Stochastic Calculus

Felix Zhou[i]

August 5, 2023

[i]based on "Introduction to Stochastic Calculus with Applications" by Fima C. Klebaner

Contents

Part I

Background

Chapter 1

Preliminaries from Analysis

As a general setting throughout this document, we consider functions of time $f : [0, T] \to \mathbb{R}$.

We assume knowledge of the Lebesgue measure and integration.

1.1 Elementary Calculus

1.1.1 Differentiation

Unless stated otherwise, we consider functions $\mathbb{R} \to \mathbb{R}$.

For $t \in \mathbb{R}$ and function g, we write

$$\begin{aligned} \Delta t &:= t' - t \\ \Delta g(t) &:= g(t') - g(t). \end{aligned}$$

Recall that a function is *continuous* at t if

$$\Delta t \to 0 \implies \Delta g(t) \to 0.$$

Moreover, g is *differentiable* at t if

$$\lim_{\Delta t \to 0} \frac{\Delta g(t)}{\Delta(t)} = C$$

and we write $g'(t) := C$.

We know that differentiability implies continuity but the converse need not hold. This is intuitive since continuity only requires $\Delta g(t) \to 0$ whenever $\Delta t \to 0$, whereas differentiability also requires that $\Delta g(t)$ converges at least at the same rate as Δt.

Theorem 1.1.1 (Mean Value)
If f is continuous on $[a,b]$ and differentiable on (a,b), there is some $c \in (a,b)$ such that
$$f(b) - f(a) = f'(c)(b-a).$$

We write $C^k(X,Y)$ to denote the set of functions $f : X \to Y$ that are k-times differentiable with a continuous k-th derivative. We also write $C := C^0$ as a shorthand for continuous functions.

1.1.2 Right & Left-Continuous Functions

Recall that a function g is *right-continuous* if

$$\lim_{t \downarrow t_0} g(t) = g(t_0)$$

and vice versa for *left-continuous* functions.

As a shorthand, we write

$$\begin{aligned} g(t-) &:= \lim_{t' \uparrow t} g(t') \\ g(t+) &:= \lim_{t' \downarrow t} g(t'). \end{aligned}$$

Definition 1.1.2 (Jump Discontinuity)
A point t is a jump discontinuity if both $g(t+), g(t-)$ exist but are not equal.

Any other discontinuity is said to be of the *second kind*.

Theorem 1.1.3
A function $g : [a,b] \to \mathbb{R}$ can have at most countably many jump discontinuities.

Theorem 1.1.4
If $f : [a,b] \to \mathbb{R}$ is differentiable with a finite derivative $f'(t)$, then for all $t \in [a,b]$, $f'(t)$ is either continuous at t or has a discontinuity of the second kind.

Proof
If $f'(t+)$ exists, then

$$\begin{aligned} f'(t) &= \lim_{\Delta t \downarrow 0} \frac{f(t+\Delta) - f(t)}{\Delta} \\ &= \lim_{\Delta \downarrow 0, t<c<t+\Delta} f'(c) \\ &= f'(t+). \end{aligned}$$

Similarly, $f'(t-) = f'(t)$ if it exists.

The result follows. $\square$

1.1.3 Functions in Stochastic Calculus

We focus on *regular* functions, ie those without discontinuities of the second kind. The class $D = D[0,T]$ of right-continuous functions on $[0,T]$ with left limits are referred to as *CADLAG* functions. Note that $C \subseteq D$. Similarly the class of regular left-continuous functions are called *CAGLAD*.

In stochastic calculus, $\Delta g(t)$ usually stands for the size of the jump at t,

$$\Delta g(t) := g(t+) - g(t-).$$

This differs from the convention for standard calculus. We will clarify if it is unclear from context.

1.2 Variation of a Function

Definition 1.2.1 (Variation)
The *variation* of a function $g : [a,b] \to \mathbb{R}$ is given by

$$V_g([a,b]) = \sup \sum_{i=1}^{n} |g(t_i^n) - g(t_{i-1}^n)|$$

where the supremum is taken over all partitions

$$a = t_0^n < t_1^n < \cdots < t_n^n = b.$$

The sums in the definition above increase as new points are added. Thus

$$V_g[a,b] = \lim_{\delta_n \to 0} \sum_{i=1}^{n} |g(t_i^n) - g(t_{i-1}^n)|$$

where $\delta_n = \max_{i \in [n]}(t_i - t_{i-1})$.

If $V_g[a,b] < \infty$, then g is said to be a function of *finite variation* on $[a,b]$. If g is a function of $t \geq 0$, then the *variation function* of g is defined as

$$V_g(t) := V_g[0,t].$$

Note that $V_g(t)$ is non-decreasing in t. We say that $g : \mathbb{R}_+ \to \mathbb{R}$ is of *finite variation (FV)* if $V_g(t) < \infty$ for all $t \geq 0$.

Example 1.2.2

1) If $g(t)$ is increasing, then $V_g(t) = g(t) - g(0)$.
2) If $g(t)$ is decreasing, then $V_g(t) = g(0) - g(t)$.

Proposition 1.2.3 (Variation of a Differentiable Function)
Suppose $g \in C^1$ and g' is absolutely integrable over $[0,t]$. Then

$$V_g(t) = \int_0^t |g'(x)| ds.$$

Proof
The Riemann integral is the supremum of Riemann sums over partitions as $\delta_n \to 0$. Apply this definition with the mean value theorem. □

If a function $g : \mathbb{R}_+ \to \mathbb{R}$ only changes by jumps, we can write

$$g(t) = \sum_{0 \leq s \leq t} \Delta g(s).$$

Proposition 1.2.4 (Variation of a Pure Jump Function)
Suppose g is

(i) regular,

(ii) left or right-continuous,

(iii) and only changes by jumps.

Then

$$V_g(t) = \sum_{0 \leq s \leq t} |\Delta g(s)|.$$

Theorem 1.2.5 (Jordan Decomposition)
Any function $g : \mathbb{R}_+ \to \mathbb{R}$ of finite variation can be expressed as the difference of two increasing functions
$$g(t) = a(t) - b(t).$$
One such decomposition is given by
$$\begin{aligned} a(t) &= V_g(t) \\ b(t) &= V_g(t) - g(t). \end{aligned}$$

Proof
It is clear the difference of two increasing functions has finite variation.

Conversely, note $V_g(t) = a(t)$ is increasing. On the other hand, fix any $s < t$. We have
$$\begin{aligned} b(t) - b(s) &= V_g[0, t] - V_g[0, s] - g(t) + g(s) \\ &\geq V_g[s, t] - |g(t) - g(s)| && V_g[0, s] + V_g[s, t] = V_g[0, t] \\ &\geq 0. \end{aligned}$$
□

Another decomposition is given by
$$g(t) = \frac{1}{2}\left[V_g(t) + g(t)\right] - \frac{1}{2}\left[V_g(t) - g(t)\right].$$

Proposition 1.2.6
Let f, g be function so finite variation. The following are all of finite variation:
(a) $f + g$
(b) fg
(c) ${}^{f}/_{g}$ given $|g| \geq C \in \mathbb{R}$

Theorem 1.2.7
A finite variation function can have no more than countable discontinuities. Moreover, all discontinuities are jumps.

Proof
The result holds for monotone functions and FV functions are differences of monotone functions.
□

Theorem 1.2.8
Let $g \in C^1$ and suppose $g'(t)$ is absolutely integrable. Then g is of finite variation.

Proof
$V_g(t) = \int_0^t |g'(t)|dt < \infty$ for all t by assumption. □

Proposition 1.2.9
Suppose $g : [a, b] \to \mathbb{R}$ is continuous. Let $t_i := \frac{i}{N}, i = 0, 1, \ldots, N$ denote the uniform partition of size $N + 1$. Then

$$V_g[a, b] = \lim_{N\to\infty} \sum_{i=1}^{N} |g(t_i) - g(t_{i-1})| =: v_n.$$

Proof (Sketch[a])
Let N_1 be sufficiently large so that some partition P_1 approximates $V_g[a, b]$ with error $\varepsilon/2$. Take N_2 to be sufficiently large so that uniform continuity holds for $\varepsilon/4N_1$ and consider the uniform partition P_2 on N_2 points. The common refinement $P = P_1 \cup P_2$ clearly approximates $V_g[a, b]$ to error $\varepsilon/2$. Using uniform continuity, we can show that by removing the points P_1 from P to yield P_2, we introduces an error of at most $2N_1 \cdot \varepsilon/4N_1 = \varepsilon/2$.

This concludes the proof. □

[a] `https://math.stackexchange.com/a/3130591`

Theorem 1.2.10 (Banach Indicatrix)
Let $g(t)$ be a continuous function on $[a, b]$ and define

$$S(c) := \{\, t \in [a, b] : g(t) = c \,\}$$
$$s(c) := |S(t)|.$$

Then the variation of g is equal to

$$\int_{-\infty}^{\infty} s(a)da.$$

Proof (Sketch[a])
We can approximate $s(c)$ using a monotonically increasing sequence $s_k(c), k \geq 1$ of indicator functions. Partition $[a, b]$ into 2^k intervals $I_1^{(k)}, \ldots, I_{2^k}^{(k)}$ of length $2^{-k}[b - a]$ with

endpoints $a = t_0^{(k)} < t_1^{(k)} < \cdots < t_{2^k}^{(k)} = b$. Define

$$s_k := \sum_{i=1}^{2^k} \mathbb{1}\{*\} f(I_i^{(k)}).$$

It can be shown that $s_k \uparrow s$ and also

$$\int s(a) = \lim_k \int s_k(a) \leq V_g[a,b].$$

Fix $\varepsilon > 0$. By continuity and the previous proposition, for sufficiently large k,

$$\int s_k(a) \geq \sum_{i=1}^{2^k} |g(t_i^{(k)}) - g(t_{i-1}^{(k)})| \geq V_g[a,b] - \varepsilon.$$

The result follows. □

[a] https://math.stackexchange.com/a/144832

1.2.1 Continuous & Discrete Parts of a Function

Let $g : \mathbb{R}_+ \to \mathbb{R}$ be right-continuous and increasing. It has at most countably many jumps and the sum of jumps is finite over finite time intervals. Define the discontinuous part g^d of g by

$$g^d(t) := \sum_{s \leq t} [g(s) - g(s-)] = \sum_{0<s\leq t} \Delta g(s)$$

and the continuous part g^c of g by

$$g^c(t) = g(t) - g^d(t).$$

By construction, $g = g^c + g^d$ with g^d only changing by jumps and g^c being continuous.

Since finite variation functions are differences of increasing functions, the decomposition extends for functions of finite variation. Note the decomposition is unique up to constants. Indeed, if $g = h^c + h^d$ then

$$h^c - g^c = g^d - h^d,$$

implying that $h^d - g^d$ is continuous. Hence h^d and g^d have the same set of jump points and so $h^d - g^d = c$ for some constant c.

1.2.2 Quadratic Variation

Definition 1.2.11 (Quadratic Variation)
Let $g : \mathbb{R}_+ \to \mathbb{R}$. Its *quadratic variation* over $[0, t]$ is given by

$$[g](t) := \lim_{\delta_n \to 0} \sum_{i=1}^{n} [g(t_i^n) - g(t_{i-1}^n)]^2$$

when it exists. The limit is taken over partitions with decreasing maximum width δ_n.

We can extend the notion of variation to Φ-variation where $\Phi : \mathbb{R}_+ \to \mathbb{R}+$ is monotonically increasing. The Φ-variation of g on $[0, t]$ is

$$V_\Phi[g] := \sup \sum_{i=1}^{n} \Phi(|g(t_i^n)| - g(t_{i-1})^n)$$

where the supremum is taken over all partitions (not just ones with decreasing width).

We note that the definition of quadratic variation is different to the Φ variation definition with $\Phi(u) := u^2$. In our setting, the limit is taken over shrinking partitions and not all possible partitions. The definition is equivalent for $\Phi(u) = u$ due to the triangle inequality but not in general.

We rarely encounter quadratic variation in calculus despite its importance in stochastic calculus since smooth functions have zero quadratic variation.

Theorem 1.2.12
If g is continuous and of finite variation, then its quadraatic variation is zero.

