AS.110.653 : Stochastic Differential Equations

# Notebook

James Guo

April 28, 2025

## Contents

*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,
- $\sigma$ represents the **diffusion** coefficient, and
- $\xi$ 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 $\xi$?
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 $\xi$ 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 $\rho$ (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 $\Delta t$.

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 ($\Delta t$ and $\Delta B_t$) centered at $\Delta t = 0$ and $\Delta B_t = 0$, yielding that:

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

where $\partial_i$ means the partial derivative with respect to the $i$-th variable.

**Remark I.2.1.** Since we are dividing $\Delta t$ on both sides, while $\Delta t \to 0$, all the terms with order greater than 1 of $\Delta t$ 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) $\mu$ and variance $\sigma^2$. 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:

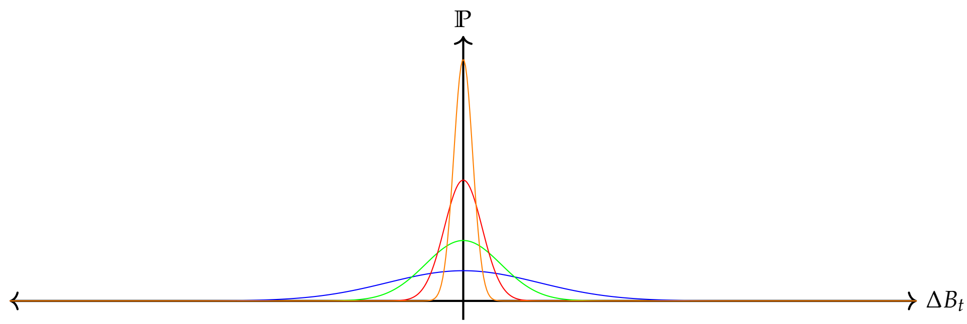


*Figure I.1. Distribution of $\Delta B_t \sim \mathcal{N}(0, \Delta t)$ with $\Delta t = 1$ (blue), $\sqrt{2}/2$ (green), $1/2$ (red), $\sqrt{2}/4$ (orange).*

**Proposition I.2.3. Taylor Expansion of SDEs.**

We consider the Taylor expansion of the discrete motion as:

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

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

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.

   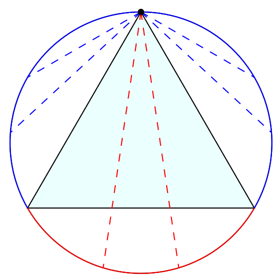

   *Figure II.1. Fixing an endpoint on the circle method.*

   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.

   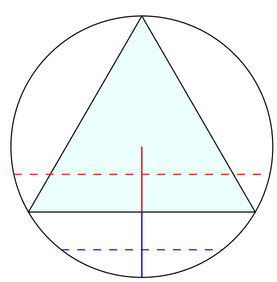

   *Figure II.2. Fixing a radius on the circle method.*

   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.

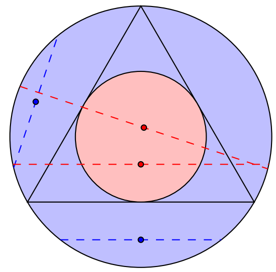

*Figure II.3. Classifying the midpoint method.*

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. $\sigma$-Algebra.**

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

1. $\varnothing \in \mathcal{F}$,
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 $\sigma$-algebra together with a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ forms a **probability measure space**, while $\mathbb{P}$ satisfies that:

1. $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\Omega) = 1$,
2. ($\sigma$-additivity): For any $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ such that they are mutually disjoint, *i.e.*, $A_i \cap A_j = \varnothing$ 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 $\sigma$-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 $\sigma$-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 $\sigma$-algebra including all sets with outer measure zero. ⌟

**Definition II.1.7. Smallest $\sigma$-algebra.**

Given any family $\mathcal{U}$ of subsets of $\Omega$, there is a smallest $\sigma$-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 $\sigma$-algebra** on $\Omega$, 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. $\sigma$-algebra Generated by a Function.**

Let $X : \Omega \to \mathbb{R}^n$ be any function, then the $\sigma$-algebra generated by $X$ is smallest $\sigma$-algebra on $\Omega$ 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 $\sigma$-algebra on $\mathbb{R}^n$. Clearly, $\mathcal{H}_X$ is $\mathcal{H}_X$-measurable, and $\mathcal{H}_X$ smallest $\sigma$-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 $\mu_X$ 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 ($\rho_X$) and the cumulative density ($\mu_X$): We can think of our Gaussian curve of $\Delta B_t \sim \mathcal{N}(0, \Delta t)$ as:

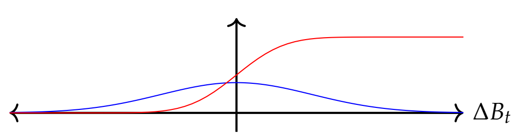


*Figure II.4. Probability density function (blue) and cumulative density function (red) of $\mathcal{N}(0,1)$.*

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 $\sigma$-algebra $\mathcal{F}$ will contain the $\sigma$-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 $\Omega$, 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 $(E, \mathcal{B}(E))$, where $E$ is a topology and $\mathcal{B}(E)$ is the $\sigma$-algebra over $E$. ⌟

**Definition II.4.2. Weak Convergence.**

Let $\{\mu_n\}_{n\in\mathbb{N}^+}$ be a sequence of finite measures on $(E, \mathcal{B}(E))$, it **converges weakly** to $\mu$ 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 $\mu_n$ is a distribution of $X_n$ and $\mu$ 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 $\mu_n$ having the distribution $\mathcal{N}(0,1)$.
- $X_n \xrightarrow{\mathbb{P}} n$ is **not true**. Suppose for all $\delta$ 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 $\sigma^2$ 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:

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

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:

$$\phi_X(u_1, u_2, \cdots, u_n) = \mathbb{E}\left[\exp\{\mathrm{i}(u_1x_1 + \cdots + u_nx_n)\}\right] = \int_{\mathbb{R}^n} e^{\mathrm{i}\langle u,x\rangle} \underbrace{\mathbb{P}(x \in dX)}_{\rho_X(x)dx \text{ if the density exists}}.$$ ⌟

**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.**

$\phi_X$ 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:

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

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

Let $X_j : \Omega \to \mathbb{R}$ be random variables for $1 \le j \le n$, then $X = (X_1, \cdots, X_n)$ is normal if and only if $Y = \lambda_1X_1 + \cdots + \lambda_nX_n$ for all $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$.

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

$$\mathbb{E}\left[\exp\left(\mathrm{i}u\sum_{j=1}^{n}\lambda_jX_j\right)\right] = \exp\left[-\frac{1}{2}\sum_{j,k}u\lambda_jc_{j,k}u\lambda_k + \mathrm{i}\sum_{j}u\lambda_jm_j\right] = \exp\left[-\frac{u^2}{2}\sum_{j,k}\lambda_jc_{j,k}\lambda_k + \mathrm{i}u\sum_{j}\lambda_jm_j\right].$$

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

($\Longleftarrow$:) 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. $\square$

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

Let $Y_0, Y_1, \cdots, Y_n$ be real random variables on $\Omega$. 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 $C = \left[x_{j,n}\right] = \lim_{k\to\infty} C_k$. $\square$

**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.

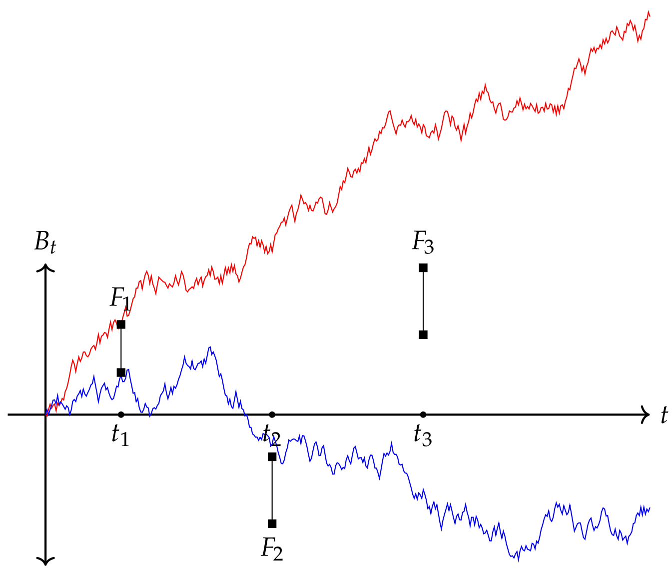


*Figure II.5. Illustration of Brownian Motion in 1D.*

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 $\alpha, \beta, D$ 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 $\sigma$-algebra generated by the random variables $\{B_i(s)\}_{\substack{1 \leq i \leq n \\ 0 \leq s \leq t}}$. ⌟

Namely, $\mathcal{F}_t$ is the smallest $\sigma$-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 $t_j \leq t$ 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 $\{N_t\}_{t \geq 0}$ be an increasing family of $\sigma$-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 $\sigma$-algebra generated is:

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

*Note that the biggest $\sigma$-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 $C^\gamma$, 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 $\sigma$-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 $\lambda$ 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 $\sigma$-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 $\Omega$ 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 $\psi$ 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 $\sigma$-algebras, where $\mathcal{N}_\infty = \sigma\big(\{\mathcal{N}_k\}_{k=1}^{\infty}\big)$, 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:

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

**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:

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

**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 $\sigma$-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 $\sigma$-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 $\sigma$-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:

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

□

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 $\omega$, 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 $\omega$ 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 $(t,\omega)$. For each $h$, let $\psi_n$ 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 $\omega$ 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 $\omega$, 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:

$$\mathbb{E}\left[\int_0^1 |\phi_n - f|^2 dt\right] to0,$$

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:

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

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 $0 \le S < U < T$, 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. $\mathbb{E}\left[\int_S^T f dB_t\right] = 0$ and $\mathbb{E}\left[\left|\int_S^T f dB_t\right|\right] = \int_S^T \mathbb{E}[|f|^2] dt$, 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 $\mathcal{M} = \{M_t\}_{t\ge 0}$ of $\sigma$-algebra $M_t \subset \mathcal{F}$ such that $0 \le s < t \implies \mathcal{M}_s \subset \mathcal{M}_t$, *i.e.*, $\{M_t\}$ is increasing. An $n$-dimensional stochastic process $\{M_t\}_{t\ge 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is a martingale with respect to a filtration $\{\mathcal{M}_t\}_{t\ge 0}$ (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 $s > t$:

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

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 $\omega$, 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 $\omega$. The limit denoted by $I_t(\omega)$ is $t$-continuous for almost all $\omega$. 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. □

**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 $\sigma$-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:

$$dX_t = udt + vdB_t.$$

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:

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

abiding to 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:

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

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:

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

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:

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

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 $\omega$. 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:

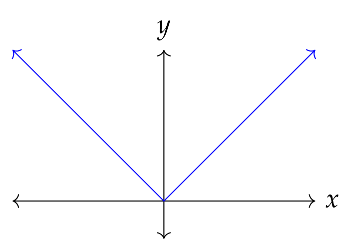


*Figure IV.1. Graph of $y = |x|$.*

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 $g_\epsilon$ 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 $\epsilon$-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:

$$\{\phi(B_{t_1},\cdots,B_{t_n}) : t_i\in[0,T], \phi\in\mathbb{C}_0^\infty(\mathbb{R}^n), n=1,2,\cdots\}$$

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 chaos 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:

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

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 $(t,\omega) \in [0,T] \times \Omega$. □

Then, we can use Itô representation theorem to prove the Martingale representation theorem.

*Proof of Theorem IV.3.3.* Without loss of generality, we assume $n = 1$. By the Itô representation theorem, we have that for all $t$, there exists a unique $f^{(t)}(s,\omega) \in L^2(\mathcal{F}, \mathbb{P})$ such that:

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

Now, assume $0 \le t_1 \le t_2$, then:

$$\begin{aligned}M_{t_1} &= \mathbb{E}[M_{t_2} \mid \mathcal{F}_{t_1}] = \mathbb{E}[M_0] + \mathbb{E}\left[ \int_0^{t_2} f^{(t_2)}(s,\omega) dB_s(\omega) \mid \mathcal{F}_{t_1} \right] \\ &= \mathbb{E}[M_0] + \int_0^{t_1} f^{(t_2)}(s,\omega) dB_s(\omega),\end{aligned}$$

by considering it as:

$$F(t_2, \widetilde{t_2}) = \int_0^t f^{(\widetilde{t_2})}(s,\omega) dB_s \text{ for any fixed } \widetilde{t_2} > 0,$$

so that, $F(t_2, \widetilde{t_2})$ is martingale if $t_2 < \widetilde{t_2}$.
Recall that:

$$M_{t_1} = \mathbb{E}[M_0] + \int_0^{t_1} f^{(t_1)}(s,\omega)dB_s(\omega).$$

So we must have that:

$$\mathbb{E}\left[\left(\int_0^{t_1} (f^{(t_2)} - f^{(t_1)})dB\right)^2\right] = \int_0^{t_1} \mathbb{E}[(f^{(t_2)} - f^{(t_1)})^2]ds = 0,$$

and therefore:

$$f^{(t_1)}(s,\omega) = f^{(t_2)}(s,\omega) \text{ for a.a. } (s,\omega) \in [0,t_1] \times \Omega.$$

Hence, we can define $f(s,\omega)$ for a.a. $s \in [0,\infty) \times \Omega$ by setting:

$$f(s,\omega) = f^{(N)}(s,\omega) \text{ if } s \in [0,N],$$

and then we have:

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

□

# V Stochastic Differential Equations

## V.1 Introduction and Examples

Recall that:

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\frac{dB_t}{dt},$$

in which $\frac{dB_t}{dt}$ is not differentiable, so we have written it in differential form:

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

and we have that:

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s.$$

**Remark V.1.1.** Recall that for Itô process, we have:

$$dX_t = udt + vdB_t,$$

in which the definition is justified. ⌟

Now, this shed the following questions:

- Can we obtain existence and uniqueness of the solution?
- Can we solve for the solution?

