If $X_n \xrightarrow{L^2} X$, then $\mathbb{E}[X_n \mid \mathcal{H}] \xrightarrow{L^2} \mathbb{E}[X \mid \mathcal{H}]$.

II.8 Martingale

Definition II.8.1. Discrete Martingale.

Let $\{X_j\}_{j=1}^{\infty}$ be random variables such that $\mathbb{E}[|X_j|] < \infty$. The the sequence $\{X_j\}_{j=1}^{\infty}$ is discrete martingale if $X_k = \mathbb{E}[X_j \mid X_1, \cdots, X_k] = \mathbb{E}[X_j \mid \mathcal{F}_k]$ a.s. for all $j \geq k$. ⌟

Martingale attempts to predict the future with present data. The *average prediction* of future is the present.

Remark II.8.2. Sometimes, we denote $X_1, \cdots, X_k$ in the conditional expectation as the σ-algebra generated by the sequence up to k, namely, $\sigma\big(\{X_i\}_{i=1}A^k\big) = \mathcal{N}_k$. ⌟

Definition II.8.3. Continuous Martingale.

Let $X(\cdot)$ be a real-valued stochastic process and $\mathcal{F}_t = \sigma\{X(s) : 0 \leq s \leq t\}$. If $\mathbb{E}[|X(t)|] < \infty$ and $X(s) = \mathbb{E}[X(t) \mid \mathcal{F}_s]$ for all $t \geq s \geq 0$, then $X(\cdot)$ is called Martingale. ⌟

Definition II.8.4. Uniform Integrable.

On $(X, \Omega, \mathbb{P})$, a family $\{f_j\}_{j \in \mathcal{J}}$ of real, measurable functions f_j on Ω is uniform integrable if:

$$\lim_{m \to \infty} \sup_{j \in \mathcal{J}} \left\{ \int_{|f_j| \geq m} |f_j| d\mathbb{P} \right\} = 0.$$ ⌟

Then, we consider the test function for an increasing, convex function.

Definition II.8.5. Uniformly Integrable Test Function.

A function $\psi : [0,\infty) \to [0,\infty)$ is uniformly integrable test function if ψ is increasing, convex, and $\lim_{x\to\infty} \frac{\psi(x)}{x} = \infty$. ⌟

For example we may have $\psi(x) = |x|^{1+\epsilon}$ for all $\epsilon > 0$ as a uniformly integrable test function.

Theorem II.8.6. Uniform Integrability and Test Function.

The family $\{f_j\}_{j\in\mathcal{J}}$ is uniformly integrable if and only if there exists a uniform integrable test function such that $\sup_{j\in\mathcal{J}}\{\int \psi(|f_i|)d\mathbb{P}\} < \infty$.

Hence, we have uniformly integrable as a stronger condition than just integrability.

Theorem II.8.7. Ultimate Generalization of Convergence Theorem.

Suppose $\{f_k\}_{k=1}^{\infty}$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

$$\lim_{k\to\infty} f_k(\omega) = f(\omega) \text{ for a.a. } \omega.$$

Then, the following are equivalences:

1. $\{f_n\}_{n=1}^{\infty}$ is uniformly integrable.
2. $f \in L^1(\mathbb{P})$ and $f_n \xrightarrow{L^1} f$.

Remark II.8.8. Note that uniformly integrable a.s. implies L^1 convergence, and Theorem II.4.11(3) dominated convergence theorem is a special case of the above equivalence. ⌟

Corollary II.8.9. Consequences of Ultimate Generalization.

- Let $\{M_k\}_{k=1}^{\infty}$ be a discrete martingale and assume that $\sup_k \mathbb{E}[|X_k|^p] < \infty$ for $p > 1$, then there exists $M \in L^1(\mathbb{P})$ such that $M_k \xleftrightarrow[a.s.]{L^1} M$.
- Let $X \in L^p(\mathbb{P})$, where $p \geq 1$ and $\{\mathcal{N}_k\}$ be an increasing family of σ-algebras, where $\mathcal{N}_\infty = \sigma(\{\mathcal{N}_k\}_{k=1}^{\infty})$, then:
$$M_k := \mathbb{E}[X \mid \mathcal{N}_k] \xleftrightarrow[a.s.]{L^p} M := \mathcal{E}[X \mid \mathcal{N}_\infty].$$

Here, we have uniform integrable $\{M_k\}$ if and only if $M_k = \mathbb{R}[X \mid \mathcal{F}_n]$ for some X and $\{\mathcal{F}_n\}$.

III Stochastic Integration

III.1 Itô Integral

Recall our model as:

$$\frac{dN}{dt} = \big(\gamma(t) + \text{noise}\big)N(t),$$

where we impose the generalization that:

$$dx_t = b(t, x_t)dt + \sigma(t, x_t)dw_t. \tag{fcn.1}$$

Remark III.1.1. An issue here is that dw_t is *ill-posed*, since $\{w_t\}$ is nowhere differentiable a.s. ⌟

However, we may consider the model on $[0, T]$, and we select $0 = t_0 < t_1 < \cdots < t_m = t$, and consider a discrete version, so we have:

$$dx_t = x_{k+1} - x_k,$$

where $x_i := x_{t_i}$, and thus our model becomes:

$$x_{k+1} - x_k = b(t_k, x_k)(t_{k+1} - t_k) + \sigma(t_k, x_k)(w_{t_{k+1}} - w_{t_k}). \tag{fcn.2}$$

Remark III.1.2. The selection of $b(t_k, x_k)$ and $\sigma(t_k, x_k)$ in (fcn.2) is the Itô integral, whereas replacing them with $b(t_k, x_k)$ and $\sigma(t_{(k+1/2}, x_{k+1/2})$ is the Stratonovich integral.
The Itô integral gives you a Martingale, whereas Stratonovich is more related to physics cases. ⌟

If we consider it as a sum, we have:

$$x_k = x_0 + \sum_{j=0}^{k-1} b(t_j, x_j)\Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, x_j)\Delta B_j.$$

In this case, we can define itô integral as $\Delta t \to 0$:

Definition III.1.3. Itô Integral.

For the above model of SDEs, we may write the *ill-defined* (fcn.1) in the integral form, namely as:

$$x_t = x_0 + \int_0^t b(s, x_s)ds + \int_0^t \sigma(s, x_s)dw_s.$$

⌟

Note that here, $\int_0^t \sigma(s, x_s)dw_s$ is a random variables, and x_s would contribute as a random variable.

Now, our goal is clear, we want to define:

$$\phi(t, \omega) = \sigma\big(t, x_t(\omega)\big),$$

and we want to define the integral:

$$\int_S^T \phi(t, \omega)dB_t(\omega).$$

To use a discrete version, we have:

$$\phi(t,\omega) = \sum_{i\geq 0} e_j(\omega)\cdot \mathbb{1}_{[j/2^n,(j+1)/2^n]}(t).$$

Note that we may define this over $[0,1]$, and we then can scale it into $[S,T]$ interval.