Proof
We have

$$\begin{aligned}
[g](t) &= \lim_{\delta_n \to 0} \sum_{i=1}^{n} [g(t_i^n) - g(t_{i-1}^n)]^2 \\
&\leq \lim_{\delta_n \to 0} \max_i |g(t_i^n) - g(t_{i-1}^n)| V_g(t).
\end{aligned}$$

But continuity on compact sets imply uniform continuity and so the limit is zero. $\square$

We refer to functions with zero quadratic variation and finite variation as functions of *zero energy*.

Definition 1.2.13 (Quadratic Covariation)
The quadratic covariation of $f, g : \mathbb{R}_+ \to \mathbb{R}$ on $[0, t]$ is given by the following limit when it exists:

$$[f, g](t) := \lim_{\delta_n \to 0} \sum_{i=1}^{n} [f(t_i^n) - f(t_i^n)][g(t_i^n) - g(t_{i-1}^n)].$$

The limit is taken over shrinking partitions.

Theorem 1.2.14
If f is continuous and g is of finite variation, then $[f, g](t) = 0$.

Theorem 1.2.15 (Polarization Identity)
Let f, g be such that their covariation is defined.

$$[f, g](t) = \frac{1}{2} \left([f + g, f + g](t) - [f, f](t) - [g, g](t) \right).$$

Clearly covariation is symmetric and it follows from the polarization identity that it is bilinear. By definition, the quadratic variation function is non-decreasing and hence of finite variation. This extends to quadratic covariation by the polarization identity.

1.3 Riemann-Stieltjes Integral

The Riemann-Stieltjes integral is an integral of the form

$$\int_a^b f(t) dg(t)$$

where g is of finite variation. Note it suffices to define the integral with respect to monotone functions as functions of finite variation are differences of monotone functions.

Definition 1.3.1 (Stieltjes Integral)
The Stieltjes integral of $f : \mathbb{R} \to \mathbb{R}$ with respect to $g : \mathbb{R} \to \mathbb{R}$ monotone over $[a, b]$ is defined as

$$\int_a^b f dg = \int_a^b f(t) dg(t) = \lim_{\delta_n \to 0} f(\xi_i^n)[g(t_i^n) - g(t_{i-1}^n)].$$

We can interpret the Riemann-Stieltjes integral as a Lebesgue integral

$$\int_{(0,t]} dg(s) = g(t) - g(0).$$

We then have

$$\int_{(0,t)} dg(s) = g(t-) - g(0).$$

If $g'(t)$ exists and $g(t) = g(0) + \int_0^t g'(s)ds$, it is possible to show that

$$\int_a^b f(t)dg(t) = \int_a^b f(t)g'(t)dt,$$

which is similar to the notion of a Radon-Nikodym derivative for absolutely continuous measures.

If $g(t) = \sum_{k=a}^{\lfloor t \rfloor} h(k)$, then

$$\int_a^b f(t)dg(t) = \sum_{k=a+1}^{b} f(k)h(k).$$

Let $g : \mathbb{R} \to \mathbb{R}$ be a function of finite variation and decompose it as

$$\begin{aligned} g &= a - b \\ a &= V_g \\ b &= V_g - g. \end{aligned}$$

Definition 1.3.2 (Stieltjes Integral)
If

$$\int_0^t |f(s)||dg(s)| := \int_0^t |f(s)|dV_g(s) < \infty,$$

then we say f is *Stieltjes integrable* with respect to g and its integral is defined by

$$\int_{(0,t]} f(s)dg(s) := \int_{(0,t]} f(s)da(s) - \int_{(0,t]} f(s)db(s).$$

We write

$$\int_a^b f(s)dg(s) := \int_{(a,b]} f(s)dg(s).$$

If f is Riemann-Stieltjes integrable with respect to g, then the *variation* of the integral is given by

$$V(t) := \int_0^t |f(s)||dg(s)| = \int_0^t |f(s)|dV_g(s).$$

In stochastic calculus, we may need to integrate with respect to functions of infinite variation. It can be shown that such integrals cannot be defined as a usual limit of approximating sums.

Theorem 1.3.3
Let δ_n denote the width of the largest interval in a partition of $[a, b]$. If

$$\lim_{\delta_n \to 0} \sum_{i=1}^{n} f(t_{i-1}^n)[g(t_i^n) - g(t_{i-1}^n)]$$

exists for any continuous function f, then g must be of finite variation on $[a, b]$.

1.3.1 Lebesgue-Stieltjes Integral

This interpretation is due to the appendix of a book[i].

Definition 1.3.4 (Borel Measure)
A *Borel measure* on $\mathbb{R}$ is a non-negative set function μ defined for all Borel sets of $\mathbb{R}$ such that

(i) $\mu(\varnothing) = 0$

(ii) $\mu(I) < \infty$ for every bounded interval I

(iii) $\mu(\cup_{i=1}^{\infty})B_i) = \sum_{i=1}^{\infty} \mu(B_i)$ for disjoint Borel sets B_i's

Theorem 1.3.5
The following hold.

(a) Let μ be a Borel measure on $\mathbb{R}$ and $G : \mathbb{R} \to \mathbb{R}$ be satisfy $G(b) - G(a) = \mu(a, b]$. Then G is right-continuous and increasing.

(b) Let $G : \mathbb{R} \to \mathbb{R}$ be right-continuous and increasing. Define $\mu(a, b] := G(b) - G(a)$. There is a unique extension of μ to a Borel measure on $\mathbb{R}$.

[i] `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118150672.app1`

Proposition 1.3.6
Let $G : \mathbb{R} \to \mathbb{R}$ be right-continuous and increasing, with μ being its associated Borel measure.

(a) $\mu(a, b) = G(b-) - G(a)$

(b) $\mu[a, b] = G(b) - G(a-)$

(c) $\mu[a, b) = G(b-) - G(a-)$

(d) $\mu\{a\} = G(a) - G(a-)$

(e) G is continuous at a if and only if $\mu\{a\} = 0$.

For a right-continous increasing function G, we can then equivalently define

$$\int_B f dG := \int_B f d\mu$$

where B is a Borel set and μ is the associated Borel measure of G.

Example 1.3.7
If $g : \mathbb{R} \to \mathbb{R}$ is right-continuous, increasing, and also differentiable on $\mathbb{R}$ except at points in a countably infinite set $x_1, x_2, \ldots,$

$$\int_0^t f(x) dg(x) = \int_0^t f(x) g(x) dx + \sum_{n: 0 < x_n \leq t} f(x_n) \Delta g(x_n).$$

Since finite variation functions are differences of increasing functions, we can extend the definition to these functions as usual. It follows from this decomposition that all standard results from Lebesgue integration, such as convergence theorems, and Fubini's theorem for iterated integrals over product spaces, hold for $\int f dg$ when f is Borel measurable and g is of bounded variation on finite intervals.

1.3.2 Integration by Parts

Theorem 1.3.8
Let $f, g : \mathbb{R} \to \mathbb{R}$ be right-continuous functions of finite variation. Then

$$\begin{aligned} f(b)g(b) - f(a)g(a) &= \int_a^b f(s-) dg(s) + \int_a^b g(s) df(s) \\ &= \int_a^b f(s-) dg(s) + \int_a^b g(s-) df(s) + \sum_{a < s \leq b} \Delta f(s) \Delta g(s). \end{aligned}$$

Remark 1.3.9 In the case that f is continuous, we recover the familiar integration by parts formula.

$$f(b)g(b) - f(a)g(a) = \int_a^b f(s)dg(s) + \int_a^b g(s)df(s).$$

Proof
By Fubini's theorem,

$$\begin{aligned}
&[f(b) - f(b)][g(a) - g(a)] \\
&= \int_a^b \int_a^b df(x)dg(y) \\
&= \int_a^b \int_a^b \mathbb{1}_{\{x<y\}} df(x)dg(y) + \int_a^b \int_a^b \mathbb{1}_{\{x\geq y\}} dg(y)df(x) \\
&= \int_a^b [f(y-) - f(a)]dg(y) + \int_a^b [g(x) - g(a)]df(x) \\
&= \int_a^b f(y-)dg(y) - f(a)[g(b) - g(a)] + \int_a^b g(x)df(x) - g(a)[f(b) - f(a)] \\
&= \int_a^b f(y-)dg(y) + \int_a^b g(x)df(x).
\end{aligned}$$

This shows the first equality. The second equality follows by the decomposition

$$\begin{aligned}
\int_a^b \Delta g(s)f(s) &= \int_a^b \Delta g(s)df^c(s) + \sum_{a<s\leq b} \Delta g(s)\Delta f(s) \\
&= \sum_{a<s\leq b} \Delta g(s)\Delta f(s).
\end{aligned}$$

□

Example 1.3.10
Let g be of finite variation with $g(0) = 0$. Then by integration by parts,

$$\int_0^t g(s-)dg(s) = \frac{g^2(t)}{2} - \frac{1}{2}\sum_{s\leq t}(\Delta g(s))^2.$$

On the other hand,

$$\begin{aligned}
\int_0^t g(s)dg(s) &= g^2(t) - \int_0^t g(s-)dg(s) \\
&= \frac{g^2(t)}{2} + \frac{1}{2}\sum_{s\leq t}(\Delta g(s))^2.
\end{aligned}$$

Thus it follows that
$$\int_0^t g(s-)dg(s) \leq \frac{g^2(t)}{2} \leq \int_0^t g(s)dg(s).$$

Remark 1.3.11 In the case that g is continuous, we have the identity

$$\int_0^t g(s)dg(s) = \frac{g^2(t)}{2}.$$

In particular, if $F(t) := \int_0^t f(s)ds$, then

$$\begin{aligned}\frac{1}{2}\int_0^t \int_0^t f(u)f(v)dudv &= \frac{1}{2}\left(\int_0^t f(s)ds\right)^2 \\ &= \frac{1}{2}F^2(t) \\ &= \int_0^t F(s)f(s)ds \\ &= \int_0^t \int_0^s f(u)f(s)duds.\end{aligned}$$

1.3.3 Change of Variables

Theorem 1.3.12
Let $f \in C^1$ and g be right-continuous and have finite variation.

$$\begin{aligned}&f(g(t)) - f(g(0)) \\ &= \int_0^t f'(g(s-))dg(s) \\ &\quad + \sum_{0<s\leq t} \left[f(g(s)) - f(g(s-)) - f'(g(s-))[g(s) - g(s-)]\right].\end{aligned}$$

If g is continuous,

$$\begin{aligned}f(g(t)) - f(g(0)) &= \int_0^t f'(g(s))dg(s) \\ &= \int_{g(0)}^{g(t)} f'(u)du.\end{aligned}$$

Example 1.3.13
Take $f(x) = x^2$. We have

$$\begin{aligned}
&g^2(t) - g^2(0) \\
&= 2\int_0^t g(s-)dg(s) \\
&\quad + \sum_{0<s\le t} [f(g(s)) - f(g(s-)) - f'(g(s-))[g(s) - g(s-)]]
\end{aligned}$$

$$\begin{aligned}
&\sum_{0<s\le t} [f(g(s)) - f(g(s-)) - f'(g(s-))[g(s) - g(s-)]] \\
&= \sum_{0<s\le t} \left[g^2(s) - g^2(s-) - 2g(s-)[g(s) - g(s-)]\right] \\
&= \sum_{0<s\le t} \left[g^2(s) + g^2(s-) - 2g(s-)g(s)\right] \\
&= \sum_{0<s\le t} [\Delta g(s)]^2.
\end{aligned}$$

1.4 Taylor's Theorem

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *differentiable* at $x \in \mathbb{R}^n$ provided that there is a vector $\boldsymbol{\nabla} f(x) \in \mathbb{R}^n$ such that

$$\Delta f(x) = \langle \boldsymbol{\nabla} f(x), \Delta x\rangle + o(\|\Delta(x)\|).$$

If f is differentiable at x, then all partial derivatives necessarily exist. It suffices to have continuous partial derivatives in order to have differentiability.

Let $f \in C^2(\mathbb{R}^n\mathbb{R})$. Then the second-order Taylor expansion is given by

$$\Delta f(x) = \langle \boldsymbol{\nabla} f(x), dx\rangle + \frac{1}{2}(dx)^T \boldsymbol{\nabla}^2 f(x + \theta\Delta x)(dx)$$

where $\boldsymbol{\nabla}^2 f(x + \theta\Delta x)$ denotes the Hessian matrix at some mid point in $[x, x + \Delta x]$.