Here, we give a counter example to existence and uniqueness.

**Example V.1.2. Non-existence Solutions.**

Consider the ODE:

$$\frac{dX_t}{dt} = X_t^2, \qquad X_0 = 1.$$

We see the solution as:

$$X_t = \frac{1}{1-t}, \text{ which is not global.}$$ ⌟

The above example corresponds to the SDE that:

$$b(t, x) = x^2, \qquad \sigma(t, x) = 0.$$

**Example V.1.3. Non-unique Solution.**

Consider the ODE:

$$\frac{dX_t}{dt} = 3X_t^{2/3}, \qquad X_0 = 0.$$

Here, we can construct the solution as:

$$X_t = \begin{cases} 0, & \text{when } t \leq a, \\ (t-a)^3, & \text{when } t > a, \end{cases} \qquad \text{for all } a \geq 0.$$

Since the choice of $a$ is arbitrary, the solution is not unique. ⌟

The above example also corresponds to the SDE that:

$$b(t,x) = x^{2/3}, \qquad \sigma(t,x) = 0.$$

Hence, we need to think about some further conditions to ensure existence and uniqueness. Then, we see some examples of solving the SDEs.

**Proposition V.1.4. Product Rule.**

We consider the **product rule** for the product of random variables, namely:

$$d(X_tY_t) = X_tdY_t + Y_tdX_t + dX_tdY_t.$$

**Example V.1.5. Geometric Brownian Motion / Population Growth.**

Here, we pose that:

$$dN_t = rN_tdt + \alpha N_tdB_t,$$

which can be transformed into:

$$\frac{dN_t}{N_t} = rdt + \alpha dB_t.$$

So, the integral will be:

$$\int_0^t \frac{dN_s}{N_s} = rt + \alpha B_t \text{ where } B_0 = 0.$$

To evaluate the integral on the left hind side we let:

$$g(t,x) = \log(x),$$

so we have the Itô formula that:

$$d\log(N_t) = \frac{1}{N_t}dN_t + \frac{1}{2}\left(-\frac{1}{N_t^2}\right)(dN_t)^2 = \frac{dN_t}{N_t} - \frac{\alpha^2}{2}dt.$$

Thus, we equivalently have:

$$\log(N_t) - \log(N_0) = \int_0^t \frac{dN_s}{N_s} - \frac{\alpha^2}{2}t = rt + \alpha B_t - \frac{\alpha^2}{2}t = \left(r - \frac{\alpha^2}{2}\right)t + \alpha B_t.$$

Therefore, we have:

$$N_t = N_0 \exp\left[\left(r - \frac{\alpha^2}{2}\right)t + \alpha B_t\right],$$

and we have existence for this example. ⌟

**Remark V.1.6.**

- $\mathbb{E}[N_t] = \mathbb{E}[N_0]\mathbb{E}\left[\exp\left[\left(r - \frac{\alpha^2}{2} + \alpha B_t\right)\right]\right] = \mathbb{E}[N_0]e^{rt}.$

- (The Law of Iterated Logarithm). Now, we have:

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log\log t}} = 1 \text{ a.s.}$$

- If $r > \frac{1}{2}\alpha^2$, then $N_t \to \infty$ a.s. as $t \nearrow \infty$.
  If $r < \frac{1}{2}\alpha^2$, then $N_t \to 0$ a.s. as $t \nearrow \infty$.
  If $r = \frac{1}{2}\alpha^2$, then $N_t$ fluctuates.

⌟

**Example V.1.7. Brownian Bridge.**

Consider the solution of the following SDE:

$$\begin{cases} dX_t = -\dfrac{X_t}{1-t}dt + dB_t & \text{for } 0 \le t < 1, \\ X_0 = 0. \end{cases}$$

Here, we claim that the solution is:

$$X_t = \begin{cases} (1-t)\int_0^t \frac{1}{1-s} dB_s, & \text{for } 0 \le t < 1, \\ 0, & \text{when } t = 1. \end{cases}$$

We verify by Itô formula:

$$dX_t = -\int_0^t \frac{1}{1-s} dB_s dt + (1-t)\frac{1}{1-t} dB_t = -\frac{X_t}{1-t} dt + dB_t.$$

Note that the Brownian motion has the two ends as zero, so it is fixed like a bridge. ⌟

**Remark V.1.8.** As $t \nearrow 1$, we have $X_t \to 0$ a.s. and $\mathbb{E}[|X_t|^2] \to 0$. ⌟

**Example V.1.9. Langevin's Equation / Ornstein-Uhlenback Process.**

Consider the differential equation:

$$\begin{cases} dX_t = -bX_t dt + \sigma dB_t, \\ X_0 = x. \end{cases}$$

The solution is:

$$X_t = e^{-bt}x + \sigma \int_0^t e^{-b(t-s)} dB_s.$$

For the solution, we notice the following:

1. For $\mathbb{E}[X_t]$, we consider:

$$\mathbb{E}[X_t] = \mathbb{E}[e^{-bt}x] + \sigma\mathbb{E}\left[\int_0^t e^{-b(t-s)} dB_s\right] = xe^{-bt},$$

   which approaches 0 as $t \to \infty$.

2. For $\mathbb{E}[X_t^2]$, we consider:

$$\begin{aligned}\mathbb{E}[X_t^2] &= \mathbb{E}\left[e^{-2bt}x^2 + 2e^{-bt}x\sigma\int_0^t e^{-b(t-s)}dB_s + \sigma^2\left(\int_0^t e^{-b(t-s)}dB_s\right)^2\right] \\ &= \mathbb{E}[x^2]e^{-2bt} + \sigma^2\int_0^t e^{-2b(t-s)}ds \\ &= \mathbb{E}[x^2]e^{-2bt} + \frac{\sigma^2}{2b}(1-2e^{-bt}).\end{aligned}$$

3. We have the variance of the process as:

$$\mathrm{Var}[X_t] = \mathbb{E}[X_t^2] - (\mathbb{E}[X_t])^2 = \frac{\sigma^2}{2b}(1-2r^{-bt}),$$

which approaches $\frac{\sigma^2}{2b}$ as $t \to \infty$.

Here, since $X_t$ has solution in its form, the distribution is:

$$\mathcal{N}(\mu_t, \sigma_t^2) = \mathcal{N}\left(0, \frac{\sigma^2}{2b}\right).$$

Here, to derive this out, we shall need the method of *integrating factor*.
For the Langevin's case, we have $F(t) = e^{bt}$ and then multiply it on both sides:

$$F_t dX_t = -bF_t X_t dt + \sigma F_t dB_t = -X_t dF_t + \sigma dB_t,$$

that is $d(F_t X_t) = \sigma F_t dB_t$ or $F_t X_t - F_0 X_0 = \sigma\int_0^t F_s dB_s$. ⌟

Particularly, in the application of physics, $\sigma$ is typically the temperature in the model.

**Example V.1.10. Gradient Flow Pertuabed by Additive Noise.**

Consider the SDE:

$$dY_t = \underbrace{rdt}_{\text{drift}} + \underbrace{\alpha Y_t dB_t}_{\text{multiplicative noise}}.$$

Here, we use the integrating factor as:

$$F_t = \exp\left(-\alpha B_t + \frac{1}{2}\alpha^2 t\right),$$

in which we have:

$$dF_t = F_t(-\alpha dB_t + \alpha^2 dt).$$

Then, we consider the product rule:

$$\begin{aligned}d(F_t Y_t) &= F_t dY_t + Y_t dF_t + dF_t dY_t \\ &= F_t dY_t + Y_t F_t(-\alpha dB_t + \alpha^2 dt) + (-\alpha F_t dB_t)(\alpha Y_t dB_t) \\ &= F_t(dY_t - \alpha Y_t dB_t) = F_t r dt.\end{aligned}$$

Hence, we write in the standard form:

$$F_tY_t = F_0Y_0 + \int_0^t rF_s ds = Y_0 + \int_0^t \exp\left(-\alpha B_s + \frac{1}{2}\alpha^2 s\right) ds.$$

Thus, with only the $Y_t$ part, we have:

$$Y_t = Y_0 \exp\left(\alpha B_t - \frac{1}{2}\alpha^2 t\right) + \int_0^t \exp\left(-\alpha(B_t - B_s) - \frac{1}{2}\alpha^2(t-s)\right) ds.$$

⌟

In particular, when we have (semi-)linear SDEs, that is:

$$b(t,x) = \tilde{b}(t)(\alpha x + \beta) \text{ and } \sigma(t,x) = \tilde{\sigma}(t)(\alpha x + \beta),$$

we can often find the solutions using the integrating factor method.

Then, we want to consider some multidimensional case.

**Example V.1.11. Multi-Dimension Brownian Motion.**

Consider the SDE of:

$$LQ_t'' + RQ_t + \frac{1}{C}Q_t = G_t + \alpha W_t.$$

For SDEs, we can only have one dimension, so we consider the vector version:

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Q_t \\ Q_t' \end{pmatrix},$$

so that we have:

$$\begin{cases} X_1' = X_2, \\ LX_2' = -RX_2 - \frac{1}{C}X_1 + G_t + \alpha. \end{cases}$$

Hence, we have:

$$dX = dX(t) = AX(t)dt + H(t)dt + KdB_t,$$

with $dX = \begin{pmatrix} dX_1 \\ dX_2 \end{pmatrix}$, $A = \begin{pmatrix} 0 & 1 \\ -\frac{1}{CL} & -\frac{R}{L} \end{pmatrix}$, $H(t) = \begin{pmatrix} 0 \\ \frac{G_t}{L} \end{pmatrix}$, and $K = \begin{pmatrix} 0 \\ \frac{\alpha}{L} \end{pmatrix}$.
We rewrite the differential form by multiplying $\exp(-At)$ on both sides:

$$\exp(-At)dX(t) - \exp(-At)AX(t)dt = \exp(-At)[H(t)dt + KdB_t],$$

in which we can consider the left hand side as $d[\exp(-At)X(t)]$, which is the product rule in multi-dimension.
Recall for the matrix exponentials, we have:

$$\exp(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!},$$

and for the right hand side, we have:

$$\exp(-At)X(t) - X(0) = \int_0^t \exp(-As)H(s)ds + \int_0^t \exp(-As)KdB_s.$$

Hence, we have:

$$X(t) = \exp(At)[X(0) + \exp(-At)KB_t] + \exp(At)\int_0^t \exp(-As)[H(s) + AKB_s]ds.$$

## V.2 Existence and Uniqueness (Strong Solution)

In general, we see some issues when there is not a unique solution, we we consider the following theorem.

**Theorem V.2.1. Existence and Uniqueness for SDEs.**

Let $T > 0$ and $b(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma(\cdot,\cdot) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ be measurable function satisfying that:

1. **Linear Growth.** $|b(t,x)| + |\sigma(t,x)| \leq C(1 + |x|)$ for $x \in \mathbb{R}^n$ and $t \in [0,T]$,
2. **Lipschitz Condition.** $|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|$ for $x,y \in \mathbb{R}^n$ and $t \in [0,T]$.

Let $Z$ be a random variable independent of the Brownian motion and $\mathbb{E}[|Z|^2] < \infty$.
Then, the SDE:

$$\begin{cases} dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, & 0 < t \leq T, \\ X_0 = Z, \end{cases}$$

has a unique $t$-continuous solution $X_t(\omega)$ with property that:

$$X_t(\omega) \text{ is adapted to } \mathcal{F}^2 = \sigma(Z, B_s, s \leq t) \text{ and } \mathbb{E}\left[\int_0^T |X_t|^2 dt\right] < \infty.$$

Note that this is different from the existence and uniqueness of ODEs, and we are only enforcing continuity, but not differentiability of the solution.

**Remark V.2.2.**

- When the **linear growth condition** does not hold, then there is no **global solution**.
  When the **linear growth condition** holds, then we have existence, that is, there exists $X = \{X_t\}_{t\in[0,T]}$ such that the stochastic integral of the SDE is adapted.
- When the **Lipschitz condition** does not hold, then there is no **uniqueness**.
  When the **Lipshitz condition** holds, then we have (pathwise) uniqueness, then for any $X, Y$ that satisfies the solution, $\mathbb{P}\{t \in [0,1] : X(t) = Y(t)\} = 1$.

Since they were applied to ODEs, then they are necessary condition.

**Remark V.2.3. Yamaha-Watanabe Theory.**

Here, this is also called the Yamaha-Watanabe result in which for:

$$\sigma(t, x) = \sqrt{x},$$

we can still guarantee the existence and uniqueness result.
In fact, it is prove that for $\sigma = x^\beta$, it satisfies if $\beta \geq 1/2$ and not if $\beta < 1/2$.
Hence, the diffusion part conditions in Theorem V.2.1 is **not** necessary. ⌟

*Proof of Theorem V.2.1.* Here, we prove the *uniqueness* and *existence* separately.