Here, we may borrow ideas from the method of separation as:

$$f(x,y) = \sum_{k=0}^{\infty} g_k(x)h_k(y).$$

Hence, we have:

$$\int_S^T \phi(t,\omega)dB_t(\omega) = \sum_{j\geq 0} e_j(\omega)(B_{t_{j-1}}, B_{t_j}).$$

Setup III.1.4. Here, we would let $S=0$ and $T=1$ for the simplicity of cases, that is:

$$t_k = \frac{k}{2^n} \text{ for } 0 \leq \frac{k}{2^n} \leq 1.$$

Otherwise, we set the value to be S on the left of 0 and T on the right of 1. ⌟

Example III.1.5. Itô and Stratonovich of Brownian Motion are Different.

We choose:

$$\phi_1(t,\omega) = \sum_{j\geq 0} B_{j/2^n}(\omega)\mathbb{1}_{[j/2^n,(j+1)/2^n]}(t).$$

Then, we have the expectation as:

$$\mathbb{E}\left[\int_0^1 \phi_1(t,\omega)dB_t(\omega)\right] = \sum_{j\geq 0}\mathbb{E}\left[B_{j/2^n}(B_{(j+1)/2^n} - B_{j/2^n})\right] = 0,$$

by independence.
On the other hand, if we choose:

$$\phi_2(t,\omega) = \sum_{j\geq 0} B_{(j+1)/2^n}(\omega)\mathbb{1}_{[j/2^n,(j+1)/2^n]}(t).$$

Then, we have the expectation as:

$$\begin{aligned}\mathbb{E}\left[\int_0^1 \phi_1(t,\omega)dB_t(\omega)\right] &= \sum_{j\geq 0}\mathbb{E}\left[B_{(j+1)/2^n}(B_{(j+1)/2^n} - B_{j/2^n})\right]\\ &= \sum_{j\geq 0}\mathbb{E}\left[(B_{(j+1)/2^n} - B_{j/2^n})(B_{(j+1)/2^n} - B_{j/2^n}) + B_{j/2^n}(B_{(j+1)/2^n} - B_{j/2^n})\right] = \sum_{j\geq 0}\Delta t_j = 1.\end{aligned}$$

Here, we can note that the results of two constructions are different. ⌟

Remark III.1.6. Location of Reference Matters.

The Itô integral selects t_j to be the left hand side, and Stratonovich selects t_j as the middle points. There results are not the same, like *Riemann* or *Lebesgue* integrals. ⌟

Setup III.1.7. Suppose $g : [0, T] \to \mathbb{R}$ is a continuous, differentiable function with $g(0) = g(1) = 0$, we define:

$$\int_0^1 g dB_t = -\int_0^1 g' B_t dt.$$

Through integration by parts, we have:

$$\int_0^1 g dB_t = g_t B_t \Big|_{t=0}^{t=1} - \int_0^1 B_t g' dt,$$

a.s. This is the Paley-Wiener-Zygmund Integral. ⌟

Proposition III.1.8. Properties with Paley-Wiener-Zygmund Integral.

Here, we consider that:

$$\mathbb{E}\left[\int_0^1 g_t dB_t\right] = 0,$$

and we have Itô isometry:

$$\mathbb{E}\left[\left(\int_0^1 g_t dB_t\right)^2\right] = \int_0^1 g^2 dt.$$

Proof. For the first expectation, we may use Fubinni as:

$$\mathbb{E}\left[\int_0^1 g_t dB_t\right] = \mathbb{E}\left[-\int_0^1 g_t' B_t dt\right] = -\int_0^1 g_t' \mathbb{E}[B_t] dt = 0.$$

For the second expectation, since g is *deterministic*, we have:

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^1 g_t dB_t\right)^2\right] &= \mathbb{E}\left[\left(\int_0^1 g_t' B_t dt\right)^2\right] = \mathbb{E}\left[\int_0^1 g_t' B_t dt \int_0^1 g_s' B_s ds\right] \\
&= \mathbb{E}\left[\int_0^1 \int_0^1 g_t' g_s' B_t B_s dt ds\right] = \int_0^1 \int_0^1 g_t' g_s' \mathbb{E}[B_t B_s] dt ds \\
&= \int_0^1 \int_0^1 g_t' g_s' \min(s,t) dt ds = \int_0^1 \left[\int_0^t g_s' s ds + \int_t^1 g_s' t ds\right] dt \\
&= \int_0^1 g_t' \left(-\int_0^t g_s ds\right) dt = \int_0^1 g_t^2 dt,
\end{aligned}$$

which completes the proof. □

Extending the definition to $g \in L^2([0,1])$, we may select a sequence of C^1 functions g_n with $g_n(0) = g_n(1) = 0$ such that:

$$\int_0^1 |g_n - g|^2 dt \to 0.$$

A specific example is the *Fourier series*.

By its convergence, it is Cauchy, we have:

$$\mathbb{E}\left[\left|\int_0^1 g_m dB_t - \int_0^1 g_m dB_t\right|^2\right] = \int_0^1 |g_m - g_n|^2 dt.$$

Hence, $\{\int_0^1 g_m dB_t\}_{m=1}^{\infty}$ is Cauchy in $L^2(\Omega, \mathbb{P})$, so we have:

$$\int_0^1 g dB_t = \lim_{n\to\infty} \int g_m dB_t \text{ in } L^2.$$

III.2 Measurability for Itô Integrals

Definition III.2.1. Filtration.

A filtration $\mathcal{F}_t$ is the σ-algebra generated by $\{B_s\}_{0\le s\le t}$. ⌟

Definition III.2.2. $\mathcal{N}_t$-adapted Process.

Let $\{\mathcal{N}_t\}_{t\ge 0}$ be an increasing family of σ-algebra. A process $g(t,\omega) : [0,\infty)\times\Omega \to \mathbb{R}$ is called $\mathcal{N}_t$-adapted if for all $t > 0$ that $\omega \mapsto g(t,\omega)$ is $\mathcal{N}_t$-measurable. ⌟

Definition III.2.3. $\mathcal{N}_t$-measurable Class.

Let $\mathcal{V} = \mathcal{V}[0,1]$ (or equivalently $\mathcal{V}[S,T]$) be the class of functions $f(t,\omega) : (0,\infty)\times\Omega \to \mathbb{R}$ that satisfied:

1. $(t,\omega) \mapsto f(t,\omega)$ is $\mathcal{B}\times\mathcal{F}$ measurable, where $\mathcal{B}$ is the Borel σ-algebra.
2. $f(t,\omega)$ is $\mathcal{F}_t$-adapted, where $\mathcal{F}_t = \sigma(\{B_s\}_{s<t})$.
3. $\mathbb{E}[\int_S^T |f(t,\omega)|^2 dt] < \infty$.