1.4.1 Differentials & Integrals

We write the *differential* $df(t)$ of a differentiable function f as the largest term in its Taylor expansion.

Thus if $dx := \Delta x$, then

$$f(x + dx) - f(x) = f'(t)dt + o(|dt|)$$

and so $df(t) = f'(t)dt$.

By the *chain rule*, if both f, g are differentiable, then $f(g(x))$ is also differentiable with

$$df(g(t)) = f'(g(t))g'(t)dt = f'(g(t))dg(t).$$

The main relationship between integral and differential calculus is the fundamental theorem of calculus:

Theorem 1.4.1
If f is differentiable on $[a, b]$ and f' is integrable on $[a, b]$, then

$$f(b) - f(a) = \int_a^b f'(s)ds.$$

For differentiable functions, differential equations of the form

$$df(t) = \varphi(t)dw(t)$$

can be written as an integral equation

$$f(t) = f(0) + \int_0^t \varphi(s)dw(s).$$

In stochastic calculus, stochastic differentials do not formally exist as the random functions $w(t)$ are not differentiable at any point. By introducing a stochastic integral, stochastic differential equations are defined as solutions to these stochastic integral solutions.

1.5 Other Results

1.5.1 Lipschitz & Hölder Continuity

Lipschitz and Hölder conditions describe subclasses of continuous functions. They appear as conditions on the coefficients in the results of the existence and uniqueness of solutions of ordinary and stochastic differential equations.

Definition 1.5.1 (Hölder Continuity)
$f : I \subseteq \mathbb{R} \to \mathbb{R}$ satisfies a Hölder condition (Hölder continuous) of order $\alpha \in (0, 1]$ over the interval I if there is a constant $K > 0$ so that for every $x, y \in I$,

$$|f(x) - f(y)| \leq K|x - y|^\alpha.$$

A *Lipschitz condition* is a *Hölder condition* with $\alpha = 1$.

We say that $f : \mathbb{R} \to \mathbb{R}$ is *smooth* on $[a, b]$ if it has a continuous derivative f' on (a, b) and the limits $f'(a+), f'(b-)$ exist.

f is *piecewise continuous* on $[a, b]$ if it is continuous on $[a, b]$ except on a finite number of points at which both left and right limits exist.

We say f is *piecewise smooth* on $[a, b]$ if it is piecewise continuous on $[a, b]$ and f' exists and is also piecewise continous on $[a, b]$.

1.5.2 Growth Conditions

Definition 1.5.2 (Polynomial Growth Condition)
We say that $f : \mathbb{R} \to \mathbb{R}$ satisfies the *polynomial growth condition* if there is some constant $K > 0, m \in \mathbb{Z}_+$ such that

$$|f(x)| \leq K(1 + |x|^m).$$

The *linear growth condition* is a polynomial growth condition with $m = 1$.

Proposition 1.5.3
Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is such that $|f(0, t)| \leq C$ for all t and $f(x, t)$ is uniformly Lipschitz condition in x. Then $f(x, t)$ satisfies the linear growth condition in x,

$$|f(x, t)| \leq K(1 + |x|).$$

Theorem 1.5.4 (Gronwall's Inequality)
Let $g, h : [0, T] \to \mathbb{R}$ be regular and non-negative. For any regular, non-negative f satisfying the following inequality for all $t \in [0, T]$

$$f(t) \leq g(t) + \int_0^t h(s)f(s)ds,$$

we have

$$f(t) \leq g(t) + \int_0^t h(s)g(s) \exp\left(\int_s^t h(u)du\right) ds.$$

In the case where g is non-decreasing, the integral above simplifies to give

$$f(t) \leq g(t) \exp\left(\int_0^t h(s)ds\right).$$

In the most basic case when $g = A, h = B$ are constants,

$$f(t) \leq A \exp(Bt).$$

1.5.3 First-Order Linear Differential Equations

A first-order linear differential equation is of the form

$$\frac{dx(t)}{dt} + g(t)x(t) = k(t).$$

These equations are solved by using the integrating factor method. Choose some $G(t)$ such that $G'(t) = g(t)$. After multiple both sides by $e^{G(t)}$, integrating, and solving for $x(t)$, we have

$$x(t) = e^{-G(t)} \int_0^t \left(e^{G(s)} k(s)\right) ds + x(0) e^{G(0)-G(t)}.$$

Note that $G(t)$ is unique up to an additive constant so that the solution above is unique.

Chapter 2

Preliminaries from Probability Theory

The treatment in this chapter is far from complete compared to the analysis preliminaries. We assume a first graduate course in probability. In particular, we assume knowledge of elementary measure-theoretic probability.

2.1 Gaussian Distributions

The density of a Gaussian random vector $X : \Omega \to \mathbb{R}^n$ is given by

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

The moment generating function of the Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ is given by

$$\mathbb{E}[e^{\langle t, X\rangle}] = e^{\mu t - \frac{1}{2}t^T\Sigma t}.$$

If Z is the random vector whose components $Z_i \sim i.i.d.\mathcal{N}(0,1)$, then

$$X = \mu + AZ$$

for some $AA^T = \Sigma$. In general, if $X \sim \mathcal{N}(\mu, \Sigma)$ and B is a matrix,

$$BX \sim \mathcal{N}(B\mu, B\Sigma B^T).$$

Definition 2.1.1
A collection of random variables is a *Gaussian process* if the joint distribution of any finite subset of its members is Gaussian.

It can be shown that if a process $X(t)$ has independent Gaussian increments, then it is a Gaussian process.

2.2 Conditional Expectation

Let $X \in L^1$ be some random variable. Given some σ-field $\mathcal{G}$, the *conditional expectation* of X with respect to $\mathcal{G}$ is some $\mathcal{G}$-measurable random variable $\mathbb{E}[X \mid \mathcal{G}]$ such that

$$\int_B X d\mathbb{P} = \int_B \mathbb{E}[x \mid \mathcal{G}] d\mathbb{P}$$

for any $\mathcal{G}$-measurable B.

Equivantly, for any bounded $\mathcal{G}$-measurable variable ξ,

$$\mathbb{E}[\xi \mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\xi X].$$

The existence of such a variable is guaranteed by the Radon-Nikodym theorem. Moreover, $\mathbb{E}[x \mid \mathcal{G}]$ is a.s. unique.

2.2.1 Discrete & Continuous Cases

The *conditional distribution function* of X given $Y = y$ is defined as

$$\mathbb{P}\{ X \leq \mid Y = y \} := \frac{\mathbb{P}\{ X \leq x, Y = y \}}{\mathbb{P}\{ Y = y \}}.$$

This is not defined if the event upon which we condition has probability 0. We can overcome this difficulty if X, Y have a joint density $p(x, y)$. In this case, we define the *conditional density* of X given $Y = y$

$$p(x \mid y) := \frac{(x, y)}{p_Y(y)}.$$

Here $p_Y(y) := \int p(x, y) dx$ is the marginal density of Y.

The conditional expectation of X given y is thus

$$\mathbb{E}[X \mid Y = y] := \int x f(x \mid y) dx.$$

By replacing y with the random variable Y, we recover the conditional expectation $\mathbb{E}[X \mid Y]$.

Remark that the conditional distributiona and density are only defined at points $f_Y(y) > 0$. We can define it arbitrarily at points $f_Y(y) = 0$ since those points amount only to a set of measure 0. Note that this recovers the fact that the conditional expectation is a.s. unique.

2.2.2 Properties of Conditional Expectation

By directly applying the definition of the conditional expectation, use by considering approximations through simple functions and applying the monotone convergence theorem, it is possible to derive the following properties.

1. If $\mathcal{G} = \{ \varnothing, \Omega \}$,
$$\mathbb{E}[X \mid \mathcal{G}] \overset{a.s.}{=} \mathbb{E}[X].$$

2. If Y is $\mathcal{G}$-measurable,
$$\mathbb{E}[YX \mid \mathcal{G}] \overset{a.s.}{=} Y\mathbb{E}[X \mid \mathcal{G}].$$

3. If $\mathcal{G}_1 \subseteq \mathcal{G}_2$,
$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] \overset{a.s.}{=} \mathbb{E}[X \mid \mathcal{G}_1].$$
This is known as the *smoothing property*. Note that by taking $G_1 = \{ \varnothing, \Omega \}$,
$$\mathbb{E}[\mathbb{E}[X \mid G_2]] \overset{a.s.}{=} \mathbb{E}[X].$$

4. If $\sigma(X) \perp \mathcal{G}$ and $\mathcal{F} \perp \mathcal{G}$,
$$\mathbb{E}[X \mid \sigma(\mathcal{F}, \mathcal{G})] \overset{a.s.}{=} \mathbb{E}[X \mid \mathcal{F}].$$

5. If g is a convex function on $\operatorname{Im}(X)$,
$$g(\mathbb{E}[X \mid \mathcal{G}]) \overset{a.s.}{\leq} \mathbb{E}[g(X) \mid \mathcal{G}].$$
In particular, by taking $g(x) := |x|$,
$$|\mathbb{E}[X \mid \mathcal{G}]| \overset{a.s.}{\leq} \mathbb{E}[|X| \mid \mathcal{G}].$$

6. Suppose $0 \overset{a.s.}{\leq} X_n \uparrow X \in L^1$. Then
$$\mathbb{E}[X_n \mid \mathcal{G}] \overset{a.s.}{\uparrow} \mathbb{E}[X \mid \mathcal{G}].$$
This is the familiar monontone convergence.

7. If $0 \overset{a.s.}{\leq} X_n$,
$$\mathbb{E}\left[\liminf_n X_n \mid \mathcal{G}\right] \overset{a.s.}{\leq} \liminf_n \mathbb{E}[X_n \mid \mathcal{G}].$$
This is known as *Fatou's Lemma*.

8. If $X_n \xrightarrow{a.s.} X$ and $|X_n| \overset{a.s.}{\leq} Y$ with $\mathbb{E}[Y] < \infty$,
$$\mathbb{E}[X_n \mid \mathcal{G}] \xrightarrow{a.s.} \mathbb{E}[X \mid \mathcal{G}].$$

The following results are commonly applied.

Theorem 2.2.1
Let $X \perp Y$ be random variables and $\phi(x, y) \in L^1$. Then

$$\mathbb{E}[\phi(X, Y) \mid Y] = G(Y)$$

where $G(y) := \mathbb{E}[\phi(X, y)]$.

Theorem 2.2.2
Let (X, Y) be a Gaussian vector. Then the conditional distribution of $X \mid Y$ is also Gaussian. Moreover, provided that $\operatorname{Cov}[\,Y, Y\,]$ is non-singular,

$$\mathbb{E}[X \mid Y] = \mathbb{E}[X] + \operatorname{Cov}[\,X, Y\,]\operatorname{Cov}[\,Y, Y\,]^{-1}(Y - \mathbb{E}[Y]).$$

If $\operatorname{Cov}[\,Y, Y\,]$ is singular, the same formula holds with the inverse replaced by the Moore-Penrose pseudo-inverse.

Theorem 2.2.3 (Best Estimator / Predictor)
Let Y be such that for any X-measurable random variable Z,

$$\mathbb{E}[(X - \hat{X})^2] \le \mathbb{E}[(X - Z)^2].$$

Then $\hat{X} \overset{a.s.}{=} \mathbb{E}[X \mid Y]$.

2.3 Continuous Time Processes

The construction of continous time stochastic processes follow the same ideas as in discrete time, but are much more involved. Consider a random element $S : \Omega \to D[0, T]$ where $D[0, T]$ is the set of all CADLAG functions on $[0, T]$.

The simplest sets for which we would like to calculate probabilities are sets of the form $\{\, \omega : S(t_1; \omega) \in [a, b] \,\}$ for some fixed $t_1 \in [0, T]$. More generally, we may also be interested how the value at t_1 affects the value at another time t_2. In general, we would like to have all *finite-dimensional distributions* of the process. That is, probabilities of the form

$$\{\, \omega : S(t_i; \omega) \in B_i, i \in [n] \,\}$$

where B_i are intervals on the line. Formally, these sets are known as *cylinder sets*.

Probability is first defined on cylinders and then extended to the field generated by cylinders. Kolmogorov's extension theorem ensures that such an extension is consistent and well-defined. A probability defined on a field of cylinder sets can then be extended uniquely to the σ-field generated by cylinder sets.