- (Uniqueness:) Consider Itô isometry and Lipschitz condition. Let $X_1(t, \omega) = X$ and $X_2(t, \omega) = \hat{X}$ be the solutions with initial values $Z = \hat{Z}$. We put:

  $$u(s, \omega) = b(s, X_s) - b(s, \hat{X}_s),$$

  and:

  $$\gamma(s, \omega) = \sigma(s, X_s) - \sigma(s, \hat{X}_s).$$

  Then, we have the expectation:

  $$\begin{aligned}
  \mathbb{E}[|X_t - \hat{X}_t|^2] &= \mathbb{E}\left[\left(Z - \hat{Z} + \int_0^t u ds + \int_0^t \gamma dB_s\right)^2\right] \\
  &\leq 3\mathbb{E}[|Z - \hat{Z}|^2] + 3\mathbb{E}\left[\left|\int_0^t u ds\right|^2\right] + 3\mathbb{E}\left[\left|\int_0^t \gamma dB_s\right|^2\right] \\
  &\leq 3\mathbb{E}[|Z - \hat{Z}|^2] + 3t\mathbb{E}\left[\int_0^t u^2 ds\right] + 3\mathbb{E}\left[\int_0^t \gamma^2 ds\right] \\
  &\leq 3\mathbb{E}[|Z - \hat{Z}|^2] + 3(1 + T)D^2 \int_0^t \mathbb{E}[|X_s - \hat{X}_s|^2] ds.
  \end{aligned}$$

  So the function $v(t) = \mathbb{E}[|X_t - \hat{X}_t|^2]$ for $0 \leq t \leq T$ satisfied:

  $$V(t) \leq F + A \int_0^t v(s) ds,$$

  where:

  $$F = 3\mathbb{E}[|Z - \hat{Z}|^2] = 0, \text{ and } A = 3(1 + T)D^2,$$

  and so by the Gronwall inequality, we conclude:

  $$V(t) \leq F \exp(At),$$

  and by Gronwall's inequality, we conclude that:

  $$V(t) \leq F \exp(At), \text{ and we have } F = 0 \text{ so that } V(t) = 0.$$

**Remark V.2.4.** Note that the $|\sigma(t, x) - \sigma(t, y)|$ is not necessary, for example, with $\sigma(t, x) = \sqrt{x}$ or Osgood's condition.

Note that with Yamada-Watanabe, we have:

$$\tilde{V}(t) = \mathbb{E}[|X_t - \hat{X}_t|],$$

so we have no Itô isometry, but we still have Itô formula similar as local time. ⌟

- (Existence:) Similar to ODEs, we use the Picard iteration.
  We define:
  $$Y_t^{(0)} = X_0 \text{ and } Y_t^{(k)} = Y_t^{(k)}(\omega)$$
  inductively, as follows:
  $$Y_t^{(k+1)} = X_0 + \int_0^t b(s, Y_s^{(k)})ds + \int_0^t \sigma(s, Y_s^{(k)})dB_s.$$
  Then the same argument as uniqueness part implies that:
  $$\mathbb{E}[|Y_t^{(k+1)} - Y_t^{(k)}|] \le 3(1+T)D^2 \int_0^t \mathbb{E}[|Y_s^{(k)} - Y_s^{(k-1)}|^2]ds,$$
  and:
  $$\mathbb{E}[|Y_t^{(1)} - Y_t^{(0)}|^2] \le wC^2t^2\mathbb{E}[(1+|X_0|)^2] + 2C^2t\big(E[|X_0|^2]+1\big) \le A_1 t,$$
  where $A_1$ depends on $C$, $T$, and $\mathbb{E}[X_0^2]$.
  By induction on $k$, we obtain that:
  $$\mathbb{E}[|Y_s^{(k)} - Y_s^{(k-1)}|^2] \le \frac{A_2^{k+1}t^{k+1}}{(k+1)!},$$
  where $A_2$ depends on $C$, $T$, and $\mathbb{E}[X_0^2]$.
  Now we have:
  $$\begin{aligned}\|Y_t^{(m)} - Y_t^{(n)}\|_{L^2(m\times\mathbb{P})} &= \left\|\sum_{k=1}^{m-1}(Y_t^{(k+1)} - Y_t^{(k)})\right\| \\ &\le \sum_{k=n}^{m-1}\|Y_t^{(k+1)} - Y_t^{(k)}\|_{L^2(m\times\mathbb{P})} = \sum_{k=n}^{m-1}\left(\mathbb{E}\left[\int_0^T |Y_t^{(k+1)} - Y_t^{(k)}|^2 dt\right]\right)^{\frac{1}{2}} \\ &\le \sum_{k=n}^{m-1}\left(\int_0^t \frac{A_2^{k+1}t^{k+1}}{(k+1)!}dt\right)^{\frac{1}{2}} = \sum_{k=n}^{m-1}\left(\frac{A_2^{k+1}T^{k+2}}{(k+2)!}\right)^{\frac{1}{2}} \to 0\end{aligned}$$
  as $m, n \to \infty$. Therefore, $\{Y_t^{(n)}\}_{n=0}^{\infty}$ is a Cauchy sequence in $L^2(m \times \mathbb{P})$. Hence, it is convergent in $L^2(m \times \mathbb{P})$, so we define:
  $$X_0 = \lim_{n\to\infty} Y_t^{(n)} \text{ in } L^2(m\times\mathbb{P})$$
  and it is $\mathcal{F}_t^2$-measurable for all $t \in [0, T]$.
  Then,we show that the function actually satisfies the SDE. For all $n$ and $t \in [0, T]$, we have:
  $$Y_t^{(n+1)} = X_0 + \int_0^t b(s, Y_s^{(n)})ds + \int_0^t \sigma(s, Y_s^{(n)})dB_s,$$

and we want to show the $L^2$ condition that:

$$\int_0^t b(s, Y_s^{(n)})ds \xrightarrow{L^2} \int_0^t b(s, X_s)ds$$

using Hölder inequality and Itô isometry.

For the $t$-continuous part, both integrals have continuous versions.

Hence, we have proven the existence and uniqueness of the solution. □

**Remark V.2.5.** The solution obtained in the previous section is called a **strong solution**, where the $\sigma$-algebra $(\mathcal{F}_t^z)$ and Brownian motion $(\{B_t\})$ is given and fixed. ⌟

**Definition V.2.6. Strong Solution.**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ and Brownian motion $\{B_t\}$. We say $X_t$ is strong solution if:

- $X_t$ is $\mathcal{F}_t^z = \sigma(\{\mathcal{F}_t\}, \mathbb{Z})$-adapted, and
- $X_t$ that satisfies the SDE and $\mathbb{P}\{\int_0^1 |b(s, X_S)| + |\sigma(s, X_s)|^2 ds < \infty\} = 1$.

⌟

**Remark V.2.7.** If the SDE satisfies the conditions in the **Existence and Uniqueness theorem**, then it has a unique strong solution. ⌟

For the **weak solution**, we need to find and construct the $\sigma$-algebra $(\tilde{\mathcal{F}}_t^z)$ and Brownian motion $(\{\tilde{B}_t\})$, which is often called a **distribution solution**. Another type of solution is a **martingale solution**.

## V.3 Weak Solution

**Definition V.3.1. Weak Solution.**

A **weak solution** of a Stochastic differential equation is a triple $\big((X, B), (\Omega, \mathcal{F}, \mathbb{P}), \mathcal{H}\big)$ such that:

1. $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space, $\mathcal{H}_t \subset \mathcal{F}$ is the filtration.
2. $X_t$ is $\mathcal{H}_t$-adapted, for the stochastic integral form of the SDE, and
3. $\mathbb{P}\left\{\int_0^t |b(s_0 X_s)|ds + |\sigma(s, X_s)|^2 ds < +\infty\right\} = 1$.

⌟

**Remark V.3.2.** Weak solution for SDEs is not the same as weak solution for PDEs, for example:

$$\partial_t u - \frac{1}{2}\Delta u = f \text{ on } \mathbb{R}^d$$

Then for all $\phi \in C^\infty(\mathbb{R}^d)$, we have:

$$\int_0^T \int_{\mathbb{R}^d} \phi \left[\partial_t u - \frac{1}{2}\Delta u\right] = \int_0^T \int_{\mathbb{R}^d} f\phi.$$

Also, for the martingale solutions, we have the test function that for all $\phi$, we have:

$$\phi(X_t) = \int_0^t \phi(X_s)ds$$

must be a martingale. ⌟

Then, we also want a weak uniqueness.

**Definition V.3.3. Weak Uniqueness.**

Let the solutions be $(X^{(1)}, B^{(1)})$ and $\{X^{(2)}, B^{(2)}\}$, and we defined **uniqueness in law** such that $\text{Law}(X^{(1)}) = \text{Law}(X^{(2)})$, namely, for all $t_1, \cdots, t_k \in [0, T]$ and for all $k \in \mathbb{N}$:

$$\mathbb{P}_1(X_{t_1}^{(1)} \in A_1, \cdots, X_{t_k}^{(1)} \in A_k) = \mathbb{P}_2(X_{t_1}^{(2)} \in A_1, \cdots, X_{t_k}^{(2)} \in A_k)$$

for all $A_1, \cdots, A_k \in \mathcal{B}(\mathbb{R}^n)$. This is equivalently that:

$$\mathbb{P}_1(X^{(1)} \in A) = \mathbb{P}_2(X^{(2)} \in A) \text{ for all } A \in \mathcal{B}(\mathbb{R}^{[0,T]}).$$

⌟

**Remark V.3.4. Gisanov's Theorem.**

For exponential martingale, martingale under $\mathbb{P}$ implies Brownian motion under a new probability measure $\mathbb{M}$. ⌟

**Proposition V.3.5. Existence and Uniqueness $\Longrightarrow$ Weak Uniqueness.**

If $b$ and $\sigma$ satisfies the linear growth and Lipschitz condition, then a solution (weak or strong) is weakly unique.

In fact, the strong/pathwise uniqueness implies weak uniqueness. It can be shown by applying the Picard iteration and induction, that is:

$$X_t^{(k+1)} = 2 + \int_0^t b(s, X_s^{(k)})ds + \int_0^t \sigma(s, X_s^{(k)})dB_s.$$

Hence, $X^{(k)}$ is weakly unique, so that $X^{(k+1)}$ is weakly unique.

**Remark V.3.6. Watanabe theorem.**

Strong uniqueness implies weak uniqueness. ⌟

**Theorem V.3.7. Convergen in Law for Transpose.**

For an Itô process $dY_t = vdB_t$ and $Y_0 = 0$, we have $V(t, \omega) \in \mathcal{V}_{\mathcal{H}}^{n \times m}$, and $VV^\intercal = I_n$ almost surely.

**Example V.3.8. Weak Solution is more General.**

We want to have a case of no strong solution but only weak solution.
The Tanaka equation in 1D is:

$$dX_t = \operatorname{sgn}(X_t)dB_t, \qquad X_0 = 0.$$

Here, we note that $\sigma(x) = \operatorname{sgn}(x)$ does not satisfy Lipschitz condition, and we want to prove that strong solution does not exist.

*Proof.* Suppose the strong solution exists for Tanaka equation, we let $\hat{B}_t$ denote the Brownian motion and $\hat{\mathcal{F}}_t = \sigma\left(\left\{\widehat{B}_s\right\}_{s\le t}\right)$.
Then, we define:

$$\begin{aligned} Y_t &= \int_0^t \operatorname{sgn}(\widehat{B}_s)d\widehat{B}_s \\ &= |\widehat{B}_t| - |\widehat{B_0}| - \widehat{L}_t(\omega), \end{aligned}$$

where we have local time as:

$$\lim_{\epsilon\to 0}\frac{1}{2\epsilon}\left|\left\{s\in[0,t] : \left|\hat{B}_s\right| \le \epsilon\right\}\right|.$$

Then $Y_t$ is adapted to $\hat{\mathcal{G}}_t = \sigma\big(\{|\hat{B}_s|\}_{s\le t}\big) \subsetneq \hat{\mathcal{F}}_t$.
But we also have $dB_t = \operatorname{sgn}(X_t)dX_t$, so $|\operatorname{sgn}(X_t)|^2 = 1$ implies that $X_t$ is a Brownian motion.
Recall the preceding theorem, $Y = \{Y_t\}$ coincides in Law with $h$-dimensional Brownian motion, and by the above argument applied to $\hat{B}_t = X_t$, $Y_t = B_t$.
Hence:

$$\sigma\left(\{B_s\}_{s\le t}\right) = \mathcal{F}_t \subsetneq \mathcal{M}_t = \sigma\big(\{X_s\}_{s\le t}\big) \subset \sigma\big(\{B_s\}_{s\le t}\big) = \mathcal{F}_t.$$

Hence, this is a contradiction as $\mathcal{F}_t \subsetneq \mathcal{F}_t$. □

Then, we want to show that a weak solution exists. We choose $X_t$ to be any Brownian motion $\hat{B}_t$. Then we define $\widetilde{B}_t$ as:

$$\widetilde{B}_t = \int_0^t \operatorname{sgn}(X_s)dX_s,$$

*i.e.*, we have:

$$d\widetilde{B}_t = \operatorname{sgn}(X_t)dX_t \qquad \text{and} \qquad dX_t = \operatorname{sgn}(X_t)d\widehat{B}_t,$$

and hence we have weak uniqueness. ⌟

# VI Diffusion Models

## VI.1 Markov Property

**Definition VI.1.1. Time-homogeneous Itô Diffusion.**

A stochastic process $X_t(\omega) = X(t,\omega) : [s,\infty) \times \Omega \to \mathbb{R}^n$ is call **time-homogeneous** diffusion if it satisfies a SDE of the form:

$$dX_t = b(X_t, t)dt + \sigma(X_t)dB_t, \quad t \geq s, \qquad X_s = 0$$

where $B_t$ is a $n$-dimensional Brownian motion, $b : \mathbb{R}^n \to \mathbb{R}^n$, and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ satisfying **linear growth** and **Lipschitz** conditions:

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x-y| \text{ for all } x, y \in \mathbb{R}^n.$$

⌟

Here, we denote the unique solution by:

$$X_t = X_t^{s,x}; \qquad t \geq s.$$

If $s = 0$, then $X_t^x = X_t^{0,s}$, and the process satisfies that:

$$\begin{aligned} X_{s+k}^{s,x} &= x + \int_s^{s+h} b(X_u^{s,x})du + \int_s^{s+k} \sigma(X_u^{s,x})dB_u \\ &= x + \int_0^h b(X_{s+v}^{s,x})dv + \int_0^h \sigma(X_{s+v}^{s,x})d\tilde{B}_v. \end{aligned}$$

where $\tilde{B}_r = B_{s+v} - B_s$ for $v \geq 0$ is a new Brownian motion.