⌟

Then, we want to define $\int_0^1 f(t,\omega) dB_t(\omega) = \mathcal{I}[F](\omega)$. Assume that $f \in \mathcal{V}$ has the form:

$$f(t,\omega) = \sum_{j\ge 0} e_j(\omega) \mathbb{1}_{[t_j, t_{j+1})}(t),$$

so we have:

$$\mathcal{I}[f](\omega) = \sum_{j\ge 0} e_j(\omega)\big(B_{t_{j+1}}(\omega) - B_{t_j}(\omega)\big).$$

Corollary III.2.4. Itô Isometry.

If $\phi(t,\omega)$ is bounded and elementary, then:

$$\mathbb{E}\left[\left|\int_0^1 \phi(t,\omega)\right|^2\right] = \mathbb{E}\left[\int_0^1 |\phi(t,\omega)|^2 dt\right].$$

Proof. Here, we denote $\Delta B_j = B_{t_{j+1}} - B_{t_j}$, then

$$\mathbb{E}[e_i e_j \Delta B_i \Delta B_j] = \begin{cases} 0, & \text{when } i \neq j, \\ \mathbb{E}[e_j^2](t_{j+1} - t_j), & \text{if } i = j. \end{cases}$$

Thus, we have:

$$\begin{aligned} \mathbb{E}\left[\left|\int_0^1 \phi(t,\omega)\right|^2\right] &= \sum_{i,j} \mathbb{E}[e_j^2](t_{j+1} - t_j) = \sum_i \mathbb{E}[e_i^2](t_{i+1} - t_i) \\ &= \mathbb{E}\left[\int_0^1 |\phi(t,\omega)|^2 dt\right]. \end{aligned}$$

□

Now, we want to use the isometry to extend definition from elementary functions to functions in class $\mathcal{V}$.

Proposition III.2.5. Approximation to Continuous Class $\mathcal{V}$ Functions.

Let $g \in \mathcal{V}$ be bounded, and $g(\cdot,\omega)$ is continuous over each ω, then there exists $\phi_n \in \mathcal{V}$ such that:

$$\mathbb{E}\left[\int_0^1 (g - \phi_n)^2 dt\right] \to 0 \text{ as } n \to \infty.$$

Proof. Let $\phi_n(t,\omega) = \sum_j g(t_j,\omega)\mathbb{1}_{[t_j,t_{j+1})}(t) \in \mathcal{V}$ and:

$$\int_0^1 (g - \phi_n)^2 dt = \sum_j \int_{t_j}^{t_{j+1}} |g(t_j,\omega) - g(t,\omega)|^2 dt \to 0$$

by the continuity and bounded convergence. □

Proposition III.2.6. Approximation to Bounded Class $\mathcal{V}$ Functions.

Let $h \in \mathcal{V}$ be bounded, then there exists $g_n \in \mathcal{V}$ such that $g_n(\cdot,\omega)$ is continuous for ω and n and:

$$\mathbb{E}\left[\int_0^1 (h - g_n)^2 dt\right] \to 0 \text{ as } n \to \infty.$$

Proof. Suppose $|h(t,\omega)| \leq M$ for all (t,ω). For each h, let ψ_n be a nonnegative continuous function on $\mathbb{R}$ such that:

- $\psi_n(x) = 0$ for $x \leq -1/n$ and $x \geq 0$, and
- $\int_{-\infty}^{\infty} \psi(x)dx = 1$.

The above is called a *good kernel* in Real analysis.
Here, we define that:

$$g_n(t,\omega) = \int_0^t \psi_n(t-s)h(s,\omega)ds.$$

- $g_n(\cdot,\omega)$ is continuous for each ω a.s., and
- $|g_n(t,\omega)| \leq M$.

Since $h \in \mathcal{V}$, we can show that $g_n(t,\cdot)$ is $\mathcal{F}_t$-measurable.

- $\int_0^1 |g_n(s,\omega) - h(s,\omega)|^2 ds \to 0$ as $n \to \infty$ for each ω, we have:
- Approximation theory and boundedness that $\mathbb{E}\left[\int_0^1 |h(t,\omega) - g_n(t,\omega)|^2 dt\right] \to 0$. □

Theorem III.2.7. Approximation to Class $\mathcal{V}$ Functions.

Let $f \in \mathcal{V}$, then there exists a sequence of $\{h_n\}_{n=1}^{\infty} \subset \mathcal{V}$ such that h_n is bounded for each n and $\mathbb{E}[\int_0^1 |f - h_n|^2 dt \to 0]$ as $n \to \infty$.

Proof. We put $h_n = \begin{cases} -n, & \text{for } f < -n, \\ f(t,\omega), & \text{for } -n \leq f \leq n, \\ n & \text{for } f > n, \end{cases}$ and this function is bounded and converges. □

In this case, we can defined:

$$\int_S^T f_n(t,\omega) dB_t(\omega) \xrightarrow{L^2(\mathbb{P})} \int_S^T f(t,\omega) dB_t(\omega)$$

Remark III.2.8. We want to define for any $f \in \mathcal{V}$ of:

$$\mathcal{I}[f](\omega) = \int_0^1 f(t,\omega) dB_t(\omega) \text{ for each } f \in \mathcal{V}.$$

Our path gets from $f \in \mathcal{V}$ and bounded function, which is from $f \in \mathcal{V}$ and bounded continuous function, from $f \in \mathcal{V}$ and elementary functions, so we think about:

$$\mathbb{E}\left[\int_0^1 |\phi_n - f|^2 dt\right] to 0,$$

so we want to define the using the elementary function. ⌟

Example III.2.9. title.