It follows from this cosntruction that:

(a) For any choice of points $0 \le t_1 \le \cdots \le t_n \le T$, $S(t_1), \ldots S(t_n)$ is a random vector.

(b) The process is determined by its finite-dimensional distributions.

2.3.1 Continuity & Regularity of Paths

It is natural to consider a stochastic process $S(t)$ as a random function in t. Realizations of S are CADLAG functions $\omega \in D[0,1]$. Finite-dimensional distributions do not determine the continuity of sample paths.

Example 2.3.1
Let $X(t) = 0$ for all $t \in [0,1]$ and $\tau \sim U[0,1]$. Define

$$Y(t) := \begin{cases} X(t), & t \neq \tau \\ 1, & t = \tau \end{cases}$$

so that all finite-dimensional distributions of $X(t), Y(t)$ are the same. Moreover, $\mathbb{P}\{\, X(t) = Y(t)\,\} = 1$ for all $t \in [0,1]$.

However, the sample paths of X are continuous, but every sample path of Y has a jump.

Definition 2.3.2 (Versions (Modifications))
Two stochastic processes are *versions (modifications)* of one another if

$$\mathbb{P}\{\, X(t) = Y(t)\,\} = 1$$

for every $t \in [0,T]$.

Thus the two processes from the previous example are versions one another. If we agree to pick any version of the process we want, we can pick the smoothest possible version of the processes.

Define $N_t := \{\, X(t) \neq Y(t)\,\}$ and remark that $\mathbb{P}(N_t) = 0$. However, there are uncountably many t's and there is no contradiction that $N := \cup_t N_t$ has probability 1. In the case that $\mathbb{P}(N) = 0$, we say X, Y are *indistinguishable (evanescent)*.

Remark that in discrete time, if X, Y are versions of one another, they are indistinguishable. In addition, if X, Y are both right-continuous, then they are indistinguishable.

We would like to work with continuous or regular versions of processes if possible. Some conditions for the existence of such versions are given below.

Theorem 2.3.3
Let $S(t)$ be a real-valued stochastic process.

1. If there $\alpha, \varepsilon, C > 0$ so that for any $0 \leq u \leq t \leq T$,

$$\mathbb{E}\left[\,|S(t) - S(u)|^{\alpha}\,\right] \leq C(t-u)^{1+\varepsilon},$$

   then there exists a version of S with Hölder continuous paths of order $h < \varepsilon/\alpha$.

2. If there are $\alpha_1, \alpha_2, \varepsilon, C > 0$ so that for any $0 \leq u \leq v \leq t \leq T$,

$$\mathbb{E}\left[\,|S(v) - S(u)|^{\alpha_1} \cdot |S(t) - S(v)|^{\alpha_2}\,\right] \leq C(t-u)^{1+\varepsilon},$$

   Then there exists a version of S that is regular and has one-sided limits at the boundaries.

This result allows us to decide on the existence of continuous (regular) versions of processes by considering the bivariate (trivariate) distributions of the process. The same result hods when the process takes values in $\mathbb{R}^d$, except that the Euclidean distance replaces the absolute value in the above functions.

Regular functions are typically considered the same if all left and right limits coincide. In this case, it can be convenient to identify any such function with its right-continous version.

Theorem 2.3.4
If the stochastic process $S(t)$ is

(i) right-continuous in probability, ie for every $t \in [0, T]$,

$$S(u) \xrightarrow{p} S(t)$$

as $u \downarrow t$

(ii) regular

then it has a right-continuous version.

It is possible to derive alternative conditions for smoothness when additional properties of the process is known to us.

2.3.2 σ-Field Generated by a Stochastic Process

We define

$$\mathcal{F}_t := \sigma\left\{\, S_u : u \leq t \,\right\}$$

as the smallest σ-field containing sets of the form $\{ S_u \in [a, b] \}$ for $0 \leq u \leq t$ and $a, b \in \mathbb{R}$. We interpret this as the information available to an observer of the process S up to time t.

2.3.3 Filtered Probability Space & Adapted Processes

A *filtration* $\mathscr{F} = \{ \mathcal{F}_t \}$ is a family of increasing σ-fields on $(\Omega, \mathcal{F})$, ie

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$$

for every $s \leq t$.

A *filtered probability space* $(\Omega, \mathcal{F}, \mathscr{F}, \mathbb{P})$ is a probability space $(\Omega, \mathcal{F}, \mathscr{F})$ paired with a filtration $\mathscr{F}$ such that

$$\mathcal{F}_0 \subseteq \mathcal{F}_t \subseteq \ldots \subseteq \mathcal{F}_T = \mathcal{F}.$$

A stochastic process on this filtered probability space is said to be *adapted* if for every t, $S(t)$ is $\mathcal{F}_t$-measurable. Intuitively, this means $\mathcal{F}_t$ contains all information about $S(t)$ (and possibly more).

2.3.4 The Usual Conditions

A filtration is said to be *right-continuous* if

$$\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s.$$

The standard assumption (referred to as the usual condition) is that filtrations are right-continuous. We interpret this as any information known immediately after t is also known at t.

Remark 2.3.5 If $S(t)$ is $\mathscr{F}$-adapted, it is also adapted to $\mathscr{G} := \{ \mathcal{G}_t \}$ for

$$\mathcal{G}_t := \mathcal{F}_{t+},$$

which is a right-continuous filtration.

The assumption of right-continuous filtration has a number of important and useful consequences. We will see that it allows us to assume that martingales, submartingales, and supermartingales have regular, right-continuous versions.

It is also assumed that any subset of a set of zero probability is $\mathcal{F}_0$-measurable. It is always possible to enlarge the σ-field to include such sets if this property does not hold in the first place.

2.3.5 Martingales & Friends

Definition 2.3.6 (Martingale)
Suppose a $\mathscr{F}$-adapted process $X(t)$ is a *martingale* if

(i) $X(t) \in L^1$ for all t

(ii) For any $s < t$,
$$\mathbb{E}[\, X(t) \mid \mathcal{F}_s \,] = X(s).$$

If in place of (ii) we have

$$\mathbb{E}[\, X(t) \mid \mathcal{F}_s \,] \leq X(s),$$

then $X(t)$ is a *supermartingale*. Similarly, if instead of (ii) we have

$$\mathbb{E}[\, X(t) \mid \mathcal{F}_s \,] \geq X(s),$$

then $X(t)$ is said to be a *submartingale*.

The following is an important example of a martingale.

Theorem 2.3.7 (Doob-Lévy Martingale)
Let $Y \in L^1$ be an integrable random variable. Then

$$M(t) := \mathbb{E}[\, Y \mid \mathcal{F}_t \,]$$

is a martingale.

Proof
By the smoothing property,

$$\begin{aligned} \mathbb{E}[\, M(t) \mid \mathcal{F}_s \,] &:= \mathbb{E}[\, \mathbb{E}[\, Y \mid \mathcal{F}_t \,] \mid \mathcal{F}_s \,] \\ &= \mathbb{E}[\, Y \mid \mathcal{F}_s \,] \\ &=: M(s). \end{aligned}$$

$\square$

Using the smoothing property, we can show that the mean of a martingale, supermatingale, and submartingale is constant, non-increasing, and non-decreasing in t, respectively.

If $X(t)$ is a supermartingale, then $-X(t)$ is a submartingale by definition.

The super and submartingale property allows us to derive conditions of the right-continuous versions of processes without the assumption of continuity in probability.

Theorem 2.3.8
Let $\mathscr{F}$ be a right-continuous filtration with each σ-field $\mathcal{F}_t$ completed by null sets from $\mathcal{F}$. A $\mathscr{F}$-adapted supermartingale $X(t)$ has a CADLAG version if and only if its mean function $\mathbb{E}X(t)$ is right-continuous. Consequently, any martingale with right-continuous filtration admits a regular right-continuous version.

In view of these results, it will often be assumed that the version of the process under consideration is CADLAG.

2.3.6 Stopping Times

Definition 2.3.9 (Stopping Time)
A non-negative random variable $\tau : \Omega \to [0, \infty]$ is a *stopping time* with respect to a filtration $\mathscr{F}$ if for each t,

$$\{\,\tau \leq t\,\} \in \mathcal{F}_t.$$

It follows from the definition that $\{\,\tau > t\,\} \in \mathcal{F}_t$ as well.

Remark 2.3.10 The event

$$\left\{\,\tau \leq t - \frac{1}{n}\,\right\} \in \mathcal{F}_{t-1/n}.$$

Since $\mathcal{F}_t$'s are increasing, we also have $\{\,\tau \leq t - 1/n\,\} \in \mathcal{F}_t$. Therefore $\{\,\tau < t\,\} \in \mathcal{F}_t$. In fact, define

$$\mathcal{F}_{t-} := \bigvee_{s<t} \mathcal{F}_s := \sigma\left(\bigcup_{s<t} \mathcal{F}_s\right).$$

The above argument show that $\{\,t < \tau\,\} \in \mathcal{F}_{t-}$.

Theorem 2.3.11
Let $\mathscr{F}$ be a right-continuous filtration. Then τ is a $\mathscr{F}$-stopping time if and only if for each t, the event $\{\,\tau < t\,\} \in \mathcal{F}_t$.

Proof
The remark above shows the implication direction. Conversely, suppose $\{\,\tau < t\,\} \in \mathcal{F}_t$ for all t. We have

$$\{\,\tau \leq t\,\} = \bigcap_{n \geq 1} \{\,\tau < t + \frac{1}{n}\,\}.$$

Since $\{\,\tau < t + 1/n\,\} \in \mathcal{F}_{t+1/n}$, we also have $\{\,\tau \leq t\,\} \in \mathcal{F}_t$ by the right-continuity of $\mathscr{F}$. $\square$

The assumption of right-continuity of $\mathscr{F}$ is important when studying hitting or exit times of a process. If $S(t)$ is a stochastic process adapted to $\mathscr{F}$, the *hitting time* of set A is defined as

$$T_A := \inf\{\, t \geq 0 : S(t) \in A \,\}.$$

The first *exit time* from a set D is defined as

$$\tau_D := \inf\{\, t \geq 0 : S(t) \notin D \,\}.$$

Observe that $\tau_D = T_{\mathbb{R}\setminus D}$.

Theorem 2.3.12
Let $S(t)$ be continuous and adapted to $\mathscr{F}$.

(a) If D is an open subset of $\mathbb{R}$, then τ_D is a stopping time.

(b) If A is closed, then T_A is a stopping time.

(c) If in addition $\mathscr{F}$ is right-continuous, then for closed sets D and open sets A, τ_D, T_A are stopping times as well.

Proof
We have

$$\{\, \tau_D > t \,\} = \bigcap_{0 \leq u \leq t} \{\, S(u) \in D \,\}.$$

This event is an uncountable intersection over all $u \leq t$ of events from $\mathcal{F}_t$. By the continuity of $S(u)$ and D being open, if there is some irrational u such that $S(u) \in D$, then there must be some close by rational point q such that $S(q) \in D$. It follows that

$$\bigcap_{0 \leq u \leq t} \{\, S(u) \in D \,\} = \bigcap_{q \in \mathbb{Q}: 0 \leq q \leq t} \{\, S(q) \in D \,\}$$

which is a countable intersection of events from $\mathcal{F}_t$ and hence belongs to $\mathcal{F}_t$. This shows that t_D is a stopping time.

Since $\mathbb{R} \setminus A$ is open and $T_A = \tau_{\mathbb{R}\setminus A}$, it follows immediately that T_A is a stopping time.

Now suppose that $\mathscr{F}$ is right-continuous. Our plan is to apply the previous theorem and show hat $\{\, \tau_D < t \,\} \in \mathcal{F}_t$. The case for T_A follows again by taking complements.