Since $\{\tilde{B}_v\}_{v\geq 0}$ and $\{B_t\}_{t\geq 0}$ have the same $\mathbb{P}^0$-distribution, it follows from the weak uniqueness of the solution, the SDE:

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \qquad X_0 = x$$

that $\{X_{s+h}^{s,x}\}_{h\geq 0}$ and $\{X_h^{0,x}\}_{h\geq 0}$ have the same $\mathbb{P}^0$-distribution, *i.e.*, $\{X_t\}_{t\geq 0}$ is time-homogeneous.

Here, we let $Q^x$ denote the probability law of a give (time-homogeneous) Itô diffusion $\{X_t\}$ when $X_0 = x \in \mathbb{R}^n$. The expectation with respect to $Q^x$ is denoted by $\mathbb{E}^x[\cdot]$, and we have:

$$\mathbb{E}_{Q^x}^x\left[f_1(X_{t_1})f_2(X_{t_2})\cdots f_k(X_{t_k})\right] = \mathbb{E}_{\mathbb{P}}\left[f_1(X_{t_1^x})f_2(X_{t_2^x})\cdots f_k(X_{t_k^x})\right]$$

for all bounded Borel functions, $f_1, \cdots, f_k$ and $t_1, \cdots, t_k \geq 0$, $k = 1, 2, \cdots$. Then, the filtration is:

$$\mathcal{F}^{(m)} = \sigma(\{B_s\}_{s\leq t}) \supset \mathcal{M}_t = \sigma(\{X_s\}_{s\leq t}).$$

**Theorem VI.1.2. Markov Property for Itô Diffusions.**

Let $f$ be a bounded Borel function: $\mathbb{R}^n \to \mathbb{R}$. Then, for $t, h \geq 0$:

$$\mathbb{E}^x[f(X_{t+h}) \mid \mathcal{F}_t^{(m)}] = \mathbb{E}^{X_t(\omega)}[f(X_h)] := \mathbb{E}^y[f(X_h)]\big|_{y=X_t(\omega)}.$$

Note that $\mathbb{E}^x$ means that we apply $Q^x$, the probability measure of $X_t^x$, while the left hand side can also be written as $\mathbb{E}[f(X_{t+h}^x \mid \mathcal{F}_t^{(m)})](\omega)$.

**Remark VI.1.3.** We can derive the equality that for $\mathcal{M}_t = \sigma(\{X_s\}_{s\leq t}) \subset \mathcal{F}_t^{(m)}$:

$$\begin{aligned}\mathbb{E}^x[f(X_{t+h}) \mid \mathcal{M}_t](\omega) &= \mathbb{E}^x\left[\mathbb{E}^x[f(X_{t+h}) \mid \mathcal{F}_t^{(m)}] \mid \mathcal{M}_t\right] \\ &= \mathbb{E}^x\left[\mathbb{E}^{X_t}[f(X_h)] \mid \mathcal{M}_t\right] = \mathbb{E}^{X_t(\omega)}[f(X_j)],\end{aligned}$$

where $\mathcal{M}_t$ is $\sigma(\{X_s\}_{s\leq t})$ and $\mathbb{E}^{X_t(\omega)}$ is only $\sigma(X_t)$-measurable.
Hence, Markov property means **memoryless**. ⌟

*Proof.* Since for $r \geq t$, we have:

$$X_r(\omega) = X_t(\omega) + \int_t^r b(X_u)du + \int_t^r \sigma(X_u)dB_u,$$

we have strong uniqueness of $X_r(\omega) = X_r^{t,X_t}(\omega)$.
If we define $F(x,t,r,\omega) = X_r^{t,x}(\omega)$, we have:

$$X_r(\omega) = F(X_t, t, r, \omega), \text{ for } r \geq t.$$

Note that $\omega \mapsto F(x,t,r,\omega)$ is independent of $\mathcal{F}_t^{(m)}$.
Hence, we can rewrite $X_r(\omega) = X_r^{t,X_t}(\omega)$ as:

$$\mathbb{E}[f\big(F(X_t, t, t+h, \omega)\big) \mid \mathcal{F}_t^{(m)}] = \mathbb{E}[f\big(F(x, 0, h, \omega)\big)]\big|_{x=X_t},$$

and we can put $g(X,\omega) = f \circ F(x,t,t+h,\omega)$, then $(x\omega) \mapsto g(x,\omega)$ is measurable. We can approximate $g$ pointwise bounded by function of form:

$$g(x,\omega) = \lim_{\ell\to\infty} \sum_{k=1}^{\ell} \phi_k(x)\psi_k(\omega).$$

Then, we can get that:

$$\begin{aligned}\mathbb{E}[g(X_t,\omega) \mid \mathcal{F}_t] &= \mathbb{E}\left[\lim_{\ell\to\infty}\sum_{k=1}^{\ell}\phi_k(x)\psi_k(\omega) \mid \mathcal{F}_t^{(m)}\right] = \lim_{\ell\to\infty}\sum_{k=1}^{\ell}\mathbb{E}[\phi_k(x)\psi_k(\omega) \mid \mathcal{F}_t^{(m)}] \\ &= \lim_{\ell\to\infty}\sum_{k=1}^{\ell}\phi_k(y)\mathbb{E}[\psi_k(\omega) \mid \mathcal{F}_t^{(m)}]\big|_{y=X_t} = \mathbb{E}\left[\lim_{\ell\to\infty}\sum_{k=1}^{\ell}\phi_k(y)\psi_k(\omega) \mid \mathcal{F}_t^{(m)}\right]\big|_{y=X_t} \\ &= \mathbb{E}[g(y,\omega) \mid \mathcal{F}_t^{(m)}]\big|_{y=X_t} = \mathbb{E}[g(y,\omega)]\big|_{y=X_t(\omega)}.\end{aligned}$$

Therefore, since $\{X_t\}$ is time-homogeneous, we have:

$$\begin{aligned}\mathbb{E}\big[f\big(F(X_t,t,t+h,\omega)\big) \mid \mathcal{F}_t^{(m)}\big] &= \mathbb{E}\big[f\big(F(y,t,t+h,\omega)\big)\big]\big|_{y=X_t(\omega)} \\ &= \mathbb{E}\big[f\big(F(y,0,h,\omega)\big)\big]\big|_{y=X_t(\omega)}.\end{aligned}$$

Hence, we have it satisfying the Markov property. □

The proof uses the **freezing technique** to write the function as a pointwise approximation.

**Remark VI.1.4. Freezing Lemma.**

Let $X : (\Omega, \mathcal{A}) \to (D, \mathcal{D})$ and $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ be two random variables. Assume that $\mathcal{X}, \mathcal{Y} \subset \mathcal{A}$ are $\sigma$-algebras such that $X$ is $\mathcal{X} \to \mathcal{D}$ measurable and $Y$ is $\mathcal{Y} \to \mathcal{E}$ measurable and $\mathcal{X} \perp \mathcal{Y}$ (they are independent), then:

$$\mathbb{E}[\Phi(X,Y) \mid \mathcal{X}] = \mathbb{E}[\Phi(x,Y)]\big|_{x=X} = \mathbb{E}[\Phi(X,Y) \mid X].$$

⌟

**Remark VI.1.5. Alternative Definition of Markov Process.**

A $\mathbb{R}^n$-valued stochastic process $\{X_t\}$ os called a Markov process with respect to $\mathcal{M}_t$ if there exists a transition probability function $p(s,t,x,dy)$ on $\mathbb{R}^n$ such that:

$$\mathbb{E}[f(X_t) \mid \mathcal{M}_s] = \mathbb{E}[f(X_t) \mid X_s] = p_{s,t}f(X_s) := \int_{\mathbb{R}} f(y)p(s,t,X_s,dy).$$

⌟

## VI.2 Stopping Time and Strong Markov Property

**Definition VI.2.1. Stopping Times.**

Let $\{N_t\}$ be an increasing family of $\sigma$-algebras. A function $\tau : \Omega \to [0, \infty)$ is called a **stopping time** with respect to $\{\mathcal{M}_t\}$ if:

$$\{\omega : \tau(\omega) \leq t\} \subset \mathcal{N}_t \text{ for all } t \geq 0.$$

⌟

**Remark VI.2.2.** If $\tau(\omega) = t_0$ for all $\omega$, then $\tau$ is a stopping time with respect to any filtration, since:

$$\{\tau \leq t\} = \begin{cases} \Omega, & \text{when } t_0 \leq t \\ \emptyset, & \text{when } t_0 > t. \end{cases}$$

⌟

**Proposition VI.2.3. Properties of Stopping Time.**

Let $\tau_1, \tau_2$ be two stopping times with respect to $\mathcal{F}_t$, then:

1. $\{J < t\} \in \mathcal{F}_t$, and $\{J = t\} \in \mathcal{F}_t$ for all $t \geq 0$.

2. $\tau_1 \wedge \tau_2 := \min\{\tau_1, \tau_2\}$ and $\tau_1 \vee \tau_2 := \max\{\tau_1, \tau_2\}$ are also stopping times.

*Proof.* 1. For all $s$, $\{J \leq s\} \in \mathcal{F}_s$, hence, for all $t$:

$$\{\tau < t\} = \bigcup_{n=1}^{\infty} \left\{\tau \leq \left(t - \frac{1}{n}\right) \vee 0\right\} \in \bigcup_{n=1}^{\infty} \mathcal{F}_{s_n} \subset \mathcal{F}_t.$$

2. For the second one, we can also represent the $\wedge$ and $\vee$ via countable unions and compliments of the $\sigma$-algebras. □

**Proposition VI.2.4. First Exit Time is Stopping Time.**

Let $U \subset \mathbb{R}^n$ be open, then the first exit time:

$$\tau_U := \inf\{t > 0 : X_t \notin U\}$$

is a stopping time with respect to $\{M_t\}$.

*Proof.* Since:

$$\{\omega : J_u \leq t\} = \bigcap_{m=1}^{\infty} \bigcup_{\substack{r \in Q \\ r < t}} \{\omega : X_r \notin K_m\} \in \mathcal{M}_t,$$

where $\{K_m\}_{m=1}^{\infty}$ is a sequence of increasing closed sets, *i.e.*, $U = \bigcup_{m=1}^{\infty} K_m$. □

**Definition VI.2.5. Stochstic Integral with respect to Stopping Time.**

If $G \in L^2$ and $\tau \leq T$ is a stopping time for some fixed $T > 0$, we define:

$$\int_0^{\tau} G dB_t = \int_0^T G \mathbb{1}_{\{t \leq \tau\}} dB_t.$$

⌟

Note that only by $\tau$ being a stopping time, we can have $\mathbb{1}_{\{t \leq \tau\}}$ measurable.

**Remark VI.2.6.** If $G \in L^2([0, T])$, and $\tau$ is the stopping time such that $0 \leq \tau \leq T$, then:

- $\mathbb{E}\left[\int_0^{\tau} G dB_t\right] = 0.$
- $\mathbb{E}\left[\left(\int_0^{\tau} G dB_t\right)^2\right] = \mathbb{E}\left[\int_0^{\tau} G^2 dt\right].$

⌟

**Definition VI.2.7. $\sigma$-algebra Generated by Infinity Stopping Time.**

Let $J$ be a stopping time with respect to $\{N_t\}$ and $N_\infty := \sigma\left(\{\bigcup_{t>0} N_t\}\right)$, then the $\sigma$-algebra $\mathcal{N}_\tau =$

$\sigma(\{X_{\tau\wedge t}\})$ is consisted of all sets $N \in \mathcal{N}_\infty$ such that $N \cap \{\tau \leq t\} \in \mathcal{N}_t$. ⌟

**Theorem VI.2.8. Strong Markov Property for Itô Diffusions.**

For a stopping time $\tau$, we have the **strong Markov property** that:

$$\mathbb{E}^x\left[f(X_{\tau+h}) \mid \mathcal{F}_\tau^{(m)}\right] = \mathbb{E}^{X_\tau}[f(X_h)] \text{ for all } h \geq 0.$$

Moreover, if $f_1, f_2, \cdots, f_k$ are bounded Borel function on $\mathbb{R}^n$, and $\tau$ is an $\mathcal{F}_t^{(m)}$-stopping time, then:

$$\mathbb{E}^x\left[f_1(X_{\tau+h_1})f_2(X_{\tau+h_2})\cdots f_k(X_{\tau+h_k}) \mid \mathcal{F}_\tau^{(m)}\right] = \mathbb{E}^{X_\tau}[f_1(X_{h_1})f_2(X_{h_2})\cdots f_k(X_{h_k})]$$

for all $h_i \geq 0$ where $1 \leq i \leq k$.

## VI.3 The Generator of an Itô Diffusion

**Definition VI.3.1. Infinitesimal Generator.**

Let $\{X_t\}$ be a time-homogeneous Itô diffusion on $\mathbb{R}^n$, then the *infinitesimal* generator $A$ of $X_t$ is defined by:

$$Af(x) = \lim_{t \searrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t} \text{ for } x \in \mathbb{R}^n.$$

In particular, the set of functions $f : \mathbb{R}^n \to \mathbb{R}$ such that the limit exists at $x$ is denoted by $\mathcal{D}_A(x)$. ⌟

Usually, we take $C^2(\mathbb{R}^n)$ for $\mathcal{D}_A(x)$ as the requirement.