Consider $B_0 = 0$, then:

$$\int_0^t B_s dB_s = \frac{B_t^2}{2} + \frac{1}{2}t.$$

- $f_s(\omega) = B_s(\omega) \in \mathcal{V}(0,t)$, and
- The Riemann integral is:

$$\int_0^t g_s dg_s = \frac{1}{2} g_t^2 \qquad \text{for } g \in C^1, \text{ and } g_0 = 0.$$

⌟

We consider:

$$\phi_n(s,\omega) = \sum B_j(\omega) \mathbb{1}_{[t_j,t_{j+1})}(s),$$

with $B_j = B_{t_j}$ and $\mathcal{F}_j = \mathcal{F}_{t_j}$-measurable. Then:

$$\begin{aligned}\mathbb{E}\left[\int_0^t (\phi_n^{(s)} B_s)^2 ds\right] &= \mathbb{E}\left[\sum_j \int_{t_j}^{t_{j+1}} (B_j - B_s)^2 ds\right] \\ &= \sum_j \int_{t_j}^{t_{j+1}} (s - t_j) ds = \frac{1}{2}\sum_j (t_{j+1} - t_j)^2 \leq |\Delta t| \sum_j (t_{j+1} - t_j) \to 0.\end{aligned}$$

So the integral is:

$$\int_0^t B_s dB_s = \lim_{\Delta t_j \to 0} \int_0^t \phi_n dB_s = \lim_{\Delta t_j \to 0} \sum_j B_j \Delta B_j \text{ in } : L^2(\mathbb{P}).$$

Now:

$$\Delta(B_j^2) = B_{j+1}^2 - B_j^2 = (B_{j+1} - B_j)^2 + 2B_j(B_{j+1} - B_j) = (\Delta B_j)^2 + 2B_j \Delta B_j.$$

Therefore, we have:

$$B_t^2 = \sum_j \Delta(B_j)^2 = \sum_j (\Delta B_j)^2 + 2B_j \Delta B_j,$$

or:

$$\sum B_j \Delta B_j = \frac{1}{2}B_t^2 - \frac{1}{2}\sum_j (\Delta B_j)^2 \to \frac{1}{2}B_t^2 - \frac{1}{2}t,$$

as we have:

$$\mathbb{E}\left[\sum_j (\Delta B_j)^2\right] = \sum_j \mathbb{E}\left[|\Delta B_j|^2\right] = \sum_j (t_{j+1} - t) = t.$$

Theorem III.2.10. Properties with Itô Integral.

Let $f, g \in \mathcal{V}(S,T)$ and $0 \leq S < U < T$, then:

1. $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$ a.s.
2. $\int_S^T (cf + g) dB_t = c\int_S^T f dB_t + \int_S^T g dB_t$, where c is a constant.
3. $\mathbb{E}\big[\int_S^T f dB_t\big] = 0$ and $\mathbb{E}\big[\big|\int_S^T f dB_t\big|\big] = \int_S^T \mathbb{E}[|f|^2] dt$, and
4. $\int_S^T f dB_t$ is $\mathcal{F}_t$-measurable.

Definition III.2.11. Martingale w.r.t. Filtration.

A **filtration** is a family $\mathcal{M} = \{M_t\}_{t\geq 0}$ of σ-algebra $M_t \subset \mathcal{F}$ such that $0 \leq s < t \implies \mathcal{M}_s \subset \mathcal{M}_t$, *i.e.*, $\{M_t\}$ is increasing. An n-dimensional stochastic process $\{M_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is a martingale with respect to a filtration $\{\mathcal{M}_t\}_{t\geq 0}$ (and with respect to $\mathbb{P}$) if:

1. M_t is $\mathcal{M}_t$-measurable for all t,
2. $\mathbb{E}[|M_t|] < \infty$ for all t, and
3. $\mathbb{E}[M_t \mid \mathcal{M}_s] = M_s$ for $t \geq s$.

⌟

Example III.2.12. Brownian Motion is Martingle w.r.t. $\mathcal{F}_t$.

Brownian motion is martingale with respect to $\mathcal{F}_t$:

1. B_t is $\mathcal{F}_t$-measurable,
2. $(\mathbb{E}[|B_t|])^2 \leq \mathbb{E}[|B_t|^2] = t < \infty$, and
3. $\mathbb{E}[B_t \mid \mathcal{F}_s] = \mathbb{E}[B_t - B_s + B_s \mid \mathcal{F}_s] = \mathbb{E}[B_t - B_s] + B_s = B_s$.

⌟

Theorem III.2.13. Doob's Martingale Inequality.

If M_t is martingale such that $t \to M_t(\omega)$ is continuous a.s., then for all $P \geq 1$, $T \geq 0$, and $\lambda > 0$, we have:

$$\mathbb{P}\left[\sup_{0\leq t\leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^p}\mathbb{E}[|M_t|^p].$$

Here, we will consider a weaker theorem to prove.

Proposition III.2.14. Discrete Doob's Martingle Inequality.

If $\{X_n\}_{n=1}^{\infty}$ is a discrete martingale, then:

$$\mathbb{P}\left\{\max_{1\leq k\leq n} X_k \geq \lambda\right\} \leq \frac{1}{\lambda}\mathbb{E}[|X_n|]\left(\text{ or } \frac{1}{\lambda^p}\mathbb{E}[|X|^p] \text{ for sub-martingale}\right), \text{ and}$$
$$\mathbb{E}\left[\max_{1\leq k\leq n} |X_k|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

Theorem III.2.15. *t*-continuous Version of Itô Integral.

Let $f \in \mathcal{V}(0,T)$, then there exists a t-continuous version of $\int_0^t f(s,\omega)dB_s(\omega)$ for $0 \leq t \leq T$, *i.e.*, there exists a t-continuous stochastic process J_t on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:

$$\mathbb{P}\left[J_t = \int_0^t f dB\right] = 1 \text{ for all } t \text{ such that } 0 \leq t \leq T.$$

Proof. Let $\phi_n = \phi(t,\omega) = \sum_j e_j^{(n)}(\omega)\mathbb{1}_{[t_j^{(n)},t_{j+1}^{(n)})}(t)$ such that:

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

Put $I_t = I_n(t,\omega) = \int_0^t \phi_n(s,\omega)dB_s(\omega)$, then$I_n(\cdot,\omega)$ is continuous. Moreover, $I_n(t,\omega)$ is a martingale with respect to $\mathcal{F}_t$ for all $s > t$:

$$\begin{aligned}
\mathbb{E}[I_s(s,\omega) \mid \mathcal{F}_t] = \mathbb{E}\left[\int_0^t \phi_n dB + \int_t^s \phi_n dB \mid \mathcal{F}_t\right] &= \int_0^t \phi_n dB_t + \mathbb{E}\left[\int_t^s \phi_n dB \mid \mathcal{F}_t\right] \\
&= \int_0^t \phi_n dB_t + \mathbb{E}\left[\sum_{t\le t_j^{(n)}\le t_{j+1}^{(n)}\le s} e_j^{(n)}\Delta B_j \mid \mathcal{F}_t\right] \\
&= \int_0^t \phi_n dB_t + \sum \mathbb{E}\,[\,\underbrace{\mathbb{E}[e_j^{(n)}\Delta B_j \mid \mathcal{F}_{t_j}^{(n)}]}_{0} \mid \mathcal{F}_t] = \int_0^t \phi_n dB_t.
\end{aligned}$$