Suppose D is closed. Then $\mathbb{R} \setminus D$ is a countable union of open sets, in fact open intervals. Thus D is a countable intersection of closed intervals $C_n = [a_n, b_n]$. It follows that

$$\begin{aligned} D &= \bigcap_{n \geq 1} [a_n, b_n] \\ &= \bigcap_{n \geq 1} \bigcap_{m \geq 1} \left(a_n - \frac{1}{m}, b_n + \frac{1}{m} \right). \end{aligned}$$

Define $D_{n,m} := (a_n - {}^1\!/_m, b_n + {}^1\!/_m)$. Then $\tau_{D_{n,m}}$ is a stopping time and so $\{ \tau_{D_{n,m}} > t \} \in \mathcal{F}_t$. But

$$\{ \tau_D \geq t \} = \bigcap_{n,m \geq 1} \{ \tau_{D_{n,m}} > t \},$$

hence $\{ \tau_D \geq t \} \in \mathcal{F}_t$ and $\{ \tau_D < t \} \in \mathcal{F}_t$ as desired. □

For general CADLAG processes, the following result holds.

Theorem 2.3.13
Suppose $S(t)$ is a CADLAG, $\mathscr{F}$-adapted process for some right-continuous $\mathscr{F}$.

(a) If $A \subseteq \mathbb{R}$ is open, then T_A is a stopping time.

(b) If A is closed, then

$$\{ t > 0 : S(t) \in A \vee S(t-) \in A \}$$

is a stopping time.

It is possible but much harder to show that the hitting time of a Borel set is a stopping time.

The following results give basic properties of stopping times.

Theorem 2.3.14
Let S, T be stoppign times. The following are all stopping times.

(a) $\min(S, T)$

(b) $\max(S, T)$

(c) $S + T$

2.3.7 σ-Field $\mathcal{F}_T$

If T is some stopping time, events observed before or at time T are described by a σ-field $\mathcal{F}_T$, defined as the collection of sets

$$\mathcal{F}_T := \{ A \in \mathcal{F} : \forall t, A \cap \{ T \leq t \} \in \mathcal{F}_t \}.$$

Theorem 2.3.15
Let S, T be stopping times. Then the following properties hold.

(a) If $A \in \mathcal{F}_S$, then $A \cap \{ S = T \} \in \mathcal{F}_T$.

(b) $\{ S = T \} \in \mathcal{F}_S \cap \mathcal{F}_T$.

(c) If $A \in \mathcal{F}_S$, then $A \cap \{ S \leq T \} \in \mathcal{F}_T$.

(d) $\{ S \leq T \} \in \mathcal{F}_S \cap \mathcal{F}_T$.

2.3.8 Fubini's Theorem

We state a particular case of Fubini's theorem that is formulated in the way that is typically applied in practice.

Theorem 2.3.16
Let $X(t)$ be a stochastic process on $[0, T]$ with regular sample paths. Then

$$\int_0^T \mathbb{E}\left[\, |X(t)| \,\right] dt = \mathbb{E}\left[\int_0^T |X(t)| dt \right].$$

Furthermore, if this quantity is finite,

$$\mathbb{E}\left[\int_0^T X(t) dt \right] = \int_0^T \mathbb{E}\left[\, X(t) \,\right] dt.$$

Part II

Stochastic Calculus

Chapter 3

Basic Stochastic Processes

3.1 Brownian Motion

The Brownian motion, also known as the Wiener process, serves as a basic model for the cumulative effect of pure noise. If $B(t)$ denotes the position of a particle at time t, the displacement $B(t) - B(0)$ is the effect of purely random bombardment by molecules of the fluide, or the effect of noise over time t.

3.1.1 Defining Properties

Definition 3.1.1 (Brownian Motion)
Brownian motion is a stochastic process $B(t)$ with the following properties:

(i) (Independent Increments) $B(t) - B(s) \perp B_u$ for all $t > s \geq u \geq 0$ and $B(t) - B(s) \perp \mathcal{F}_s$ for all $t > s$.

(ii) (Normal Increments) $B(t) - B(s) \sim \mathcal{N}(0, t-s)$.

(iii) (Continuity of Paths) The sample paths of $B(t)$ are continuous functions of t.

The initial position is not specified in the definition but we can take it to be 0. We write P_x to denote the probability of events when the process states at x. The first two properties determine all the finite-dimensional distributions, all of whom are Gaussian. The time interval on which Brownian motion is defined is $[0, T]$ for some $\in (0, \infty]$.

Remark that we can deduce the existence of a continuous version of Brownian motion from the first two properties. Indeed,

$$\mathbb{E}\left[\, |B(t) - B(s)|^4 \,\right] = 3(t-s)^2.$$

Note that a more general model of Brownian motion is a pair $(B(t), \mathscr{F})$ where $B(t)$ is an $\mathscr{F}$-adapted process satisfying the defining properties.

Example 3.1.2
Although $B(t) - B(s)$ is independent of the past,

$$B(t) - 2B(s) = [B(t) - B(s)] - B(s)$$

is not.

Example 3.1.3
We write $W \overset{d}{=} B$ to be an independent copy.

By computation,

$$\begin{aligned}
&\mathbb{P}\{\, B(0) \leq 0, B(1) \leq 0, B(2) \leq (0) \,\} \\
&= \mathbb{P}\{\, B(1) \leq 0, B(2) \leq 0 \,\} \\
&= \mathbb{P}\{\, B(1) \leq 0, B(2) - B(1) \leq -B(1) \,\} \\
&= \mathbb{P}\{\, B(1) \leq 0, W(1) \leq -B(1) \,\} \\
&= \int_{-\infty}^{0} P(W(1) \leq -x) f(x) dx \\
&= \int_{0}^{\infty} \Phi(x) f(-x) dx && \text{change of variable} \\
&= \int_{0}^{\infty} \Phi(x) f(x) dx && f \text{ symmetric} \\
&= \int_{0}^{\infty} \Phi(x) d\Phi(x) \\
&= \int_{1/2}^{\infty} y dy && \text{change of variable} \\
&= \frac{3}{8}.
\end{aligned}$$

3.1.2 Transition Probability Functions

If the process started at $x \in \mathbb{R}$, $B(0) = x$ and $B(t) \sim \mathcal{N}(x, t)$. More specifically, the conditional distribution of $B(t+s)$ given $B(s) = x$ is $\mathcal{N}(x, t)$.

Define the *transition function*

$$\begin{aligned}
P(y, t, x, s) &:= \mathbb{P}\{\, B(t+s) \leq y \mid B(s) = x \,\} \\
&= P_x\{\, B(t) \leq y \,\}.
\end{aligned}$$

The density function of this distribution is the *transition density* of Brownian motion,

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

The finite-dimensional distributions can be computed using the transition density since the increments of Brownian motion are independent. Indeed,

$$\begin{aligned} &P_x\{\, B(t_1) \le x_1, B(t_2) \le x_2, \dots, B(t_n) \le x_n \,\} \\ &= \int_{-\infty}^{x_1} p_{t_1}(x, y_1)dy_1 \int_{-\infty}^{x_2} p_{t_2 - t_1}(y_1, y_2)dy_2 \cdots \int_{-\infty}^{x_n} p_{t_{n-1}-t_n}(y_{n-1}, y_n)dy_n. \end{aligned}$$

3.1.3 Space Homogeneity

Definition 3.1.4 (Space-Homogeneous)
A stochastic process is called *space-homogeneous* if its finite dimensional distributions are translation invariant, ie

$$\begin{aligned} &\mathbb{P}\{\, X(t_1) \le x_1, X(t_2) \le x_2, \dots, X(t_n) \le x_n \mid X(0) = 0 \,\} \\ &= \mathbb{P}\{\, X(t_1) \le x_1 + x, X(t_2) \le x_2 + x, \dots, X(t_n) \le x_n + x \mid X(0) = x \,\}. \end{aligned}$$

Brownian motion is space-homogeneous.

3.1.4 Brownian Motion as a Gaussian Process

Recall a process is Gaussian if all finite-dimensional distributions are Gaussian.

Example 3.1.5
Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent. The distribution of $(X, X+Y)$ is bivariate normal with mean vector $(\mu_1, \mu_1 + \mu_2)$ and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix}.$$

Indeed, suppose $Z \sim \mathcal{N}(0, I_2)$. Then

$$(X, X+Y) = \mu + AZ$$

where

$$A = \begin{bmatrix} \sigma_1 & 0 \\ \sigma_1 & \sigma_2 \end{bmatrix}.$$

Similar to this example, we can express any finite-dimensional distribution of Brownian motion using the independence of increments. Let $Y_1 := B(t_1)$ and $Y_k := B(t_k) - B(t_{k-1})$ for $k > 1$. By construction, the Y_k's are all independent. Moreover, $Y_1 \sim \mathcal{N}(0, t_1)$ and $Y_k \sim \mathcal{N}(0, t_k - t_{k-1})$ for $k > 1$. Remark that $Y_1 = \sqrt{t_1} Z_1$ and $Y_k = \sqrt{t_k - t_{k-1}} Z_k, k > 1$. Thus $(B(t_k))_{k \in [n]}$ is a linear transformation of $Z \sim \mathcal{N}(0, I_n)$.

Definition 3.1.6 (Covariance Function)
The *covariance function* of a process $X(t)$ is defined as

$$\begin{aligned}\gamma(s,t) &:= \operatorname{Cov}[\, X(t), X(s)\,] \\ &= \mathbb{E}[\,(X(t) - \mathbb{E}[X(t)])(X(s) - \mathbb{E}[X(s)])\,] \\ &= \mathbb{E}[\, X(t)X(s)\,] - \mathbb{E}[X(t)]\mathbb{E}[X(s)].\end{aligned}$$

The next result characterizes Brownian motion as a particular Gaussian process.

Theorem 3.1.7
A stochastic process satisfies the defining properties of a Brownian motion started at zero if and only if it is a Gaussian process with zero mean function and covariance function $\min(t, s)$.

Proof
Since the mean of Brownian motion is zero

$$\gamma(s,t) = \mathbb{E}[\, B(t)B(s)\,].$$

If $t < s$, then $B(s) = B(t) + B(s) - B(t)$ and

$$\begin{aligned}\mathbb{E}[\, B(t)B(s)\,] &= \mathbb{E}[B^2(t)] + \mathbb{E}[\, B(t)(B(s) - B(t))\,] \\ &= \mathbb{E}[B^2(t)] \\ &= t\end{aligned}$$

and vice versa if $t > s$.

Conversely, let t be arbitrary and $s \geq 0$. Suppose $X(t)$ is a Gaussian process so its joint distribution of $X(t), X(t+s)$ must be a bivariate normal with zero mean by assumption. But then the vector $(X(t), X(t+s) - X(t))$ is also bivariate normal. For any $u \leq t$,

$$\begin{aligned}\operatorname{Cov}[\, X(u), X(t+s) - X(t)\,] &= \operatorname{Cov}[\, X(u), X(t+s)\,] - \operatorname{Cov}[\, X(u), X(t)\,] \\ &= \min(u, t+s) - \min(u, t) && \text{assumption} \\ &= 0.\end{aligned}$$

But a bivariate normal distribution has independent marginals if and only if they are independent. Moreover,

$$\begin{aligned}\operatorname{Cov}[\, X(t+s) - X(t), X(t+s) - X(t) \,] &= (t+s) - 2t + t \\ &= s.\end{aligned}$$

Hence the increment $X(t+s) - X(t) \sim \mathcal{N}(0, s)$ and is independent of $X(t)$. Therefore, it is a Brownian motion. $\square$

Example 3.1.8
We find the distribution of $\sum_{i=1}^{n} B(i)$. The random vector $X = (B(i))_{i\in[n]}$ is a multivariate normal vector with mean zero and covariance matrix with entries $\Sigma_{ij} = \min(i, j)$. Then the desired random random is equal to $\mathbb{1}^T X$ and hence has a normal distribution with mean zero and covariance $\mathbb{1}^T \Sigma \mathbb{1}$.

Example 3.1.9
Suppose we wish to find the distribution of $\frac{1}{n}\sum_{i=1}^{n} B(i)$. Define $Y = (B(i/n))_{i\in[n]}$ and remark that $Y = \frac{1}{\sqrt{n}} X$ where X is from the previous example. Thus the covariance matrix of Y is given by $1/\sqrt{n}\Sigma$ and $\mathbb{1}^T Y$ has mean zero and variance $1/\sqrt{n}\mathbb{1}^T\Sigma\mathbb{1}$.

Example 3.1.10
We wish to find the probability

$$\mathbb{P}\left\{ \omega : \int_0^1 B(t;\omega)dt > x \right\}.$$

First note that since Brownian motion has continuous paths, the Riemann integral is well-defined for any particular sample path. In order to find the desired probability, we remark that the distribution of $\int_0^1 B(t)dt$ can be obtained as a limit of the distributions of the approximating sums $\sum_i B(t_i)\Delta$, where the points t_i partition $[0, 1]$ and $\Delta := t_{i+1} - t_i$.