Then, we will first see some consequences of the definition.

**Theorem VI.3.2. Dynkin's Formula.**

Let $f \in C_0^2(\mathbb{R}^n)$. Suppose $\tau$ is stopping time, $\mathbb{E}^x[\tau] < \infty$, then:

$$\mathbb{E}^x[f(X_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^\tau Af(X_s)ds\right],$$

where $\mathbb{E}^x$ is the expectation with respect to the law $R^x$ of $X_t$ starting from $x$:

$$R^x[Y_{t_1} \in F_1, \cdots, Y_{t_k} \in F_k] = \mathbb{P}^0[Y_{t_1}^x \in F_1, \cdots, Y_{t_k}^x \in F_k].$$

**Theorem VI.3.3. Expression of Infinitestimal Generator.**

Let $X_t$ be Itô diffusion:

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

then:
$$Af(x) = \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x) + \frac{1}{2}\sum_{i,j}\left(\sigma(x)\sigma^\intercal(x)\right)_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

**Example VI.3.4. Infinitesimal Generator as Laplace-Beltrami Operator.**

The $n$-dimensional Brownian motion $\{B_t\}$:
$$dX_t = dB_t,$$
where $b = 0$ and $\sigma = \mathrm{Id}_n$. So the generator of $X_t = B_t$ is:
$$Af = 0 + \frac{1}{2}\sum_{i,j}\delta_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x_i^2}.$$

Hence, $A = \frac{1}{2}\Delta$, which is half of the Laplace-Beltrami Operator operator.k ⌟

The above example is an effective connection between SDEs and Laplace equation in PDEs.

**Example VI.3.5. Infinitesimal Generator with Heat Operator.**

Let $\{B_t\}$ denote a 1-dimensional Brownian motion and $X = (X_1, X_2)$ be the solution of:
$$\begin{cases} dX_1 = dt, & X_1(0) = t_0, \\ dX_2 = dB_t, & X_2(0) = x_0, \end{cases}$$
*i.e.*, $dX = bdt + \sigma dB_t$, $X(0) = (t_0, x_0)$, where $b = (1, 0)$ and $\sigma = (0, 1)$.
Hence, we have:
$$Af = \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2},$$
for $f = f(t, x) \in C_0^2(\mathbb{R}^n)$, and hence is the heat operator in the heat equation. ⌟

This example is a connection between SDEs and heat equation in PDEs.

**Example VI.3.6. Probabilistic Approximation of PDEs.**

Let $U \subset \mathbb{R}^n$ be a smooth bounded domain and $\partial U$ is smooth consider:
$$\begin{cases} -\frac{1}{2}\Delta u = 1, & \text{in } U^\circ, \\ u = 0, & \text{on } \partial U. \end{cases}$$

We claim $u(x) = \mathbb{E}[\tau_{x,U}]$ where $\tau_{x,U}$ is the first time $X_t^x$ hits $\partial U$. ⌟

*Proof.* By Dykin's formula (Theorem VI.3.2), $Au = \frac{1}{2}\Delta u$, we have:

$$\mathbb{E}[u(X_{\tau x})] - \mathbb{E}[u(X_0)] = \mathbb{E}\left[\int_0^{\tau_x} \frac{1}{2}\Delta u(X_s)ds\right].$$

Since $\frac{1}{2}\Delta u = -1$, we get:

$$u(x) - \mathbb{E}[u(X_{\tau_x})] = \mathbb{E}\left[\int_0^{\tau_x} 1ds\right] = \mathbb{E}[\tau_x].$$

Hence, we have $u(x) = \mathbb{E}[\tau_x]$. □

**Remark VI.3.7.** Here, this can be done using a Monte Carlo approximation, considering a random starting point within $U^\circ$ and let a random particle to get around. ⌟

**Proposition VI.3.8. Itô Process Expectation.**

Let $Y_t = Y_t^x$ be an Itô process in $\mathbb{R}^n$ of the form:

$$Y_t^x(\omega) = \int_0^t u(s,\omega)ds + \int_0^t v(s,\omega)dB_s(\omega).$$
$$dY_t = u(t,\omega)dt + v(t,\omega)dB_t \text{ with } Y_0 = x,$$

where $B$ is $m$-dimensional Brownian motion. Let $f \in C_0^2(\mathbb{R}^n)$, and $\tau$ be the stopping time such that $\mathbb{E}^x[\tau] < +\infty$. Assume $u, v$ are bounded, then:

$$\mathbb{E}^x[f(Y_t)] = f(x) + \mathbb{E}^x\left[\int_0^t \left(\sum_i u_i(s,\omega)\frac{\partial f}{\partial x_i}(Y_s) + \frac{1}{2}\sum_{i,j}(vv^\intercal)_{i,j}(s,\omega)\frac{\partial^2 f}{\partial x_i \partial x_j}(Y_s)ds\right)\right].$$

*Proof.* Put $Z = f(Y)$ and apply Itô lemma, we have:

$$\begin{aligned} dZ &= \sum_i \frac{\partial f}{\partial x_i}(Y)dY_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}(Y)dY_i dY_j \\ &= \sum_i \frac{\partial f}{\partial x_i}(Y)u_i dt + \sum_i \frac{\partial f}{\partial x_i}(Y)(vdB)_i + \frac{1}{2}\sum_{i,j}\frac{\partial f}{\partial x_i \partial x_j}(Y)(VdB)_i(vdB)_j. \end{aligned}$$

Note that:

$$(vdB)_i(vdB)_j = \left(\sum_k v_{i,j}dB_k\right)\left(\sum_h v_{j,h}dB_h\right) = \sum_k v_{i,k}v_{j,k}dt.$$

This gives that:

$$f(Y_t) = f(Y_0) + \int_0^t \left(\sum_i u_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(vv^\intercal)_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}\right)ds + \sum_{i,k}\int_0^t v_{i,k}\frac{\partial f}{\partial x_i}dB_k.$$

Hence, we have:

$$\mathbb{E}_x[f(X_\tau)] = f(x) + \mathbb{E}^x\left[\int_0^t \left(\sum_i u_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(vv^\intercal)_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}\right)ds\right] + \sum_{i,k}\mathbb{E}^x\left[\int_0^t v_{i,k}\frac{\partial f}{\partial x_i}dB_k\right].$$

We want to show that the last expectation is zero.
If $g$ is a bounded Borel measurable function such that $|g| \leq M$, then for all $\ell \in \mathbb{N}$:

$$\mathbb{E}^x\left[\int_0^{\tau\wedge\ell} g(Y_s)dB_s\right] = \mathbb{E}^x\left[\int_0^{\ell} \mathbb{1}_{\{s\leq\tau\}}g(Y_s)dB_s\right] = 0.$$

Moreover, we have:

$$\begin{aligned}\mathbb{E}^x\left[\left(\int_0^{\tau} g(Y_s)dB_s - \int_0^{\tau\wedge\ell} g(Y_s)dB_s\right)^2\right] &= \mathbb{E}^x\left[\int_{\tau\wedge\ell}^{\tau} g^2(Y_s)ds\right] \\ &\leq M^2\mathbb{E}^x[\tau - \tau\wedge\ell] \to 0 \text{ as } \ell\to\infty,\end{aligned}$$

and hence we have the desired part to be zero. □

Then, we can think of Theorem VI.3.2 and Theorem VI.3.3 as the corollaries of the above proposition.

*Proof of Theorem VI.3.3.* Directly from Proposition VI.3.8 by setting:

$$u(t,\omega) = b\big(X_t(\omega)\big), \qquad v(t,\omega) = \sigma\big(X_t(\omega)\big).$$

□

*Proof of Theorem VI.3.2.* Consequence of Proposition VI.3.8 by replacing $Af(X_s)$ as generator. □

**Example VI.3.9. Green's Formula for Harmonic PDE.**

Let $U \subset \mathbb{R}^n$ be a smooth bounded domain and $g : \partial U \to \mathbb{R}$ a continuous function. Consider:

$$\begin{cases}\Delta u = 0, & \text{in } U^\circ, \\ u = g, & \text{on } \partial U\end{cases}$$

Let $X_t(\omega) = B_t(\omega) + x$. Then $u(x) = \mathbb{E}[g(X_{\tau_x})]$, where $\tau_x$ is the first time such that $X$ hits $\partial U$.
Then, we have that:

$$\begin{aligned}\mathbb{E}[g(X_{\tau_x})] = \mathbb{E}[u(X_{\tau_x})] &= \mathbb{E}[u(X_0)] + \mathbb{E}\left[\int_0^{\tau_x} \frac{1}{2}\Delta u(X_s)ds\right] \\ &= \mathbb{E}[u(x)] = u(x).\end{aligned}$$

⌟

The main thing about the application is how to construct the stopping time model.

**Example VI.3.10. Particle Escaping from Ball.**

Consider $n$-dimensional Brownian motion $B = (B_1, \cdots, B_n)$, starting at $a = (a_1, \cdots, a_n) \in \mathbb{R}^n$, and $|a| < R$, which is the expected value of first exit time $\tau_k$ of $B$ from the ball:

$$K = K_r = \{x \in \mathbb{R}^n : |x| < R\}.$$

Here, let $n \in \mathbb{N}$ be fixed and apply the Dykin's formula with $X = B$, $\tau = a_k = \min\{k, \tau_k\} = \tau_k \wedge k$, we have:

$$\mathbb{E}^a[|B_{\sigma_k}|^2] = \mathbb{E}^a[f(B_{\sigma_k})] = f(a) + \mathbb{E}^a\left[\int_0^{\sigma_k} \frac{1}{2}\Delta f(B_s)ds\right]$$
$$= |a|^2 + \mathbb{E}^a\left[\int_0^{\sigma_k} nds\right] = |a|^2 + h\mathbb{E}^a[\sigma_k].$$

Then $\mathbb{E}^a[\tau_k] = \frac{1}{n}[R^2 - |a|^2]$. So for $k \nearrow \infty$, we have $\mathbb{E}^a[\tau_k] < \infty$.
Note that when the dimension is big, it takes less time to approach the boundary.
Then, consider $|b| > R$, what is the probability that $B$ starting at $b$ ever hits $K$?
Let $\alpha_k$ be the first exit time from the annulus, we have:

$$\mathbb{A}_k = \{x : R < |x| < 2^k R\},$$

and we put:

$$T_k = \inf\{t > 0 : B_t \in K\}.$$

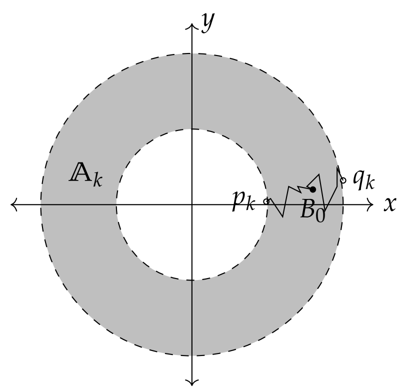


*Figure VI.1. Brownian motion escaping from the annulus $\mathbb{A}_k$.*

Let $f = f_{R,k}$ be $C^2$ with compact supremum and if $R \leq |x| < 2^k R$. Eventually, we conclude that:

$$f(x) = \begin{cases} -\log|x|, & \text{when } n = 2, \\ |x|^{2-n}, & \text{when } n = 3. \end{cases}$$

It should be noticed that this is the solution to $\Delta f = 0$ in $\mathbb{A}_k$.
Then, by the Dynkin's formula, we have:

$$\mathbb{E}^b[f(B_{\sigma_k})] = f(b).$$

As we put $p_k = \mathbb{P}^b[|B_{\sigma_k} = R]$ and $q_k = \mathbb{P}^b[|B_{\sigma_k}| = 2^k R]$.

- For $n = 2$, we get that:

$$-\log R \cdot (1 - q_n) - (\log R + k\log 2)q_k = -\log|b|$$
$$-\log k + \log Rq_k - \log Rq_k - k\log 2q_k = -\log|b|$$
$$q_k = \frac{\log|b|}{k\log 2 + \log k} \searrow 0 \text{ as } k \nearrow \infty.$$

Thus, we have that:

$$\mathbb{P}^b[\tau_k < \infty] = 1,$$

*i.e.*, the Brownian motion is recurrent in $\mathbb{R}^2$.

- For $n > 2$, we have:

$$p_k R^{2-n} + q_k (2^k R)^{2-n} = |b|^2 2 - n.$$

Hence, as $k \nearrow \infty$, we have:

$$\mathbb{P}_k = \mathbb{P}^b[\tau_k < \infty] = \left(\frac{|b|}{k}\right)^{2-n},$$

*i.e.*, the Brownian motion is transient in $\mathbb{R}^k$ for $k > 2$.

In particular:

- When $n = 2$, the Brownian motion is recurrent.
- When $n \geq 3$, the Brownian motion is transient.

Hence, for a *random walk*, it is almost surely to return in $\mathbb{R}^2$ (like a *drunk man*), but not in higher dimensions (like a *drunk bird*). ⌟

# VII Topics in Diffusion Theory

## VII.1 Kolmogorov's Backward/Forward Equation

Let $X_t$ be Itô diffusion in $\mathbb{R}^n$ with generator $A$, we choose the function $f \in C_0^2(\mathbb{R}^n)$ with $\tau = t$ in Dynkin's formula, and we see that:

$$u(t,x) = \mathbb{E}^x[f(X_t)] = f(x) + \int_0^t \mathbb{E}[Af(X_s)]ds,$$

and it is a diffusion with respect to $t$ and:

$$\frac{\partial u}{\partial t} = \mathbb{E}^x[Af(X_t)] = A\mathbb{E}^x[f(X_t)] = Au.$$

**Theorem VII.1.1. Kolmogorov's Backward Equation.**

Let $f \in C_0^2(\mathbb{R}^n)$, with $u$ defined as above, then $u(t,\cdot) \in \mathcal{D}_A$ for all $t$ and:

$$\begin{cases} \dfrac{\partial u}{\partial t} = Au, & \text{for } t > 0, x \in \mathbb{R}^n, \\ u(0,x) = f(x), & \text{for } x \in \mathbb{R}^n. \end{cases}$$

Moreover, if $\omega(t,x) \in C^{1,2}(\mathbb{R} \times \mathbb{R}^n)$ is bounded and satisfies the PDEs, then $\omega(t,x) = u(t,x)$.