Hence, $I_n - I_m$ is also $\mathcal{F}_t$-martingale, so by the martingale inequality, it follows that:

$$\mathbb{P}\left[\sup_{0\le t\le T} |I_n(t,\omega) - I_m(t,\omega)| > \epsilon\right] \le \frac{1}{\epsilon^2}\mathbb{E}[|I_n(T,\omega) - I_m(t,\omega)|^2] = \frac{1}{\epsilon^2}\mathbb{E}\left[\int_0^T (\phi_n - \phi_m)^2 ds\right] \to 0 \text{ as } m,n \to \infty.$$

Hence, we can choose a subsequence h_k where $k \nearrow \infty$ such that:

$$\mathbb{P}\left[\sup_{0\le t\le T} |I_{n_{k+1}}(t,\omega) - I_{n_k}(t,\omega)| > 2^{-k}\right] \le 2^{-k}.$$

Thus, by the Borel-Cantelli lemma, we have:

$$\mathbb{P}[\sup_{0\le t\le T} |I_{n_{k+1}}(t,\omega) - I_{n_k}(t,\omega)| > 2^{-k} \text{ for infinitely many } k] = 0.$$

Hence, for almost all ω, there exists $k_1(\omega)$ such that:

$$\sup_{0\le t\le T} |I_{n_{k+1}}(t,\omega) - I_{n_k}(t,\omega)| \le 2^{-k} \text{ for } k \ge k_1(\omega).$$

Therefore, $I_{n_k}(t,\omega)$ is uniform convergent for $t \in [0,T]$ for almost all ω. The limit denoted by $I_t(\omega)$ is t-continuous for almost all ω. However, we also know $I_n(t,\omega) \to I(t,\omega) = I_t$ in $L^2(\mathbb{P})$, we must have $I_t = J_t$ a.s. $\square$

Corollary III.2.16. Itô Integral is Martingale.

Let $f(t,\omega) \in \mathcal{V}(0,T)$, then $M_t(\omega) = \int_0^t f(s,\omega)dB_s$ is martingale with respect to $\mathcal{F}_t$.

III.3 Extensions of Itô Integral

Here, we first extend the class $\mathcal{V}$ to be $\mathcal{W}_{\mathcal{H}}$.

Definition III.3.1. $\mathcal{W}_{\mathcal{H}}$ Class of Processes.

$\mathcal{W}_{\mathcal{H}}$ denotes the class of processes $f(t, \omega)$ such that:

1. $(t, \omega) \mapsto f(t, \omega)$ is $\mathcal{B} \times \mathcal{F}$-measurable.
2. There exists an increasing family of σ-algebra $\mathcal{H}_t$ such that:
   - B_t is a martingale with respect to $h\mathcal{H}_t$, and
   - f_t is t-adapted.
3. $\mathbb{P}\left[\int_S^T |f(s, \omega)|^2 ds < \infty\right] = 1$.

⌟

For $f \in \mathcal{W}_{\mathcal{H}}$, we can still define:

$$\int_S^T \phi_n(t, \omega) d\mathbb{P}(\omega) \xrightarrow{\mathbb{P}} \int_S^T f(t, \omega) d\mathbb{P}(\omega).$$

Note that the convergence is not in L^2, but in probability, which is weaker. Also, with this class of functions, the integral is not necessarily a martingale.

Remark III.3.2. This definition is applied to define higher derivatives on stochastic integrals. ⌟

IV Itô Formula

IV.1 Itô Lemma

Here, we introduce the Itô lemma as a "chain rule" in stochastic setting.

Recall that:
$$\frac{1}{2}B_t^2 = \frac{1}{2}t + \int_0^t B_s dB_s.$$
Consider $f(B_t) = f \circ B_t$, we want to investigate $df(B_t)$.

Remark IV.1.1. This differs from the usual chain rule, and $df(B_t)$ can be expressed as a combination of dt and dB_t. ⌟

Definition IV.1.2. Itô Process.

Let B_t be Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, a Itô process is a stochastic integral X_t of the form:

$$X_t(\omega) = X_0(\omega) + \int_0^b u(s,\omega)ds + \int_0^t v(s,\omega)dB_s,$$

where:

1. $v \in \mathcal{W}_{\mathcal{H}}$,
2. $\mathbb{P}[\int_0^t |v(s,\omega)|^2 ds < \infty \text{ for all } t \geq 0] = 1$,
3. u is $\mathcal{H}_t$-adapted, and
4. $\mathbb{P}[\int_0^t |u(s,\omega)| ds < \infty \text{ for all } t \geq 0] = 1$. ⌟

In the differential form, we rewrite:
$$dx_t = udt + vdB_t.$$

Remark IV.1.3. We can construct for x_t on $[0, T]$ that:

$$dx_t = [B_T - t]^s udt + [B_T - t]^s vdB_t.$$
⌟

Theorem IV.1.4. Itô Lemma in 1-D.

Let X_t be a Itô process, and:
$$dW_t = udt + vdB_t.$$

Let $g(t,x) \in C^2([0,\infty) \times \mathbb{R})$, then for any $Y_{t(\omega)} = g(t, x_{t(\omega)})$, it is a Itô process and:

$$dY_t = \frac{\partial g}{\partial t}(t,x_t)dt + \frac{\partial g}{\partial x}(t,x_t)dx_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,x_t)\underbrace{(dx_t)^2}_{v^2 dt},$$

with the rules:

$$(dx_t)^2 = (dx_t)\cdot(dx_t) \text{ is computed by } dt \cdot dt = dt \cdot dB_t = 0, \text{ and } dB_t \cdot dB_t = dt.$$

Proof. Recall that:

$$dx_t = udt + vdB_t.$$

Consider Itô formula, we want to show:

$$g(t,x_t) = g(0,x_0) + \int_0^t \left(\frac{\partial g}{\partial t}(s,x_s) + u_s\frac{\partial g}{\partial x}(s,x_s) + \frac{1}{2}v_s^2 \cdot \frac{\partial^2 g}{\partial x^2}(s,x_s)\right)ds + \int_0^t v_s\frac{\partial g}{\partial x}(s,x_s)dB_t.$$

Consider that $v_s = v(s,\omega)$ and $u_s = u(s,\omega)$ are elementary processes:

$$\begin{aligned}
g(t,x_t) &= g(0,x_0) + \sum_j \Delta g(t_j,x_j) \\
&= g(0,x_0) + \sum_j \frac{\partial g}{\partial x}\Delta t_j + \sum_j \frac{\partial g}{\partial x}\Delta x_j + \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial t^2}(\Delta t_j)^2 \\
&\quad + \sum_j \frac{\partial^2 g}{\partial t \partial x}(\Delta t_j)(\Delta x_j) + \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}(\Delta x_j)^2 + \sum_j R_j.
\end{aligned}$$