Since the limit of Gaussian distributions is Gaussian, $\int_0^1 B(t)dt$ has a normal distribution with zero mean. It remains only to compute its variance. By Fubini's theorem,

$$\begin{aligned}
\operatorname{Var}\left[\int_0^1 B(t)dt\right] &= \operatorname{Cov}\left[\int_0^1 B(t)dt, \int_0^1 B(s)ds\right] \\
&= \mathbb{E}\left[\int_0^1 B(t)dt \int_0^1 B(s)ds\right] \\
&= \int_0^1\int_0^1 \mathbb{E}\left[\, B(t)B(s)\,\right] dtds \\
&= \int_0^1\int_0^1 \operatorname{Cov}\left[\, B(t), B(s)\,\right] dtds \\
&= \int_0^1\int_0^1 \min(t,s)dtds \\
&= \frac{1}{3}.
\end{aligned}$$

Thus $\int_0^1 B(t)dt \sim \mathcal{N}(0, 1/3)$ and the desired probability is calculated explicitly.

Note the application of Fubini's theorem is justified since

$$\begin{aligned}
&\int_0^1\int_0^1 \mathbb{E}\left[\, |B(t)B(s)|\,\right] dtds \\
&\leq \int_0^1\int_0^1 \mathbb{E}\left[\, |B(t)|\,\right]\cdot\mathbb{E}\left[\, |B(s)|\,\right] dtds && \text{Cauchy-Schwartz} \\
&= \int_0^1\int_0^1 \sqrt{\operatorname{Var}[B(t)]\cdot\operatorname{Var}[B(s)]}dtds \\
&< \infty.
\end{aligned}$$

3.1.5 Brownian Motion as a Random Series

The process

$$\frac{t}{\pi}\xi_0 + \sqrt{\frac{2}{\pi}}\sum_{j\geq 1}\frac{\sin(jt)}{j}\xi_j$$

where $\xi_j \sim_{iid} \mathcal{N}(0,1), j \geq 0$ is a Brownian motion on $[0,\pi]$. Convergence of the series is understood a.s. This representation resembles Weirstrauss's continuous but nowhere differentiable function. The claim can be shown by showing that the partial sums converge uniformly and verifying the process is Gaussian with zero mean and covariance $\min(s,t)$.

A more general representation of a Brownian motion is given by using an orthonormal sequence of functions on $[0, T]$, say $h_j(t)$. Then take

$$B(t) = \sum_{j \geq 0} \xi_j H_j(t).$$

Here $\int_0^t h_j(s)ds$ is a Brownian motion on $[0, T]$.

3.2 Properties of Brownian Motion Paths

3.2.1 Quadratic Variation of Brownian Motion

The quadratic variation of Brownina motion is a random variable given by

$$\begin{aligned}[B, B](t) &= [B, B]([0, t]) \\ &= \lim_{\delta_n \to 0} \sum_{i=1}^{n} |B(t_i^n) - B(t_{i-1}^n)|^2.\end{aligned}$$

Here the limit is taken in probability over all shrinking partitions of $[0, t]$.

Theorem 3.2.1
The quadratic variation of Brownian motion over $[0, t]$ is t.

Proof (Sketch[a])
Define

$$T_n := \sum_{i \in [n]} |B(t_i^n) - B(t_{i-1}^n)|^2.$$

By the independence of normal increments, we see that $\mathbb{E}[T_n] = t$. We claim that $\mathrm{Var}[T_n - t] \xrightarrow{p} 0$ so the result follows by Chebyshev's inequality.

First we note that

$$\begin{aligned}\mathbb{E}\left[\, (T_n - t)^2 \,\right] &= \mathbb{E}\left[\left(\sum_{i \in [n]} |B(t_i^n) - B(t_{i-1}^n)|^2 - (t_i^n - t_{i-1}^n) \right)^2 \right] \\ &= \sum_{i \in [n]} \mathbb{E}\left[\left(|B(t_i^n) - B(t_{i-1}^n)|^2 - (t_i^n - t_{i-1}^n) \right)^2 \right].\end{aligned}$$

This is because each summand is an independent zero mean random variable thus any interaction terms when expanding the square becomes zero.

By computation,

$$
\begin{aligned}
&\mathbb{E}\left[\,(T_n - t)^2\,\right] \\
&= \sum_{i\in[n]} \mathbb{E}\left[\,|B(t_i^n) - B(t_{i-1}^n)|^4\,\right] \\
&\quad - 2\sum_{i\in[n]} (t_i^n - t_{i-1}^n)\mathbb{E}\left[\,|B(t_i^n) - B(t_{i-1}^n)|^2\,\right] \\
&\quad + \sum_{i\in[n]} (t_i^n - t_{i-1}^n)^2 \\
&= 2\sum_{i\in[n]} (t_i^n - t_{i-1}^n)^2 && \mathbb{E}[Z^4] = 3\sigma^4, Z \sim \mathcal{N}(0,\sigma^2) \\
&\le 2\delta_n T.
\end{aligned}
$$

As $\delta_n \to 0$, this variance also tends to 0. This concludes the proof by our initial remark.□

[a] https://math.uchicago.edu/~may/REU2019/REUPapers/Carlstein.pdf

We also note that it is possible to show that $T_n \xrightarrow{a.s.} t$ for any sequence of partitions which are successive refinements and satisfy $\delta_n \to 0$.

Remark 3.2.2 In the proof above, we actually showed the stronger statement that $T_n \xrightarrow{2} t$.

By varying t, the quadratic variation process of Brownian motion is t. Remark that the classic quadratic variation of Brownian motion paths, defined as the supremum over all partitions, not just shrinking ones, is infinite.

3.2.2 Properties of Brownian Paths

Let us think of $B(t)$ as a distribution over sample paths. Almost surely, a sample path satisfies the following.

1. Is a continuous function of t.
2. Not monotone in any interval, regardless of the length.
3. Not differentiable at any point.
4. Has infinite variation on any interval, regardless of length.
5. Has quadratic variation on $[0, t]$ equal to t, for any t.

Note that a continuous function with bounded derivative is of finite variation. Thus it follows from property 4 that $B(t)$ can not have a bounded derivative on any interval. This is not yet the non-differentiability at any point. We show a simpler statement below.

Theorem 3.2.3
For any t, the trajectories of Brownian motion are not differentiable at t almost surely.

Proof
We remark that

$$\frac{B(t+\Delta)-B(t)}{\Delta} \overset{d}{=} \frac{\sqrt{\Delta}Z}{\Delta} = \frac{Z}{\sqrt{\Delta}}.$$

Here $Z \sim \mathcal{N}(0,1)$. But

$$\mathbb{P}\left\{ \left|\frac{Z}{\sqrt{\Delta}}\right| > K \right\} \to 1$$

for any K as $\Delta \to 0$, the ratio converges to ∞ in distribution and so the derivative cannot exist almost surely. □

3.3 Three Martingales of Brownian Motion

Recall a stochastic process is a *martingale* if for any t, $X(t) \in L^1$ and for any $s > 0$,

$$\mathbb{E}\left[\, X(t+s) \mid \mathcal{F}_t \,\right] \overset{a.s.}{=} X(t).$$

where $\mathcal{F}_t = \sigma\left\{\, X(u) : u \le t \,\right\} = \sigma(X(t))$.

Remark 3.3.1 Intuitively, $\mathcal{F}_t$ is the information available to an observer at time t. A set $A \in \mathcal{F}_t$ only if one can decide whether or not A has occured by observing the process up to time t.

Remark 3.3.2 Since the conditional expectation given a σ-field is defined as a random variable, all relations such as equalities and inequalities must be understood in the almost surely sense. Thus the "a.s." will frequently be dropped for brevity.

Examples of martingales constructed from Brownian motion are stated below.

Theorem 3.3.3
Let $B(t)$ be a Brownian motion. The following are martingales.
(a) $B(t)$
(b) $B(t)^2 - t$
(c) $\exp(uB(t) - {}^{u^2 t}/_2)$ for any u

Proof
First, we remark that

$$\begin{aligned} &\mathbb{E}\left[\, g(B(t+s)-B(t)) \mid \mathcal{F}_t \,\right] \\ &= \mathbb{E}\left[\, g(B(t+s)-B(t)) \,\right] && B(t+s)-B(t) \perp \mathcal{F}_t \\ &= \mathbb{E}[g(X(s))]. \end{aligned}$$

<u>$B(t)$</u>: By definition, $B(t) \sim \mathcal{N}(0,t)$ so that $B(t) \in L^1$. By the indepedence of increments,

$$\begin{aligned} \mathbb{E}\left[\, B(t+s) \mid \mathcal{F}_t \,\right] &= \mathbb{E}\left[\, B(t) \mid \mathcal{F}_t \,\right] + \mathbb{E}\left[\, B(t+s)-B(t) \mid \mathcal{F}_t \,\right] \\ &= B(t) + 0. \end{aligned}$$

<u>$B(t)^2 - t$</u>: By definition, $\mathbb{E}[\, B^2(t) \,] = t < \infty$. We now perform a similar calculation.

$$\begin{aligned} &\mathbb{E}\left[\, B^2(t+s) \;\middle|\; \mathcal{F}_t \,\right] \\ &= \mathbb{E}\left[\, (B(t) + B(t+s) - B(t))^2 \;\middle|\; \mathcal{F}_t \,\right] \\ &= B^2(t) + 2B(t) \cdot \mathbb{E}\left[\, B(t+s) - B(t) \mid \mathcal{F}_t \,\right] + \mathbb{E}\left[\, (B(t+s)-B(t))^2 \;\middle|\; \mathcal{F}_t \,\right] \\ &= B^2(t) + s. \end{aligned}$$

Subtracting $(t+s)$ from both sides yields the martingale property.

<u>$\exp(uB(t) - u^2 t/2)$</u>: Since $B(t) \sim \mathcal{N}(0,t)$, inspecting its moment generating function yields integrability.

$$\mathbb{E}\left[\, \exp(uB(t)) \,\right] = \exp(tu^2/2) < \infty.$$

Finally, we apply our initial remark.

$$\begin{aligned} &\mathbb{E}\left[\, \exp(uB(t+s)) \mid \mathcal{F}_t \,\right] \\ &= \mathbb{E}\left[\, \exp(uB(t) + u(B(t+s)-B(t))) \mid \mathcal{F}_t \,\right] \\ &= \exp(uB(t)) \cdot \mathbb{E}\left[\, u(B(t+s)-B(t)) \mid \mathcal{F}_t \,\right] && \sigma(B(t)) = \mathcal{F}_t \\ &= \exp(uB(t)) \cdot \mathbb{E}\left[\, u(B(t+s)-B(t)) \,\right] && \text{remark} \\ &= \exp(uB(t)) \exp(u^2 s/2). \end{aligned}$$

The martingale property is obtained by multiplying both sides by $\exp(-u^2(t+s)/2)$. □

All three martingales are central in theory. Lévy's characterization states that $X(t)$ is a continuous martingale such that $X^2(t) - t$ is a martingale if and only if $X(t)$ is a Brownian motion. The third martingale is known as the *exponential martingale* and can be used to establish distributional properties of the process.

3.4 Markov Property of Brownian Motion

Definition 3.4.1 (Markov)
A stochastic process $X(t)$ is *Markov* if for any $s > 0$ and t,

$$\mathbb{P}\{ X(t+s) \le y \mid \mathcal{F}_t \} \overset{a.s.}{=} \mathbb{P}\{ X(t+s) \le y \mid X(t) \}.$$

Theorem 3.4.2
Brownian motion is Markov.

Proof
It suffices to show that the moment generating function for $B(t+s) \mid \mathcal{F}_t$ is the same as $B(t+s) \mid \mathcal{B}(t)$.