Here, we can think of $A$ as an operator acting on $u$, and the argument argues about existence and uniqueness of the solution.

*Proof.* (Existence:) Let $g(x) = u(t,x)$, then since $t \mapsto u(t,x)$ is differentiable, we have:

$$\begin{aligned} \frac{\mathbb{E}^x[g(X_r)] - g(x)}{r} &= \frac{1}{r}\mathbb{E}^x\left[\mathbb{E}^{X_r}[f(X_t)] - \mathbb{E}^x[f(X_t)]\right] \\ &= \frac{1}{r}\mathbb{E}^x\left[\mathbb{E}^x[f(X_{t+r}) \mid \mathcal{F}_r] - \mathbb{E}^x[f(X_t) \mid \mathcal{F}_r]\right] && \text{(Markov \& towering property)} \\ &= \frac{1}{r}\mathbb{E}^x[f(X_{t+r}) - f(X_t)] && \text{(Towering property)} \\ &= \frac{u(t+r,x) - u(t,x)}{r} = \frac{\partial}{\partial t}u(t,x). \end{aligned}$$

Here, we consider, the left hand side as:

$$\frac{\mathbb{E}^x[g(X_r)] - g(x)}{r} = Ag(x) = A_x u(t,x).$$

(Uniqueness:) Assume that $\omega(t,x) \in C^{1,2}(\mathbb{R} \times \mathbb{R}^n)$ and it satisfies the PDE, then:

$$\tilde{A}_{s,x}\omega = -\frac{\partial \omega}{\partial t} + A_x\omega = 0 \text{ for } t > 0 \text{ and } x \in \mathbb{R}^n,$$

and $\omega(0,x) = f(x)$ for $x \in \mathbb{R}^n$.
We fix $(s,x) \in \mathbb{R} \times \mathbb{R}^n$, we define the process $Y_t$ in $\mathbb{R}^{n+1}$ by:

$$Y_t = (s-t, X_t^{0,x}),$$

and then $Y_t$ has the generator $\tilde{A}$, by Dykin, we ahve:

$$\mathbb{E}^{s,t}[\omega(Y_{t\wedge\tau_R})] = \omega(s,x) + \mathbb{E}^{s,x}\left[\int_0^{t\wedge\tau_R} \tilde{A}_{s,x}\omega(Y_r)dr\right],$$

where $\tau_R = \inf\{t > 0 : |X_t| \geq R\}$.
Then, as $R \nearrow \infty$, we get:

$$\mathbb{E}^{s,x}[\omega(Y_t)] = \omega(s,x).$$

By choosing $t = s$, this implies that:

$$\begin{aligned}\omega(s,x) &= \mathbb{E}^{s,x}[\omega(Y_s)] = \mathbb{E}[\omega(0, X_s^{0,x})] \\ &= \mathbb{E}[f(X_s^{0,x})] = \mathbb{E}^x[f(X_s)] = u(s,x).\end{aligned}$$

□

**Remark VII.1.2. Differential Operator Notation.**

Here, we have:

$$Au(x) = L(x,D)u(x) = \sum_i b_i(x)\frac{\partial u}{\partial x_i}(x) + \frac{1}{2}\sum_{i,j}\left(\sigma(x)\sigma^\intercal(x)\right)_{i,j}\frac{\partial^2 u}{\partial x_i \partial x_j}(x),$$

where $L(x,D)$ is a differential operator. ⌟

**Remark VII.1.3. Backwards/Forward Probability for Markov Process.**

We consider $X$ has a transition probability density $p(t,x,y)$.

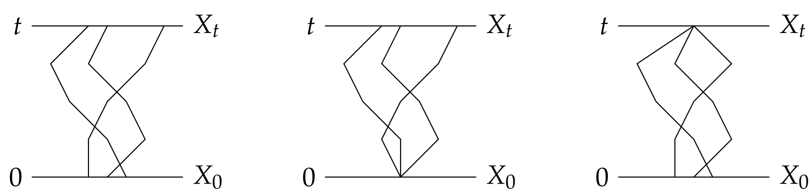


*Figure VII.1. A trajectories (left) from 0 of $X_0$ to t of $X_t$ with forward (middle) and backward (right) equation.*

The (backwards) density equation satisfies that:

$$\partial_t p(t,x,y) = L(x,D_x)p(t,x,y) \text{ for all } t > 0 \text{ and } y \in \mathbb{R}^n.$$

Here, we have $L^*\left(y, D_y u(y)\right)$ as the adjoint of $L(x,D_x)$, that is:

$$L^*\left(y, D_y u(y)\right) = -\sum_i \frac{\partial}{\partial y_i}\left(b_i(y)u(y)\right) + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial y_i \partial y_j}\left((\sigma\sigma^\intercal)_{i,j}(y)u(y)\right).$$

⌟

**Remark VII.1.4. Commutativity of Operator.**

Here, we can assume the transition measure of $X_t$ with density $p_t(x,y)$, *i.e.*, having that:

$$u(t,x) = \mathbb{E}^x[f(X_t)] = \int_{\mathbb{R}^n} f(y)p_t(x,y)dy.$$

Then, we can have that:

$$Au = A_x \int_{\mathbb{R}^n} f(y) p_t(x,y) dy = \int_{\mathbb{R}^n} f(y) A_x p_t(x,y) dy.$$

Here, in particular, we can consider $A$ as the partial derivative operator, so:

$$\frac{\partial u}{\partial t} = \int_{\mathbb{R}^n} f(y) \frac{\partial}{\partial t} P_t(x,y) dy,$$

and we can consider that for all $y \in \mathbb{R}^n$ that:

$$\frac{\partial}{\partial t} p_t(x,y) = A_x P_t(x,y).$$

⌟

**Remark VII.1.5. Forward Equation.**

By Dynkin, we have that:

$$\begin{aligned}\int_{\mathbb{R}^n} f(y) p_t(x,y) dy &= f(x) + \int_0^t \int_{\mathbb{R}^n} A_y f(y) p_s(x,y) dy ds \\ &= f(x) + \int_0^t \langle Af(\cdot), p_s(x,\cdot)\rangle_{L^2(\partial g)} \, ds = f(x) + \int_0^t \langle f(\cdot), A^* p_s(x,\cdot)\rangle_{L^2(\partial g)} \, ds,\end{aligned}$$

where we consider the equation with the adjoint as:

$$A_y^* \phi(y) = \sum_{i,j} \frac{1}{2} \frac{\partial^2}{\partial y_i \partial y_j} [(\sigma\sigma^\intercal)_{i,j} \phi] - \sum_i \frac{\partial}{\partial y_i} (b_i \phi).$$

Here, we note that:

$$A_y f(y) = b(y) f'(y) + a(y) f''(y),$$

and with $a = \frac{1}{2}\sigma^2$, we have:

$$\begin{aligned}\int_{\mathbb{R}} [b(y) f'(y) + a(y) f''(y)] p_t(x,y) dy &= \int_{\mathbb{R}} [b(y) p_t(x,y)] f'(y) dy + \int_{\mathbb{R}} [a(y) p_t(x,y)] f''(y) dy \\ &= -\int_{\mathbb{R}} \frac{\partial}{\partial y} [b(y) p_t(x,y)] f(y) dy + \int_{\mathbb{R}} \frac{\partial}{\partial y} [a(y) p_t(x,y)] f(y) dy.\end{aligned}$$

Hence, we have that:

$$\frac{\partial}{\partial t} p_t(x,y) = A_y^* p_t(x,y) \text{ for all } x, y \in \mathbb{R}^n \text{ and } t > 0.$$

⌟

**Example VII.1.6. Semilinear Hear Equation.**

Consider the following PDE:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{1}{2} \beta^2 x^2 \dfrac{\partial^2 u}{\partial x^2} + \alpha x \dfrac{\partial u}{\partial x}, & \text{for } t > 0, x \in \mathbb{R}, \\ u(0,x) = f(x), & \text{for } x \in \mathbb{R}, f \in C_0^2(\mathbb{R}). \end{cases}$$

Here, we suppose that $u$ is a good function, and we want:

$$Au(x) = \frac{1}{2} \beta^2 x^2 \frac{\partial^2 u}{\partial x^2} + \alpha x \frac{\partial u}{\partial x},$$

where we want to find the diffusion as:

$$b(x) = \beta x \qquad \text{and} \qquad \sigma(x) = \alpha x,$$

which is:

$$dX_t = \alpha X_t dt + \beta X_t dB_t,$$

and we notice that this is exactly in Example V.1.5 of Geometrical Brownian motion, and the solution is:

$$X_t = X_0 \exp\left[\left(\alpha - \frac{\beta^2}{2}\right) t + \beta B_t\right].$$

Here, by the backwards equation, we have:

$$u(t,x) = \mathbb{E}^x[f(X_t)] = \mathbb{E}[f(X_t^x)] = \mathbb{E}\left[f\left(x \exp\left[\left(\alpha - \frac{\beta^2}{2}\right) t + \beta B_t\right]\right)\right].$$

Here, we consider the transition density of the Brownian motion as:

$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{|x-y|^2}{2t}\right),$$

so we have the density as:

$$\begin{aligned}
u(t,x) &= \mathbb{E}\left[f\left(x \exp\left[\left(\alpha - \frac{\beta^2}{2}\right) t + \beta B_t\right]\right)\right] \\
&= \int_{\mathbb{R}} f\left(x \exp\left[\left(\alpha - \frac{\beta^2}{2}\right) t\right]\right) p_t(0,y) dy \\
&= \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f\left(x \exp\left[\left(\alpha - \frac{\beta^2}{2}\right) t\right]\right) \exp\left(-\frac{|x-y|^2}{2t}\right) dy.
\end{aligned}$$

Also, just to note, as $t \searrow 0$, we have $p_t(x,y) \to \delta(y-x)$. ⌟

## VII.2 Resolvent Operator

**Definition VII.2.1. Resolvent Operator.**

For $\alpha > 0$, $g \in C_b(\mathbb{R}^n)$, we define the **resolvent operator** $R_\alpha$ by:

$$R_\alpha g(x) = \mathbb{E}^x\left[\int_0^\infty e^{-\alpha t} g(X_t) dt\right].$$

⌟

**Proposition VII.2.2. Properties of Resolvent Operator.**

Let $g \in C_b(\mathbb{R}^n)$, then the resolvent operator satisfies that:

1. $R_\alpha g$ is a bounded continuous function.

2. Let $g$ be a lower bounded measurable function and define $u(x) = \mathbb{E}^x[g(X_t)]$:

- If $g$ is lower semi-continuous, *i.e.*, for all $x_0$, there exists a sequence $x_n \to x_0$ such that:

$$\liminf_{n\to\infty} f(x_n) \geq f(x_0),$$

then $u$ is also lower semi-continuous.

- If $g$ is bounded continuous, then $u$ is continuous.

These properties lead to stronger statement of the invertibility of the operator.

**Theorem VII.2.3. Identity of Operators.**

1. If $f \in X_0^2(\mathbb{R}^n)$, then $R_\alpha(\alpha - A)f = f$ for all $\alpha > 0$.
2. If $g \in C_b(\mathbb{R}^n)$, then $R_\alpha g \in \mathcal{D}_A$, and $(\alpha - A)R_\alpha g = g$ for all $\alpha > 0$.

*Proof.* 1. If $f \in C_0^2(\mathbb{R}^n)$, then:

$$\begin{aligned}
R_\alpha(\alpha - A)f(x) &= (\alpha R_\alpha f - R_\alpha A f)(x) \\
&= \alpha \int_0^\infty e^{-\alpha t}\mathbb{E}^x[f(X_t)]dt - \int_0^\infty e^{-\alpha t}\mathbb{E}^x[Af(X_t)]dt \\
&= -e^{-\alpha t}\mathbb{E}^x[f(X_t)]\Big|_{t=0}^\infty + \int_0^\infty e^{-\alpha t}\frac{d}{dt}\mathbb{E}^x[f(X_t)]dt - \int_0^\infty e^{-\alpha t}\mathbb{E}^x[Af(X_t)]dt \\
&= \mathbb{E}^x[f(X_0)] = f(x).
\end{aligned}$$

2. Suppose $g \in C_b(\mathbb{R}^n)$, then by the Markov property:

$$\begin{aligned}
\mathbb{E}^x[R_\alpha g(X_t)] &= \mathbb{E}^x\left[\mathbb{E}^{X_t}\left[\int_0^\infty e^{-\alpha s}g(X_s)ds\right]\right] = \mathbb{E}^x\left[\mathbb{E}^x\left[\theta_t\left(\int_0^\infty e^{-\alpha s}g(X_s)ds\right) \mid \mathcal{F}_t\right]\right] \\
&= \mathbb{E}^x\left[\mathbb{E}^x\left[\int_0^\infty e^{-\alpha s}g(X_{t+s})ds \mid \mathcal{F}_t\right]\right] = \mathbb{E}^x\left[\int_0^\infty e^{-\alpha s}g(X_{t+s})ds\right] = \int_0^\infty e^{-\alpha s}\mathbb{E}^x[g(X_{t+s})]ds,
\end{aligned}$$

where $\theta_t$ denotes the shift operator, shifting $X_s$ to $X_{t+s}$.
Here, we have, by definition that:

$$A(R_\alpha g) = \lim_{t\to 0}\frac{1}{t}\left[\mathbb{E}^x[R_\alpha g(X_t)] - R_\alpha g(x)\right].$$

Then, we use integration by parts to obtain that:

$$\begin{aligned}
E^x[R_\alpha g(X_t)] &= \alpha\int_0^\infty e^{-\alpha s}\int_t^{t+s}\mathbb{E}^x[g(X_v)]dvds \\
&= \lim_{t\to 0}\frac{1}{t}\left\{\alpha\int_0^\infty e^{-\alpha s}\int_t^{t+s}\mathbb{E}^x[g(X_v)]dvds\frac{\partial}{\partial t}G(t,x)\big|_{t=0} - \alpha\int_0^\infty e^{-\alpha s}\int_0^s\mathbb{E}^x[g(X_v)]dvds\right\}
\end{aligned}$$

Hence, we have $A(R_\alpha, g) = \alpha R_\alpha g - g$. □

## VII.3 The Feynman-Kac Formula

We can find a generalization of the Kolmogorov's backward equation.