If $\Delta t_j \to 0$, we have:

$$\begin{aligned}
\sum_j \frac{\partial g}{\partial t}\Delta t_j &= \sum_j \frac{\partial g}{\partial t_j}\Delta t_j \xrightarrow{\text{a.s.}} \int_0^t \frac{\partial g}{\partial t}(s,x_s)ds \\
\sum_j \frac{\partial g}{\partial x} &= \sum_j \frac{\partial g}{\partial x}(t_j,x_j)\Delta x_j \xrightarrow{L^2} \int_0^t \frac{\partial g}{\partial x}(s,x_s)dX_s.
\end{aligned}$$

Then, we get:

$$\sum_j \frac{\partial^2 g}{\partial x^2}(\Delta x_j)^2 = \sum_j \frac{\partial^2 g}{\partial x^2}\big[\underbrace{u_j^2(\Delta t_j)^2}_{(1)} + \underbrace{2u_j v_j(\Delta t_j)\Delta B_j}_{(2)} + \underbrace{v_j^2(\Delta B_j)^2}_{(3)}\big].$$

We note that for (1), we have it as:

$$u_j^2(\Delta t_j)^2 = \sup_j(\Delta t_j)\sum_j \frac{\partial^2 g}{\partial x^2}u_j^2\Delta t_j = 0.$$

For (3), we have:

$$\sum_j \frac{\partial g}{\partial x}v_j^2(\Delta B_j)^2 \xrightarrow{L^2} \int_0^t \frac{\partial^2 g}{\partial x}v^2 dx \text{ as } \Delta t_i \to 0.$$

By putting $a_j = a(t_j)$, then:

$$\mathbb{E}\left[\left(\sum_j a_j((\Delta B_j)^2 - \Delta t_j)\right)^2\right] = \sum_{i,j} \mathbb{E}\left[a_i a_j\big((\Delta B_j)^2 - \Delta t_j\big)\big((\Delta B_j)^2 - \Delta t_j\big)\right].$$

Suppose $i < j$, we have the two terms independent, so the terms vanishes since $\mathbb{E}[(\Delta B_i)^2 - \Delta t_j] = 0$. If $i > j$, we have:

$$\begin{aligned}\sum_j \mathbb{E}[a_j^2((\Delta B_j)^2 - \Delta t_j)^2] &= \sum_j \mathbb{E}[a_j^2]\mathbb{E}[(\Delta B_j)^4 - 2(\Delta_j)^2\Delta t_j + (\Delta t_j)^2] \\ &= \sum_j \mathbb{E}[a_j^2] \cdot (3(\Delta t_j)^2 - 2(\Delta t_j)^2 + (\Delta t_j)^2) = 2\sum_j \mathbb{E}[a_j^2] \cdot (\Delta t_j)^2 \to 0.\end{aligned}$$

Hence, we have:

$$\sum_j a_j(\Delta B_j)^2 \to \int_0^t a(s)ds \text{ in } L^2(\mathbb{P}).$$

□

Example IV.1.5. Worked Example of Evaluting Itô Integral, I.

Consider $I_t = \int_0^t B_s dB_s$, we choose $x_t = B_t$ and $g(t,x) = \frac{1}{2}x^2$. Then for:

$$Y_t = g(t, B_t) = \frac{1}{2}B_t^2,$$

by applying the Itô lemma:

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial^2 x}(dx_t)^2 = B_t dB_t + \frac{1}{2}(dB_t)^2 = B_t dB_t + \frac{1}{2}dt.$$

⌟

Example IV.1.6. Worked Example of Evaluting Itô Integral, II.

Consider $I_t = \int_0^t s dB_s$, we let $g(t,x) = tx$ and $Y_t = g(t, B_t) = tB_t$. Then by Itô lemma:

$$dY_t = B_t dt + t dB_t + 0 = B_t dt + t dB_t.$$

Hence, in the integral form:

$$tB_t = \int_0^t B_s ds + \underbrace{\int_0^t s dB_s}_{I_t}.$$

Therefore, we have:

$$I_t = tB_t - \int_0^t B_s ds.$$

⌟

Theorem IV.1.7. Integration by Parts.

Suppose $f(t,\omega)$ is continuous and of bounded variation with respect to $s \in [0,t]$ for almost all ω. Then:

$$\int_0^t f(s)dB_s = f(t)B_t - \int_0^t B_s df_s,$$

or equivalently:

$$I_t = \int_0^t B_s df_s = f(t)B_t - \int_0^t f(s)dB_s.$$

IV.2 Multidimensional Itô Formula

Then, our next step is to Itô formula for multi-dimensions. We let $B(t,\omega) = \big(B_1(t,\omega), \cdots, B_m(t,\omega)\big)$ denote m-dimensional (coordinately i.i.d.) Brownian motion, We can form the following Itô process:

$$\begin{aligned}
dX_1 &= u_1 dt + v_{1,1} dB_1 + \cdots + v_{1,m} dB_m. \\
dX_2 &= u_2 dt + v_{2,1} dB_1 + \cdots + v_{2,m} dB_m. \\
&\vdots \\
dX_n &= u_n dt + v_{n,1} dB_1 + \cdots + v_{n,m} dB_m.
\end{aligned}$$

Here, we have $\{u_i\}_{i=1}^n$ and $\{v_{i,j}\}_{i,j=1}^{n,m}$.

We note that the first order Itô formula does not apply to this process.

In a matrix notation, we have:

$$dX(t) = udt + vdB(t),$$

where $X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \end{pmatrix}$, $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$, $v = \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,m} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n,1} & v_{n,2} & \cdots & v_{n,m} \end{pmatrix}$, and $dB(t) = \begin{pmatrix} dB_1(t) \\ dB_2(t) \\ \vdots \\ dB_m(t) \end{pmatrix}$.

Theorem IV.2.1. Itô Formula for Higher Dimensions.

Let $X(t)$ be the n-dimensional Itô process as above. Let:

$$g(t,x) = \big(g_1(t,x), \cdots, g_p(t,x)\big) \in C^2 \text{ on } [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^m,$$

Then the process $Y(t,\omega) = g\big(t, X(t)\big)$ satisfies that:

$$dY_k = \frac{\partial g_k}{\partial t}(t,x)dt + \frac{\partial g_k}{\partial x_i}(t,x)dX_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,x) dX_i dX_j \text{ for } 1 \le k \le p.$$

Here, we follow the rule: $dB_i dB_j = \delta_{i,j} dt$ and $d_t dB_i = dB_i dt = 0$.