Indeed,

$$\begin{aligned}
&\mathbb{E}[\exp(uB(t+s)) \mid \mathcal{F}_t] \\
&= \exp(uB(t)) \cdot \mathbb{E}[\exp(uB(t+s) - B(t)) \mid \mathcal{F}_t] \\
&= \exp(uB(t)) \cdot \mathbb{E}[\exp(uB(t+s) - B(t))] && B(t+s) - B(t) \perp \mathcal{F}_t \\
&= \exp(uB(t)) \cdot \mathbb{E}[\exp(uB(t+s) - B(t)) \mid B(t)] && B(t+s) - B(t) \perp B(t) \\
&= \mathbb{E}[\exp(uB(t+s)) \mid B(t)]. && B(t) \in \sigma(B(t))
\end{aligned}$$

$\square$

Recall the transition probability function of a Markov process $X(t)$ is defined as

$$P(y,t,x,s) := \mathbb{P}\{ X(t) \le y \mid X(s) = x \}.$$

It is possible to choose these functions so that for nay fixed x, they are true probabilities on the line. In the case of Brownian motion,

$$P(y,t,x,s) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(u-x)^2}{2(t-s)} \right) du.$$

In other words,

$$P(y,t,x,s) = \mathbb{P}\{ B(t-s) \le y \mid B(0) = x \}.$$

This property states that Brownian motion is *time-homogeneous*, that its distributions do not change with a shift in time.

3.4.1 Stopping Times & Strong Markov Property

Recall that a random time T is a *stopping time* for $B(t)$ if for any $t \ge 0$, $\{T \le t\} \in \mathcal{F}_t$. Intuitively, this means that it is possible to deduce whether T has occured or not by time t

through observing $B(s), 0 \leq s \leq t$.

The following are all examples of stopping times and random times.

1. Any deterministic time is a stopping time since $\{ T \leq t \}$ is either $\varnothing$ or Ω.
2. The first hitting time of $a \in \mathbb{R}$ is a stopping time since $\{ T > t \} = \{ \forall u \leq t, B(u) < a \}$.
3. The time T when Brownian motion reaches its maximum on the interval $[0, 1]$ is not a stopping time.
4. The time T of the last zero before time $t = 1$ is not a stopping time.

The *strong Markov* property is similar to the Markov property, except that the fixed time t is replaced by a stopping time.

Theorem 3.4.3
Brownian motion is strongly Markov: For any finite stopping time T, the regular conditional distribution $B(T + t) \mid \mathcal{F}_T$ for $t \geq 0$ is $P_{B(T)}$.

$$\mathbb{P}\{ B(T+t) \leq y \mid \mathcal{F}_T \} \overset{a.s.}{=} \mathbb{P}\{ B(T+t) \leq y \mid B(T) \}.$$

The proof of the strong Markov property cand be shown using the exponential martingale and the optional stopping theorem.

Corollary 3.4.4
Let T be a finite stopping time. Define a new process in $t \geq 0$ as

$$\widehat{B}(t) := B(T+t) - B(T).$$

Then $\widehat{B}(t)$ is a Brownian motion started at zero and is independent of $\mathcal{F}_T$.

3.5 Hitting Times & Exit Times

Let T_x denote the first time $B(t)$ hits $x \in \mathbb{R}$,

$$T_x := \inf \{ t > 0 : B(t) = x \}.$$

Moreover, denote the time to exit an interval by $\tau := \min(T_a, T_b)$.

Theorem 3.5.1
Let $x \in (a, b)$ and $\tau := \min(T_a, T_b)$. Then $E_x[\tau] < \infty$ and $P_x\{ \tau < \infty \} = 1$.

Proof
First, we remark that

$$\{\, \tau > 1 \,\} = \{\, \forall 0 \le s \le 1, B(s) \in (a,b) \,\} \subseteq \{\, B(1) \in (a,b) \,\}.$$

Thus

$$P_x \{\, \tau > 1 \,\} \le P_x \{\, B(1) \in (a,b) \,\} = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{(y-x)^2}{2}\right) dy.$$

The RHS is a continuous function of $x \in [a,b]$, hence it reaches its maximum $\theta < 1$, ie,

$$\theta := \max_{y \in (a,b)} P_y \left\{ \forall 0 \le s \le 1, \widehat{B}(s) \in (a,b) \right\} < 1.$$

In similar fashion, using the fact that $B(t)$ is Markov,

$$\begin{aligned}
&P_x \{\, \tau > n \,\} \\
&= P_x \{\, \forall 0 \le t \le n-1, B(t) \in (a,b) \land \forall n-1 \le s \le n, B(s) \in (a,b) \,\} \\
&= P_x \{\, \tau > n-1 \land \forall n-1 \le s \le n, B(s) \in (a,b) \,\} \\
&= P_x \left\{ \tau > n-1 \land \forall 0 \le s \le 1, B(n-1) + \widehat{B}(s) \in (a,b) \right\} \\
&= E_x \left[\mathbb{1} \{\, \tau > n-1 \,\} \cdot \mathbb{1} \left\{ \forall 0 \le s \le 1, B(n-1) + \widehat{B}(s) \in (a,b) \right\} \right] \\
&= E_x \left[E_x \left[\mathbb{1} \{\, \tau > n-1 \,\} \cdot \mathbb{1} \left\{ \forall 0 \le s \le 1, B(n-1) + \widehat{B}(s) \in (a,b) \right\} \mid \mathcal{F}_{n-1} \right] \right] \\
&= E_x \left[\mathbb{1} \{\, \tau > n-1 \,\} \cdot E_x \left[\mathbb{1} \left\{ \forall 0 \le s \le 1, B(n-1) + \widehat{B}(s) \in (a,b) \right\} \mid \mathcal{F}_{n-1} \right] \right] \\
&= E_x \left[\mathbb{1} \{\, \tau > n-1 \,\} \cdot P_x \left\{ \forall 0 \le s \le 1, B(n-1) + \widehat{B}(s) \in (a,b) \;\middle|\; B(n-1) \right\} \right] \\
&\le P_x \{\, \tau > n-1 \,\} \cdot \theta \\
&\le \dots \\
&\le \theta^n.
\end{aligned}$$

Since τ is a non-negative random variable,

$$E_x[\tau] \le \sum_{n \ge 0} P_x(X > n) \le \frac{1}{1-\theta} < \infty.$$

Note that this implies that $\tau \overset{a.s.}{<} \infty$ or else if there is any positive weight on events where $\tau = \infty$, the expectation cannot be finite. $\square$

The next result gives the recurrence property of Brownian motion.

Theorem 3.5.2
Let $a, b \in \mathbb{R}$. The following hold for hitting times of Brownian motion.
(a) $P_a \{ T_b < \infty \} = 1$
(b) $P_a \{ T_a < \infty \} = 1$

Proof
Note that (a) implies (b) since

$$P_a \{ T_a < \infty \} \geq P_a \{ T_b < \infty \} P_b \{ T_a < \infty \} = 1.$$

Thus it suffices to show (a).

By the previous result, either T_a, T_b occurs with probability 1. By symmetry,

$$P_{(a+b)/2} \{ T_a < T_b \} = \frac{1}{2}.$$

Consider now

$$P_0 \left\{ T_{-(2^n-1)} < T_1 \right\}.$$

Since the paths of Brownian motion are continuous, in order to reach -2^n+1, the path must reach $-1, -3$, etc. Hence by the Markov property,

$$\begin{aligned}
&P_0 \{ T_{-2^n+1} < T_1 \} \\
&= P_0 \{ T_{-1} < T_1 \} \cdot P_{-1} \{ T_{-3} < T_1 \} \cdot \dots \cdot P_{-2^{n-1}+1} \{ T_{-2^n+1} < T_1 \} \\
&= \frac{1}{2^n}.
\end{aligned}$$

Let A_n denote the event that Brownian motion hits -2^n+1 before 1. Then we showed that $\mathbb{P}(A_n) = 2^{-n}$. Observe that $A_n \subseteq A_{n-1}$, thus

$$\bigcap_{i=1}^n A_i = A_n$$

and

$$\mathbb{P}\left(\bigcap_{i \geq 1} A_i \right) = \lim_{n \to \infty} \mathbb{P}(A_n) = \lim_n 2^{-n} = 0.$$

It follows that

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n^c \right) = 1.$$

In other words, one of the events complementary to some A_n occurs a.s. so there is some n such that Brownian motion hits 1 before it hits -2^n+1. Thus $P_0 \{ T_1 < \infty \} = 1$. $\square$

We note that this fact can also be shown using the martingale property of Brownian motion.

Chapter 4

Brownian Motion Calculus

4.1 Itô Integral

We aim to define a notion of stochastic integral

$$\int_0^T X(t)dB(t),$$

also denoted $\int XdB$ or $X\cdot B$. Firstly, we would like $\int_0^T dB(t) = B(T)-B(0)$. More generally, if $X(t)$ is a simple function, say

$$X = c_1\mathbb{1}_{(0,a]} + c_2\mathbb{1}_{(a,T]},$$

the integral should be the sum of integrals over the two subintervals.

4.1.1 Simple Processes

Let us first consider integrals for deterministic simple processes $X(t)$. Let $0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of $[0,T]$. Recall these are functions of the form

$$X(t) = c_0\mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{n-1} c_i\mathbb{1}_{(t_i,t_{i+1}]}(t).$$

The Itô integral is defined as a sum

$$\int_0^T X(t)dB(t) := \sum_{i=0}^{n-1} c_i[B(t_{i+1}) - B(t_i)].$$

Remark that $\int X dB(t)$ is a random variable. Moreover, by the independence increments of Brownian motion, the integral is a Gaussian random variable with mean zero and variance

$$\begin{aligned}
\operatorname{Var}\left[\int_0^T X(t)dB(t)\right] &= \operatorname{Var}\left[\sum_{i=0}^{n-1} c_i[B(t_{i+1}) - B(t_i)]\right] \\
&= \sum_{i=0}^{n-1} \operatorname{Var}\left[c_i[B(t_{i+1}) - B(t_i)]\right] \\
&= \sum_{i=0}^{n-1} c_i^2(t_{i+1} - t_i).
\end{aligned}$$

By taking limits of simple deterministic processes, we can obtain more general but still deterministic random variables as integrals. For instance, if $X(t)$ is deterministic and "integrable" under some conditions,

$$\int_0^T X(t)dB(t) \sim \mathcal{N}\left(0, \int_0^T X^2(t)dt\right).$$

In order to integrate random processes, we allow the coefficients c_i to be random variables ξ_i. In order to obtain convenient properties of the integral, the random variable ξ_i's are allowed to depend on $B(s)$ for $s \leq t_i$, but not future values. In other words, the integrand process $X(t)$ must be $\{\mathcal{F}_t\}$-*adapted* where $\mathcal{F}_t = \sigma\{B(s) : s \leq t\}$.

Remark 4.1.1 If we wish to have more useful properties such as Fubini's theorem, adaptedness is insufficient. We instead consider *progressive* processes, ie, for every $t \in [0, T]$, $X : [0, t] \times \mathcal{F}_t \to \mathbb{R}$ is measurable. Luckily, any adapted CADLAG process is progressive. Thus if we restrict ourselves to CADLAG process (as we do), there is no need to consider progressiveness.

Definition 4.1.2 (Simple Adapted Process)
A process $X(t)$ is a *simple adapted process* if there are times $0 = t_0 < t_1 < \cdots < t_n = T$ and random variables $\xi_0, \xi_1, \ldots, \xi_{n-1}$ such that

$$X(t) = \xi_0 I\mathbb{1}_0(t) + \sum_{i=0}^{n-1} \xi_i \mathbb{1}_{(i,t_{i+1}]}(t).$$

Here ξ_0 is constant, ξ_i is $\mathcal{F}_{t_i}$-measurable, and $\mathbb{E}[\xi^2] < \infty$ for $i = 0, \ldots, n-1$.

For simple adapted processes, the Itô integral is defined as a random sum

$$\int_0^T X(t)dB(t) := \sum_{i=0}^{n-1} \xi_i[B(t_{i+1}) - B(t_i)].$$

When the ξ_i's are random, the integral need not have a normal distribution.

Remark 4.1.3 Simple adapted processes are defined as left-continuous step functions. One can also take right-continuous functions. However, when the stochastic integral is defined with respect to general martingales rather than just Brownian motion, only left-continuous functions are taken.

4.1.2 Properties of the Itô Integral for Simple Adapted Processes

We now establish the main properties of the Ito integral for simple processes which carry over to the Itô integral of general processes.