**Theorem VII.3.1. Feynman-Kac Formula.**

Let $f \in C_0^2(\mathbb{R}^n)$ and $q \in C(\mathbb{R}^n)$. Assume $q$ is lower bounded. Put $v(t,x) = \mathbb{E}^x\left[\exp\left(-\int_0^t q(X_s)ds\right) f(X_t)\right]$, where $X_t$ is an Itô diffusion with generator $A$, then $v(t,x)$ satisfies the PDE:

$$\begin{cases} \dfrac{\partial v}{\partial t} = Av - qv, & \text{when } t > 0, x \in \mathbb{R}^n, \\ v(0,x) = f(x), & \text{for } x \in \mathbb{R}^n. \end{cases}$$

Moreover, if $w(t,x) \in \mathcal{C}^{1,2}(\mathbb{R} \times \mathbb{R}^n)$ is bounded on $K \times \mathbb{R}^n$ for each compact $K \subset \mathbb{R}$, and $w$ solves the PDE, then $w(t,x) = v(t,x)$.

*Proof.* (Existence:) Let $Y_t = f(X_t)$ and $Z_t = \exp\left(-\int_0^t q(X_s)ds\right)$, then:

$$dZ_t = -Z_t q(X_t)dt \qquad \text{and so} \qquad d(Y_tZ_t) = Y_t dZ_t + Z_t dY_t.$$

Next, we consider $\frac{\partial}{\partial r} v(r,x)$:

$$\begin{aligned} \frac{1}{r}\left[\mathbb{E}^x[v(t,X_r)] - v(t,x)\right] &= \frac{1}{r}\left[\mathbb{E}^x\mathbb{E}^x[Z_t f(X_t)] - \mathbb{E}^x[Z_t f(X_t)]\right] \\ &= \frac{1}{r}\left\{\mathbb{E}^x E^x\left[\exp\left(-\int_0^t q(X_s s + r)ds\right) f(X_{t+r}) \middle| \mathcal{F}_r\right] - \mathbb{E}^x[Z_t f(X_t) \mid \mathcal{F}_r]\right\} \\ &= \frac{1}{r}\mathbb{E}^x\left[Z_{t+r}\exp\left(\int_0^r q(X_s)ds\right) f(X_{t+r}) - Z_t f(X_t)\right] \\ &= \frac{1}{r}\left\{E^x[Z_{t+r}f(X_{t+r}) - Z_t f(X_t)] + \mathbb{E}^x\left[f(X_{t+r})Z_{t+r}\exp\left(\int_0^r q(X_s)ds - 1\right)\right]\right\} \\ &\to \frac{\partial}{\partial t}v(t,x) + \mathbb{E}^x[f(X_t)Z_t]q(x). \end{aligned}$$

Hence, we have the left part as:

$$Av = \frac{\partial}{\partial t}v + qv.$$

(Uniqueness:) Here, we consider the new generator as:

$$\hat{A}w(t,x) = -\frac{\partial w}{\partial t} + Aw - q(w),$$

with the new process $H_t = (s - t, X_t^{0,x}, Z_t)$. $\square$

**Remark VII.3.2.** When $q \equiv p$, we can consider $A = \frac{1}{2}\Delta$ as the Laplacian, so:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{1}{2}\Delta u + \rho u, & \text{for } t > 0, x \in \mathbb{R}^n, \\ u(0,x) = f(x), & \text{for } x \in \mathbb{R}^n. \end{cases}$$

By the Feynman-Kac theorem, we have:

$$\begin{aligned} u(t,x) &= \mathbb{E}^x \left[ \exp\left( \int_0^t \rho ds \right) f(B_t) \right] \\ &= \mathbb{E}[\exp(\rho t) f(B_t^x)] = e^{\rho t} \frac{1}{(2\pi t)^{n/2}} \int_{\mathbb{R}} \exp\left( -\frac{(x-y)^2}{2t} \right) f(y) dy. \end{aligned}$$

┘

**Example VII.3.3. Feynman-Kac for Laplace Equation.**

Suppose $U$ is a domain with smooth boundary. Consider the PDE:

$$\begin{cases} -\frac{1}{2}\Delta u + cu = f, & \text{in } U^\circ \\ u = 0, & \text{on } \partial U. \end{cases}$$

We have the Feynman-Kac representation as:

$$u(x) = \mathbb{E}\left[ \int_0^{\tau_x} f(X_t) \exp\left( -\int_0^t c\big(X(s)\big) ds \right) dt \right],$$

and so we have $X_t = B_t + x$ for $x \in U^\circ$, and $\tau_x$ as the first hitting time of $X_t$ on $\partial U$. ┘

**Remark VII.3.4. Feynman-Kac Backward Equation.**

Let $f \in C_0^2(\mathbb{R}^n)$ and $q \in C(\mathbb{R}^n)$, and assume that $q$ is lower bounded. Consider:

$$\begin{cases} \frac{\partial w}{\partial t} + Aw = cw + f, & \text{for } x \in \mathbb{R}^n, t \in [0,T], \\ w(x,T) = \phi(x), & \text{for } x \in \mathbb{R}^n. \end{cases}$$

Then, the Feynman-Kac indicates that:

$$w(x,t) = \mathbb{E}^{x,t}\left[ \phi(X_t) \exp\left( -\int_t^T q(X_s) ds \right) \right] - \mathbb{E}^{x,t}\left[ \int_t^T f(X_s) \exp\left( -\int_t^s q(X_u) du \right) ds \right].$$

In particular, this has more application in finance, related to price option. ┘

In particular, Feynman-Kac is the generalization of the Kolmogrov property, and it is Kolmogrov when $q \equiv 0$.

## VII.4 The Martingale Problem

Consider the Itô diffusion $X = \{X_t\}_{t \geq 0}$ that:

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t.$$

The Itô generator is given by $A$ and we have:

$$f(X_t) - f(x) = \int_0^t Af(X)sds + \int_0^t \nu f^\intercal(X_s)\sigma(X_s)dB_s.$$

Note that the generator is:

$$Au(x) = \sum_i b_i(x)\frac{\partial u}{\partial x_i}(x) + \frac{1}{2}\sum_{i,j}(\sigma\sigma^\intercal)_{i,j}\frac{\partial^2 u}{\partial x_i \partial x_j}.$$

Here, we can define:

$$M_t = f(X_t) - \int_0^t Af(X_s)ds = f(x) - f(x) - \int_0^t \nabla f^\intercal(X_s)\sigma(X_s)dB_s.$$

Since the Itô integrals are martingales, we have for all $s > t$ that:

$$\mathbb{E}^x[M_s \mid \mathcal{F}_t^{(m)}] = M_t.$$

Here, $\mathcal{F}_t^{(m)} = \sigma(\{B_s : s \leq t\})$. Moreover, if we consider $\mathcal{M}_t = \sigma(\{X_s : s \leq t\})$, then $\mathcal{M}_t \subset \mathcal{F}_t^{(m)}$.

It follows that:

$$\mathbb{E}^x[M_s \mid \mathcal{M}_t] = \mathbb{E}^x\left[\mathbb{E}^x[M_s \mid \mathcal{F}_t^{(m)}] \mid \mathcal{M}_t\right] = \mathbb{E}^x[M_t \mid \mathcal{M}_t] = M_t,$$

by using the towering property, the martingale property of $M_s$ with respect to $\mathcal{F}_t^{(m)}$, and measurability of the function.

**Theorem VII.4.1. Martingale with respect to Itself.**

If $X_t$ is an Itô diffusion in $\mathbb{R}^n$ with generator $A$, then for all $f \in C_0^2(\mathbb{R}^n)$, the process:

$$M_t = f(X_t) - \int_0^t Af(X_s)ds$$

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

Recall that if we identify each $\omega \in \Omega$ with the function:

$$\omega_t = \omega(t) = X_t^x(\omega),$$

we can set the probability space $(\Omega, \mathcal{M}, Q^x)$ is identified with $\left((\mathbb{R}^n)^{[0,\infty)}, \mathcal{B}, \tilde{Q}^x\right)$, and we can reformulate Theorem VII.4.1.

**Theorem VII.4.2. Generealization of Martingle with respect to Itself with Measure Space.**

If $\tilde{Q}^x$ is a probability measure $\mathcal{B}$ induced by the law $Q^x$ of an Itô diffusion $X_t$, then for all $f \in C_0^2(\mathbb{R}^n)$, the process:

$$M_t = f(X_t) - \int_0^t Af(X_s)ds$$

is a $\tilde{Q}^x$-martingale with respect to the $\sigma$-algebra $\mathcal{B}_t = \sigma(\{(\mathbb{R}^n)^{[0,t]}\})$.

**Definition VII.4.3. Martingale Problem.**

Let $L$ be a semi-elliptic differential operator of the form:

$$L = \sum_i b_i \frac{\partial}{\partial x_i} + \sum_{i,j} a_{i,j} \frac{\partial^2}{\partial x_i \partial x_j},$$

where the coefficients $b_i$, $a_{i,j}$ are locally bounded Borel measurable function $\mathbb{R}^n$. Then, we say a probability measure $\tilde{\mathbb{P}}^x$ on $\big((\mathbb{R}^n)^{[0,\infty)}, \mathcal{B}\big)$ solves the **martingale problem** for $L$ if the process:

$$\begin{cases} M_t = f(\omega_t) - \int_0^t Lf(\omega_s) ds \text{ almost surely with respect to } \tilde{\mathbb{P}}^x, \\ M_0 = f(x) \end{cases}$$

is a $\tilde{\mathbb{P}}^x$-martingale. ⌟

**Remark VII.4.4.**

- The $\tilde{Q}^x$ solves the martingale problem for the operator $A$.
- When $X_t$ is a weak solution to the SDE:

  $$dX_t = b(X_0)dt + \sigma(X_t)dB_t,$$

  then $\tilde{Q}^x$ solves the martingale problem associated with $A$ if and only if $X_t$ is a weak solution of the above Itô diffusion.
- (Stroock & Varadhan, 1979; Rogers & Williams, 1987). $\tilde{Q}^x$ is the unique solution of the martingale problem for the operator $L$ given by:

  $$L = \sum_i b_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j} (\sigma\sigma^\intercal)_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}.$$

- The Lipschitz-continuity of the coefficient of $L$ is not necessary for the uniqueness. ⌟

## VII.5 Itô Process and Diffusion

The question is now posed:

When is an Itô process a diffusion?

**Example VII.5.1. The Bessel Process.**

The process:

$$R_t(\omega) = |B(t,\omega)| = \left( \sum_{i=1}^n |B_i(t,\omega)|^2 \right)^{1/2}$$

such that the equation is:

$$dR_t = \sum_{i=1}^{n} \frac{B_i}{R_t} dB_t + \frac{n-1}{2R_t} dt.$$

The process is a Itô process as it is with respect $dB_i$ and $dt$.
In terms of the diffusion theory, this does not seem like an Itô diffusion, since the $dB_t$ part function is with respect to $B_t$ and $R_t$.
However, we define:

$$Y_t = \int_0^t \sum_{i=1}^{n} \frac{B_i}{|B|} dB_t \overset{d}{\sim} \text{1-dimensional } \tilde{B}_t.$$

Hence, we have:

$$dR_t = d\tilde{B}_t + \frac{n-1}{R_t} dt.$$

⌟

In this case, we can associate $R_t$ with a generator:

$$Af(x) = \frac{1}{2} f''(x) + \frac{n-1}{2x} f'(x).$$

**Theorem VII.5.2. Itô Process and Brownian Motion.**

An Itô process:

$$dY_t = v dB_t, \qquad Y_0 = 0 \qquad \text{with } v(t,\omega) \in \mathcal{V}_{\mathcal{H}}^{n\times m}$$

coincides in law with $n$-dimensional Brownian motion if and only if $vv^\intercal(t,\omega) = I_n$ for almost all $(t,\omega)$-$dt \times d\mathbb{P}$.

Up to here, we have only identified with a Brownian motion, not with an Itô diffusion yet.

**Theorem VII.5.3. Itô Process and Diffusion.**

Let $X_t$ be an Itô diffusion given by:

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dB_t, & \text{for } b \in \mathbb{R}^n, \sigma \in \mathbb{R}^{n\times n}, \\ X_0 = x, \end{cases}$$

and let $Y_t$ be an Itô process given by:

$$dY_t = u(t,\omega)dt + v(t,\omega)dB_t.$$

Then $\{X_t\} \simeq \{Y_t\}$ if and only if $b(Y_t^x) = \mathbb{E}^x[u(t,\cdot) \mid \sigma(\{Y_s : s \le t\})]$ and $vv^\intercal(t,x)(Y_t^x)$.