Remark IV.2.2. When $m = n = 1$, this is the 1-dimensional Itô formula. In particular:

$$(dX_i)^2 = (u_i dt + v_{i,1} dB_1 + \cdots + v_{1,m} dB_m)^2 = v_{i,1}^2 dt + v_{i,2}^2 dt + \cdots + v_{1,m}^2 dt.$$

⌟

For the $dB_i dB_j = 0$ part, we formally have for $i \neq j$:

$$\mathbb{E}[dB_i dB_j] = \mathbb{E}[(B_i(t) - B_i(t-\Delta t))(B_j(t) - B_j(t-\Delta))] = \mathbb{E}[(B_i(t) - B_i(t-\Delta t))]\mathbb{E}[(B_j(t) - B_j(t-\Delta))] = 0.$$

For the case in which $i = j$:

$$\mathbb{E}[dB_i dB_i] = \mathbb{E}[(B_i(t) - B_i(t-\Delta t))^2] = \Delta t.$$

Example IV.2.3. *n*-dimensional Bessel Process.

Let $B = (B_1, \cdots, B_n)$ be standard n-dimensional Brownian motion with $n \geq 2$ and consider:

$$R(t,\omega) = |B(t,\omega)| = \sqrt{B_1^2(t,\omega) + \cdots + B_n^2(t,\omega)},$$

i.e., $R(t,\omega)$ measures the distance of the Brownian motion to the origin. We consider the function $g(t,x) = |x| = \sqrt{x_1^2 + \cdots + x_n^2}$.
By applying the multi-dimensional Itô formula, we find:

$$\frac{\partial g}{\partial x_i} = \frac{1}{2\sqrt{x_1^2 + \cdots + x_n^2}} 2x_i = \frac{x_i}{\sqrt{x_1^2 + \cdots + x_n^2}} = \frac{x_i}{R}.$$

Then, for the second partials, we have:

$$\frac{\partial^2 g}{\partial x_i^2} = \frac{1}{\sqrt{x_1^2 + \cdots + x_n^2}} - \frac{x_i^2}{(x_1^2 + \cdots + x_n^2)^{3/2}} = \frac{R^2 - x_i^2}{R^3}.$$

Note that if $i \neq j$, the differential form is zero, so we have:

$$dR = \sum_{i=1}^{n} \frac{B_i dB_i}{R} + \frac{1}{2}\sum_{i=1}^{n} \frac{R^2 - x_i^2}{R^3} dt = \sum_{i=1}^{n} \frac{B_i dB_i}{R} + \frac{n-1}{2R} dt.$$

⌟

Note that the function is not differentiable, the function is not differentiable at $x = 0$, but $B_t = 0$ has probability 0.

Example IV.2.4. Tanaka's Formula and Local Time.

We try to apply Itô formula to:

$$g(B_t) = |B_t| \text{ with } g(x) = |x|.$$

In this case, we note that the graph is:

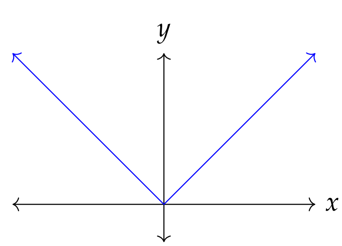

Figure IV.1. Graph of $y = |x|$.

First, we consider the derivative:

$$g'(x) = \operatorname{sgn}(x) = \begin{cases} 1, & \text{when } x \geq 0, \\ -1, & \text{when } x < 0. \end{cases}$$

The second derivative is:

$$g''(x) = \delta_0(x).$$

In this case, g is not C^2 at 0, and we have:

$$\begin{aligned} |B_t| = g(B_t) &= \int_0^t g'(B_s)ds + \frac{1}{2}\int_0^t g''(B_s)ds \\ &= \int_0^t \operatorname{sgn}(B_s)dB_s + \frac{1}{2}\int_0^t \delta_0(B_s)ds. \end{aligned}$$

Alternatively, we defined g_ϵ for $\epsilon > 0$ near zero:

$$g_\epsilon(x) = \begin{cases} |x|, & \text{when } |x| \geq \epsilon, \\ \frac{1}{2}\left(\epsilon + \frac{x^2}{\epsilon}\right), & \text{when } |x| < \epsilon. \end{cases}$$

We may note that $g_\epsilon \to g(x)$ as $\epsilon \to 0$.
Then, we consider $Y_t^{(\epsilon)} = g_\epsilon(X_t)$, and by the Itô formula, we get:

$$dY_t^{(\epsilon)} = g'_\epsilon(B_t)dB_t + \frac{1}{2}g''_\epsilon(B_t)dt.$$

We note that:

$$g'_\epsilon = \begin{cases} 1, & \text{when } x \geq \epsilon, \\ \frac{x}{\epsilon}, & \text{when } -\epsilon < x < \epsilon, \\ -1, & \text{when } x \leq -\epsilon. \end{cases}$$

Then, the second derivative is:

$$g''_\epsilon(x) = \begin{cases} 0, & \text{when } |x| \geq \epsilon, \\ \frac{1}{\epsilon}, & \text{when } |x| < \epsilon, \end{cases} = \frac{1}{\epsilon}\mathbb{1}_{|x|<\epsilon}.$$

Then, we have:

$$\begin{aligned} Y_t^{())} = g_\epsilon(B_t) = g_\epsilon(B_0) &+ \int_0^t g'_\epsilon(B_s)dB_s + \frac{1}{2\epsilon}\int_0^t \mathbb{1}_{|B_s|<\epsilon}ds \\ = g_\epsilon(B_0) &+ \int_0^t g'_\epsilon(B_s)dB_s + \frac{1}{2\epsilon}|\{s \in [0,t] : |B_s| < \epsilon\}|. \end{aligned}$$

Note that the last term measures how long the Brownian motion stays on the ϵ-neighborhood of 0, and the division makes it the density.

Then, we use Itô isometry to get that:

$$\begin{aligned}\mathbb{E}\left[\left|\int_0^t g_\epsilon'(B_s)\mathbb{1}_{|B_s|<\epsilon}dB_s\right|^2\right] &= \mathbb{E}\left[\int_0^t \left|\frac{B_s}{\epsilon}\right|^2 \mathbb{1}_{|B_s|<\epsilon}ds\right] \\ &\leq \mathbb{E}\left[\int_0^t \mathbb{1}_{|B_s|<\epsilon}\right]ds = \int_0^t \mathbb{P}[|B_s|<\epsilon]ds \xrightarrow{\epsilon\to 0} 0.\end{aligned}$$

Therefore, we have $\int_0^t g_\epsilon'(B_s)\mathbb{1}_{|B_s|<\epsilon}dB_s$ converges to 0 in the L^2 sense. Therefore, we can reduce our formula into:

$$\begin{aligned}Y_t^{(\epsilon)} &= g_\epsilon(B_0) + \int_0^t \operatorname{sgn}(B_s)\mathbb{1}_{|B_s|\geq\epsilon}dB_s + \frac{1}{2\epsilon}|\{s\in[0,t]:|B_s|<\epsilon\}| \\ &\xrightarrow{\epsilon\to 0} g(B_0) + \int_0^t \operatorname{sgn}(B_s)ds + \underbrace{\lim_{\epsilon\to 0}\frac{1}{2\epsilon}|\{s\in[0,t]:|B_s|<\epsilon\}|}_{L_t(\omega)}.\end{aligned}$$

Hence, we have the Tanaka formula as:

$$|B_t| = \int_0^t \operatorname{sgn}(B_s)dB_s + L_t,$$

and L_t is the local time of Brownian motion at 0. ⌟

Note that when we have $g(x) = |x-a|$ for $a\in\mathbb{R}$, then we shall have L_t as the local time of Brownian motion at a.