Note that here, we want to fundamentally have $u(t,\omega) \to b(Y_t)$ and $v(t,\omega) \to \sigma(Y_t)$.

**Remark VII.5.4.** Let $\{Y_t\}$ be Itô process as above, then there exists some $\mathcal{N}_t := \sigma(\{Y_s : s \le t\})$-adapted process $w(t,\omega)$ such that:

$$vv^\intercal(t,\omega) = w(t,\omega).$$

Hence, this explains why the second expectation is not conditional, as it is adapted. ⌟

In general, however, $u(t, \cdot)$ and $v(t, \cdot)$ are not measurable with respect to $\mathcal{N}_t$.

## VII.6 The Girsanov Theorem

The main point of Girsanov theorem is that the diffusion does not have that much of an impact.

**Theorem VII.6.1. Lêvy Characterization of Brownian Motion.**

Let $X(t) = \big(X_1(t), X_2(t), \cdots, X_n(t)\big)$ be a continuous stochastic process on $(\Omega, \mathcal{H}, \mathbb{Q})$ with values in $\mathbb{R}^n$. Then the following are equivalent:

1. $X(t)$ is a Brownian motion with respect to $Q$.
2. $X(t)$ is a martingale with respect to $\mathbb{Q}$ and $X_i(t) - X_j(t) - \delta_{i,j}t$ is a martingale with respect to $\mathbb{Q}$ for all $i, j \in \{1, 2, \cdots, n\}$.

Then, we consider the following abstraction of Bayes' rule.

**Proposition VII.6.2. Conditional Bayes' Rule.**

Let $\mu$ and $\nu$ be two probability measure on $(\Omega, \mathcal{G})$ such that:

$$d\nu(\omega) = f(\omega)d\mu(\omega)$$

for some $f \in L^1(\mu)$, *i.e.*, $\int_\Omega f d\mu = 1$.
Let $X$ be a random variable on $(\Omega, \mathcal{G})$ such that:

$$\mathbb{E}_\nu[|X|] = \int_\Omega |X(\omega)| f(\omega) d\mu(\omega) < +\infty.$$

Let $\mathcal{H}$ be a $\sigma$-algebra. Then:

$$\mathbb{E}_\nu[X \mid \mathcal{H}] \cdot \mathbb{E}[f \mid \mathcal{H}] = \mathbb{E}_\mu[fX \mid \mathcal{H}].$$

Note that is $\mathbb{E}_\mu[f \mid \mathcal{H}]$ is nonzero, we have:

$$\mathbb{E}_\nu[X \mid H] = \frac{\mathbb{E}_\mu[fX \mid \mathcal{H}]}{\mathbb{E}_\mu[f \mid \mathcal{H}]}.$$

*Proof.* Here, we have:

$$\nu(d\omega) = f(\omega)\mu(d\omega).$$

If $\mathcal{H}$ is $\sigma$-algebra, then for all $H \in \mathcal{H}$:

$$\int_H \mathbb{E}_\mu[f \mid \mathcal{H}]d\mu = \int_H f d\mu = \nu(H),$$

hence we can equivalently say that $\nu \mid_{\mathcal{H}} = \mathbb{E}_\mu[f \mid \mathcal{H}]\mu \mid_{\mathcal{H}}$.
Hence, we have:

$$\mathbb{E}_\nu[\mathbb{1}_H X] = \mathbb{E}_\mu[\mathbb{1}_H f X] = \mathbb{E}_\mu\left[\mathbb{1}_H \mathbb{E}_\mu[fX \mid \mathcal{H}]\right] = \mathbb{E}_\mu\left[\mathbb{1}_H \frac{\mathbb{E}_\mu[fX \mid \mathcal{H}]}{\mathbb{E}_\mu[X \mid \mathcal{H}]} \cdot \mathbb{E}_\mu[f \mid \mathcal{H}]\right] = \mathbb{E}_\nu\left[\mathbb{1}_H \mathbb{E}_\mu[fX \mid \mathcal{H}]\right]$$

□

**Definition VII.6.3. Absolutely Continuous Measure.**

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be a filtered probability space. Fix $T > 0$ and let $\mathbb{Q}$ be another probability measure on $\mathcal{F}_T$, we say $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P} \mid_{\mathcal{F}_T}$, denoted $\mathbb{Q} \ll \mathbb{P}$, if:

$$\mathbb{P}(H) = 0 \Longrightarrow \mathbb{Q}(H) = 0 \text{ for all } H \in \mathcal{F}_T.$$

⌟

By **Radon-Nikodym theorem**, there exists a $\mathcal{F}_T$-measurable random variable $Z_T(\omega) \geq 0$ such that:

$$d\mathbb{Q}(\omega) = Z_T(\omega)d\mathbb{P}(\omega) \text{ on } \mathcal{F}_T \iff \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_i} = Z_T.$$

**Proposition VII.6.4. Weak Potential Converse of Girsanov Theorem.**

Suppose $\mathbb{Q} \ll \mathbb{P} \mid_{\mathcal{F}_T}$, and $\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_T$ on $\mathcal{F}_T$, then $\mathbb{Q} \mid_{\mathcal{F}_t} \ll \mathbb{P} \mid_{\mathcal{F}_t}$ for all $t \in [0, T]$.
We define:

$$Z_t = \frac{d\mathbb{Q} \mid_{\mathcal{F}_t}}{d\mathbb{P} \mid_{\mathcal{F}_t}} =: \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t},$$

then $Z_t$ is a martingale with respect to $\mathcal{F}_t$ and $\mathbb{P}$.

*Proof.* For any $F \in \mathcal{F}_t$, then:

$$\begin{aligned}\mathbb{E}_\mathbb{P}\left[\mathbb{1}_F \mathbb{E}_\mathbb{P}[Z_T \mid \mathcal{F}_t]\right] &= \mathbb{E}_\mathbb{P}\left[\mathbb{E}_\mathbb{P}[\mathbb{1}_F Z_T \mid \mathcal{F}_t]\right] = \mathbb{E}_\mathbb{P}[\mathbb{1}_F Z_T] \\ &= \mathbb{E}_\mathbb{Q}[\mathbb{1}_F] = \mathbb{E}_\P[\mathbb{1}_F Z_t].\end{aligned}$$

Hence, we have $Z_t$ as a martingale and:

$$\mathbb{E}_\mathbb{P}[Z_t \mid \mathcal{F}_t] = Z_t$$

almost surely on $\mathbb{P} \mid_{\mathcal{F}_t}$. □

**Theorem VII.6.5. Girsanov Theorem I.**

Let $Y(t) \in \mathbb{R}^n$ be an Itô process:

$$dY(t) = a(t, \omega)dt + dB(t) \text{ for } t \leq T \text{ and } Y(0) = 0.$$

We define:

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

Assume that $M(t)$ is a martingale with respect to $\mathcal{F}_t^{(m)}$ and $\mathbb{P}$. We define a probability measure $\mathbb{Q}$ on $\mathcal{F}_T^{(m)}$ with:

$$dQ(\omega) = M_T(\omega)d\mathbb{P}(\omega).$$

Then, $\mathbb{Q}$ is a probability measure on $\mathcal{F}_T^{(m)}$, and $Y_t$ is an $n$-dimensional Brownian motion with respect to $\mathbb{Q}$ for $0 \le t \le T$, that is, on $(\Omega, \mathcal{F}, \mathcal{N}_t, \mathbb{Q})$.

**Remark VII.6.6.** Girsanov theorem states that for all $F_1, \cdots, F_k \subset \mathbb{R}^n$ and $t_1, \cdots, t_k \le T$, we have:

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

Particularly, we consider:

$$L_t := \left.\frac{d\mathbb{Q}_Y}{s\mathbb{P}_B}\right|_{\mathcal{F}_t} = \mu_t = \exp\left[-\int_0^t a(s)dY_s + \frac{1}{2}a^2(s)ds\right],$$

which is called a likelihood process. ⌟

**Example VII.6.7. Maximum Likelihood Estimate for the OU process.**

Consider the OU process:

$$dX_t = -\alpha X_t dt + dW_t,$$

where we have $\alpha$ unknown. Given $\{X_t\}_{t\in[0,T]}$, how do we estimate $\alpha$? By Girsanov, we can have the **maximum likelihood** (MLE). By the Girsanov, we have:

$$M_t = \exp\left[-\alpha\int_0^T X_t dX_t - \frac{\alpha^2}{2}\int_0^T X_t^2 dt\right]$$

Then, we consider the MLE:

$$L_t = \left.\frac{d\mathbb{P}_X}{d\mathbb{P}_W}\right|_{\mathcal{F}_t} = -\alpha\int_0^T X_t dX_t - \frac{\alpha^2}{2}\int_0^T X_t^2 dt.$$

Then, we have that:

$$\frac{\partial L_t}{\partial \alpha} = 0,$$

hence leading to:

$$\hat{\alpha} = -\frac{\int_0^T X_t dX_t}{\int_0^T X_t^2 dt} = \frac{\frac{1}{2}(X_T^2 - X_0^2 - T)}{\int_0^T X_t^2 dt}.$$

In fact, if we want to find $\alpha$, we have:

$$\alpha = \frac{-\int_0^T X_t dX_t + \int_0^T X_t dB_t}{\int_0^T X_t^2 dt}.$$

Hence, we consider here the right component as the unbiased part. ⌟

Another example could be the problem of unknown parameter for:

$$dX_t = \alpha dt + dW_t$$

with $\alpha$ being unknown. Given $\{X_t\}$, we want to estimate $\alpha$, and we have the solution like:

$$X_t = \alpha t + W_t.$$

**Proposition VII.6.8. Partual Converse of Girsanov.**

Suppose $\mathbb{Q} \ll |\mathbb{P}|_{\mathcal{F}_T}$ with $\frac{d\mathbb{Q}}{d\P} = Z_T$ on $\mathcal{F}_T$, then $\mathbb{Q}\big|_{\mathcal{F}_t} \ll \mathbb{P}\big|_{\mathcal{F}_t}$ for all $t \in [0,T]$ and $Z_t = \frac{d\mathbb{Q}\big|_{\mathcal{F}_t}}{d\mathbb{P}_{\mathcal{F}_t}}$ is a martingale with respect to $\mathcal{F}_t$ and $\mathbb{P}$.

Here, we give the proof of Girsanov theorem .

*Proof of.* Since $M_t$ is martingale, $\mathbb{E}[M_t] = \mathbb{E}\big[\mathbb{E}[M_T \mid \mathcal{F}_t]\big] = \mathbb{E}[M_t] = \mathbb{E}[M_T]$, then:

$$\mathbb{Q}(\Omega) = \mathbb{E}_{\mathbb{Q}}[1] = \mathbb{E}_{\mathbb{P}}[M_T] = 1.$$

Hence $\mathbb{Q}$ is a probability measure. Without loss of generality, we assume that $a(s,\omega)$ is bounded. We need to verify that:

1. $Y(t)$ is a martingale with respect to $\mathbb{Q}$, and
2. $Y_i(t) - Y_j(t) - \delta_{i,j}t$ is also martingale with respect to $\mathbb{Q}$.

To verify 1, we put $k(t) = \mu(t)Y(t)$ and use the Itô formula to obtain:

$$\begin{aligned} dK_i(t) &= M(t)dY_i(t) + Y_i(t)d\mu(t) + dY_i(t)d\mu(t) \\ &= M(t)\left[dB_i(t) - Y_i(t)\sum_{k=1}^{n} a_k(t)dB_k(t)\right] = M(t)Y^{(i)}(t)dB(t). \end{aligned}$$

Here, we have:

$$Y_j^{(i)}(t) = \begin{cases} -Y_i(t)a_j(t), & i \neq j, \\ 1 - Y_i(t)a_j(t), & i = j. \end{cases}$$

Hence, $K_i(t)$ is martingale with respect to $\mathbb{P}$.
Then, by the Bayes' rule, we get:

$$\mathbb{E}_{\mathbb{Q}}[Y_i(t) \mid \mathcal{F}_s] = \frac{\mathbb{E}[M(t)Y_i(t) \mid \mathcal{F}_s]}{\mathbb{E}_{\mathbb{P}}[M(t) \mid \mathcal{F}_s]} = \frac{\mathbb{E}[K_i(t) \mid \mathcal{F}_s]}{M(s)} = \frac{K_i(s)}{M(s)} = Y_i(s).$$

Hence, $Y_i(t)$ is a martingale with respect to $\mathbb{Q}$. □

**Remark VII.6.9.** Recall the Novikov condition is sufficient to guarantee that $\{M_t\}_{t \leq T}$ is a martingale:

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T a^2(s,\omega)ds\right)\right] < +\infty.$$

Also, note that since $M_t$ is a martingale, we have that:

$$M_T d\mathbb{P} = M_t d\mathbb{P} \text{ on } \mathcal{F}_t^{(n)}.$$

Hence, by Girsanov, we have that for all $F_1, \cdots, F_k \subset \mathbb{R}^n$ and $t_1, \cdots, t_k$, hence:

$$\mathbb{Q}[Y(t_1) \in F_1, \cdots, Y(t_k) \in F_k] = \mathbb{P}[B(t_1) \in F_1, \cdots, B(t_k) \in F_k],$$

which leads to:

$$\begin{aligned}\frac{d\mathbb{Q}_Y}{d\mathbb{P}_B} &= M_T \text{ on } \mathcal{F}_T^{(n)} = M_t \text{ on } \mathcal{F}_t^{(n)} \\ &= \exp\left[-\int_0^t a(s,\omega)dB_s - \frac{1}{2}\int_0^t a^2(s,\omega)ds\right] = \exp\left[-\int_0^t a(s,\omega)dY_s + \frac{1}{2}\int_0^t a^2(s,\omega)ds\right].\end{aligned}$$

┘

**End of April 28, 2025**