IV.3 Martingale Representation Theory

The idea is that we have the Itô integral as:

$$X_t = X_0 + \int_0^t v(s,\omega)dB(s)$$

in n-dimension is martingale with respect to the filtration $\mathcal{F}_t^{(n)}$.

Given a martingale $\{M_t\}_{t\geq 0}$, can we have:

$$M_t = \mathbb{E}[M_0] + \int_0^t f(s,\omega)dB(s)?$$

Proposition IV.3.1. Step Random Variable is Dense.

Fix $T>0$, the set of random variables:

$$\{\phi(B_{t_1},\cdots,B_{t_n}) : t_i\in[0,T], \phi\in C_0^\infty(\mathbb{R}^n), n=1,2,\cdots\}$$

is dense in $L^2(\mathcal{F}_t,\mathbb{P})$.

Proof. Doob-Dynkin Formula (Proposition II.2.3). □

Proposition IV.3.2. Linear Span of Class of Functions is Dense.

The linear span of the random variables of the type:

$$\exp\left[\int_0^T h(t)dB_t(\omega) - \frac{1}{2}\int_0^T h^2(t)dt\right], \qquad h \in L^2([0,T])$$

is dense in $L^2(\mathcal{F}, \mathbb{P})$.

Then, we introduce the main theorem.

Theorem IV.3.3. Martingale Representation Theorem.

Let $B(t) = \big(B_1(t), B_2(t), \cdots, B_n(t)\big)$ be n-dimensional Brownian motion. Suppose M_t is an $\mathcal{F}_t^{(n)}$-martingale and $M_t \in L^2(\mathbb{P})$ for all $t \geq 0$, then there exists a unique stochastic process $g(t,\omega)$ such that $g \in \mathcal{V}^{(n)}(0,t)$, and:

$$M_t(\omega) = \mathbb{E}[M_0] + \int_0^t g(s,\omega)dB(s,\omega) \text{ a.s. for all } t \geq 0.$$

The above theorem is a consequence of the following.

Theorem IV.3.4. Itô Representation Theorem.

Let $F \in L^2(\mathcal{F}_T^{(n)}, \mathbb{P})$, then there exists a unique stochastic process $f(t,\omega) \in \mathcal{V}^{(n)}(0,T)$ such that:

$$F(\omega) = \mathbb{E}[F] + \int_0^T f(t,\omega)dB(t).$$

Remark IV.3.5. Iterative Itô Representation Theorem.

Consider we apply Itô representation theorem multiple times:

$$\begin{aligned} F(T,\omega) &= \mathbb{E}[F] + \int_0^T \mathbb{E}[f] + \int_0^t g(s,\omega)dB(s)dB(t) \\ &= \mathbb{E}[F] + \int_0^T \mathbb{E}[f]dB_s + \iint_{0<s<t<T} g(s,\omega)dB(s)dB(t) \\ &= \sum_{n=0}^{\infty} C_n I_n(T,\omega), \end{aligned}$$

which is called the Itô-Wiener expansion. ⌟

Proof of Theorem IV.3.4. Without loss of generality, let $n = 1$. First, we assume:

$$F(\omega) = \exp\left[\int_0^T h(t)dB_t(\omega) - \frac{1}{2}\int_0^T h^2(t)dt\right] \text{ for some } h \in L^2([0,T]).$$

We define:

$$Y_t(\omega) = \exp\left[\int_0^t h(s)dB_s(\omega) - \frac{1}{2}\int_0^t h^2(s)ds\right] \text{ for } 0 \le t \le T.$$

By the Itô formula, we have:

$$dY_t = Y_t\left[\left(h(t)dB_t - \frac{1}{2}h^2(t)\right)dt + \frac{1}{2}Y_t\big(h(t)dB_t\big)^2\right] = Y_t h(t)dB_t.$$

Hence, it is equivalently:

$$Y_t = 1 + \int_0^t Y_s h(s)dB_s \text{ and } F_T = 1 + \int_0^T Y_s h(s)dB_s.$$

Second, we assume if $F \in L^2(\mathcal{F}_T, \mathbb{P})$, then there exists unique F_n in the exponential-martingale form such that $F_n \to F$ in $L^2(\mathcal{F}_T, \mathbb{P})$ sense. We have:

$$F_n(\omega) = \mathbb{E}[F_n] + \int_0^T f_n(s,\omega)dB_s(\omega), \text{ with } f_n \in \mathcal{V}([0,T]).$$

Then, by the Itô isometry, we have:

$$\begin{aligned}\mathbb{E}[|F_n - F_m|^2] &= (\mathbb{E}[|F_n - F_m|])^2 + \int_0^T \mathbb{E}[|f_n - f_m|^2]dt \\ &\le \mathbb{E}[|F_n - F_m|^2] + \int_0^T \mathbb{E}[|f_n - f_m|^2]dt \to 0 \text{ as } n,m \to \infty.\end{aligned}$$

Hence, we have $f_n \to f$ in L^2 to $f \in \mathcal{V}[0,T]$ by completeness, so:

$$F = \lim_{n\to\infty} F_n = \lim_{n\to\infty}\left(\mathbb{E}[F_n] + \int_0^T f_n dB\right) = \mathbb{E}[F] + \int_0^T f dB.$$

Hence, we have prove the existence, and we shall now think about uniqueness.
Consider Itô Isometry, there exists f_1, f_2 such that:

$$F(\omega) = \mathbb{E}[F] + \int_0^T f_1(t,\omega)dB_t = \mathbb{E}[F] + \int_0^T f_2(t,\omega)dB_t,$$

and hence:

$$0 = \mathbb{E}\left[\left|\int_0^T \big(f_1(t,\omega) - f_2(t,\omega)\big)dB_t\right|^2\right] = \mathbb{E}\left[|f_1(t,\omega) - f_2(t,\omega)|^2\right]dt.$$

Hence, we have $f_1(t,\omega) = f_2(t,\omega)$ almost anywhere for $(t,\omega) \in [0,T] \times \Omega$. □

End of March 5, 2025