1. (Linearity) If $X(t), Y(t)$ are simple processes and α, β are constants,
$$\int_0^T (\alpha X(t) + \beta Y(t))dB(t) = \alpha \int_0^T X(t)dB(t) + \beta \int_0^T Y(t)dB(t).$$
2. For the indicator function of an interval $\mathbb{1}_{(a,b]}$,
$$\int_0^T \mathbb{1}_{(a,b]}(t)X(t)dB(t) = \int_a^b X(t)dB(t).$$
3. (Zero Mean) $\mathbb{E}[\, \int_0^T X(t)dB(t)\,] = 0$.
4. (Isometry)
$$\mathbb{E}\left[\left(\int_0^T X(t)dB(t) \right)^2 \right] = \int_0^T \mathbb{E}\left[\, X^2(t)\,\right] dt.$$

The first two properties can be verified directly from the definition.

In order to show that the integral has zero mean, we first show that it has a mean.

$$\begin{aligned} \mathbb{E}\left[\, |\xi_i(B(t_{i+1}) - B(t_i))|\,\right] &\le \sqrt{\mathbb{E}\left[\, \xi_i^2\,\right] \mathbb{E}\left[\, (B(t_{i+1}) - B(t_i))^2\,\right]} \qquad \text{Cauchy-Schwartz} \\ &< \infty. \end{aligned}$$

This implies that

$$\begin{aligned} \mathbb{E}\left[\left| \int_0^T X(t)dB(t) \right| \right] &\le \sum_{i=0}^{n-1} \mathbb{E}\left[\, |\xi_i(B(t_{i+1}) - B(t_i))|\,\right] \\ &< \infty. \end{aligned}$$

This shows that the integral has expectation. By the martingale property and the fact that ξ_i's are $\mathcal{F}_{t_i}$-measurable,

$$\mathbb{E}\left[\, \xi_i(B(t_{i+1}) - B(t_i)) \mid \mathcal{F}_{t_i}\,\right] = \xi_i \mathbb{E}\left[\, B(t_{i+1}) - B(t_i) \mid \mathcal{F}_{t_i}\,\right] = 0.$$

Thus taking an expectation over each summand yields 0 in total.

To show the isometry property, first expand the square

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^T X(t) dB(t) \right)^2 \right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[\xi_i^2 (B(t_{i+1}) - B(t_i))^2 \right] + 2 \sum_{i<j} \mathbb{E}\left[\xi_i \xi_j (B(t_{i+1}) - B(t_i))(B(t_{j+1}) - B(t_j)) \right].
\end{aligned}$$

Using the martingale property of Brownian motion,

$$\begin{aligned}
\sum_{i=0}^{n-1} \mathbb{E}\left[\xi_i^2 (B(t_{i+1}) - B(t_i))^2 \right] &= \sum_{i=0}^{n-1} \mathbb{E}\left[\mathbb{E}\left[\xi_i^2 (B(t_{i+1}) - B(t_i))^2 \mid \mathcal{F}_{t_i} \right] \right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[\xi_i^2 \mathbb{E}\left[(B(t_{i+1}) - B(t_i))^2 \mid \mathcal{F}_{t_i} \right] \right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[\xi_i^2 \right] (t_{i+1} - t_i) \\
&=: \int_0^T \mathbb{E}\left[X^2(t) \right] dt.
\end{aligned}$$

By conditioning on $\mathcal{F}_j$ for $i < j$, we can show that

$$\mathbb{E}\left[\xi_i \xi_j (B(t_{i+1}) - B(t_i))(B(t_{j+1}) - B(t_j)) \right] = 0.$$

4.1.3 Adapted Processes

Definition 4.1.4 (Stochastic Integral)
Suppose $X_n \uparrow X$ in the sense that X_n are monotonically increasing and satisfy

$$\int_0^T \mathbb{E}\left[|X_n(t) - X(t)|^2 \right] \to 0.$$

By the completeness of L^2 and the isometry property,

$$\int_0^T X_n(t) dB(t) \xrightarrow{2} J.$$

The random variable J is taken to be the integral of $X(t)$.

Remark 4.1.5 In the case that $\int_0^T X^2(t)dt$ is finite but $\int_0^T \mathbb{E}[\, X^2(t)\,]\, dt$ is not, we can still approximate this process by taking limits in probability rather than mean squared. The sequence of corresponding Itô integrals is a Cauchy sequence in probability and converges in probability to a limit $\int_0^T X(t)dB(t)$ which we take as the Itô integral.

Example 4.1.6 ($\int_0^T B(t)dB(t)$)
Let $0 = t_0^n < t_1^n < \cdots < t_n^n = T$ be a partition of $[0, T]$. Define

$$X_n(t) := \sum_{i=0}^{n-1} B(t_i^n)\mathbb{1}_{(t_i^n, t_{i+1}^n]}(t).$$

Then each $X_n(t)$ is a simple adapted process. By the continuity of $B(t)$ (uniform continuity on $[0, T]$),

$$X_n(t) \xrightarrow{a.s.} B(t)$$

as $n \to \infty$ since $\delta_n \to 0$. The Itô integral of $X^n(t)$ is given by

$$\int_0^T X_n(t)dB(t) = \sum_{i=0}^{n-1} B(t_i^n)(B(t_{i+1}^n) - B(t_i^n)).$$

We claim that this sequence of integrals converge in mean squared to

$$J := \frac{1}{2}B^2(T) - \frac{1}{2}T.$$

By adding and subtracting $B^2(t_{i+1}^n)$, we see that

$$B(t_i^n)(B(t_{i+1}^n) - B(t_i^n)) = \frac{1}{2}[B^2(t_{i+1}^n) - B^2(t_i^n) - (B(t_{i+1}^n) - B(t_i^n))^2]$$

and

$$\begin{aligned}\int_0^T X_n(t)dB(t) &= \frac{1}{2}\sum_{i=0}^{n-1}(B^2(t_{i+1}^n) - B^2(t_i^n)) - \frac{1}{2}\sum_{i=0}^{n-1}(B(t_{i+1}^n) - B(t_i^n))^2 \\ &= \frac{1}{2}B^2(T) - \frac{1}{2}B^2(0) - \frac{1}{2}\sum_{i=0}^{n-1}(B(t_{i+1}^n) - B(t_i^n))^2.\end{aligned}$$

By our computation of the quadratic variation of Brownian motion, the second sum converges in mean squared to T. This concludes the proof.

We state the following remarks.

1. If $X(t)$ is a function of finite variation, the stochastic integral can be defined using

integration by parts:

$$\int_0^T X(t)dB(t) = X(T)B(T) - X(0)B(0) - \int_0^T B(t)dX(t).$$

However, this approach fails when $X(t)$ depends on $B(t)$.

2. Brownian motion has no derivative, but it has a generalized derivative as a *Schwartz distribution*. It is defined by the following relation. For any smooth function g with compact support,

$$\int g(t)B'(t)dt = -\int B(t)g'(t)dt.$$

Again, this approach fails when $g(t)$ depends on $B(t)$.

Theorem 4.1.7
Let $X(t)$ be a regular adapted process such that $\int_0^T X^2(t)dt \overset{a.s.}{<} \infty$. Then the Itô integral $\int_0^T X(t)dB(t)$ is well-defined and has the following properties.

1. Linearity
2. $\int_0^T X(t)\mathbb{1}_{(a,b]}dB(t) = \int_a^b X(t)dB(t)$.

If $\int_0^T \mathbb{E}\left[\, X^2(t)\, \right] dt \overset{a.s.}{<} \infty$, the two following properties also hold.

3. (Zero Mean) The integral has mean 0.
4. (Isometry)

$$\mathbb{E}\left[\left(\int_0^T X(t)dB(t)\right)^2\right] = \int_0^T \mathbb{E}\left[\, X^2(t)\, \right] dt.$$

Note that the Itô integral need not have mean or variance but then it does, the mean is zero and the variance can be computed using the isometry property.

Corollary 4.1.8
If X is a continuous adapted process, then the Itô integral $\int_0^T X(t)dB(t)$ exists. In particular, $\int_0^T f(B(t))dB(t)$ is well-defined for any continuous $f : \mathbb{R} \to \mathbb{R}$.

Remark 4.1.9 It follows from the definition of the Itô integral that the sums approximate the Itô integral $\int_0^T X(t)dB(t)$.

$$\sum_{k=0}^{n-1} X(t_i^n)[B(t_{i+1}^n) - B(t_i^n)].$$

In an approximation of the Stieltjes integral by sums, the function on an interval $[t_i, t_{i+1}]$ of the partition is replaced by its value at some midpoint $\theta_i \in [t_i, t_{i+1}]$. For the Itô integral, it was important to choose the left endpoint, otherwise the process may not be adapted. It is possible to define anotehr integral by choosing $\theta_i := \lambda t_i + (1-\lambda)t_{i+1}$ for some $\lambda \in (0,1)$. When $\lambda = 1/2$, the *Stratonovich stochastic integral* results.

Remark 4.1.10 The Itô integral does not have the monotonicity property. Indeed,

$$\int_0^t 1 \cdot dB(t) = B(1) \sim \mathcal{N}(0,1),$$

which is negative with probability $1/2$.

Example 4.1.11
Consider $f(t) = \exp(t)$. We have

$$\begin{aligned}\mathbb{E}\left[\int_0^1 \exp(2B(t))dt\right] &= \int_0^1 \mathbb{E}\left[\, \exp(2B(t)) \,\right] dt \\ &= \int_0^1 \exp(2t)dt \\ &= \frac{1}{2}(e^2 - 1) \\ &< \infty.\end{aligned}$$

Thus the stochastic integral has mean zero and variance $(e^2-1)/2$.

Example 4.1.12
For $f(t) = t$,

$$\int_0^1 \mathbb{E}[\, B^2(t) \,]\, dt = \int_0^t t dt = 1/2 < 0.$$

Thus the integral $\int_0^1 B(t)dB(t)$ has mean zero and variance $1/2$.

Example 4.1.13
Take $f(t) = \exp(t^2)$. It can be computed that $\int_0^1 \mathbb{E}\left[\, \exp(2B^2(t)) \,\right] dt = \infty$. Thus we cannot claim the Itô has finite moments. In fact, using martingale inequalities, it can be shown that the expectation of the Itô integral does not exist.

Example 4.1.14
Let $J := \int_0^1 t dB(t)$. Since $\int_0^1 t^2 dt < \infty$, the Itô integral is defined. Since the integrand is non-random,

$$\mathbb{E}\left[\, J \,\right] = 0$$

$$\mathbb{E}\left[\, J^2 \,\right] = \int_0^1 t^2 dt = \frac{1}{3}.$$

Example 4.1.15
Consider $\int_0^1 (1-t)^{-\alpha} dB(t)$. In order for the integral to be defined, we must have

$$\int_0^1 (1-t)^{-2\alpha} dt < \infty.$$

This gives $\alpha < 1/2$.

The following result is a consequence of the isometry property.

Theorem 4.1.16
Let $X(t), Y(t)$ be regular adapted processes such that $\mathbb{E}[\int_0^T X^2(t)dt] < \infty$ and $\mathbb{E}[\int_0^T Y^2(t)dt] < \infty$. Then

$$\mathbb{E}\left[\int_0^T X(t)dB(t) \cdot \int_0^T Y(t)dB(t)\right] = \int_0^T \mathbb{E}[X(t)Y(t)]\, dt.$$

Proof
Write I_1, I_2 as the Itô integral of X, Y respectively. We have

$$\begin{aligned}
&\mathbb{E}[I_1 I_2] \\
&= \frac{1}{2}\mathbb{E}\left[(I_1+I_2)^2\right] - \frac{1}{2}\mathbb{E}\left[I_1^2\right] - \frac{1}{2}\mathbb{E}\left[I_2^2\right] \\
&= \frac{1}{2}\int_0^T \mathbb{E}\left[(X(t)+Y(t))^2\right] dB(t) \\
&\quad + \frac{1}{2}\int_0^T \mathbb{E}\left[X^2(t)\right] dB(t) + \frac{1}{2}\int_0^T \mathbb{E}\left[Y^2(t)\right] dB(t) && \text{isometry} \\
&= \int_0^T \mathbb{E}[X(t)Y(t)]\, dt. && \text{linearity}
\end{aligned}$$

□

4.2 Itô Integral Process

Let X be a regular adapted process such that $\int_0^T X^2(s)ds \overset{a.s.}{<} \infty$. Thus $\int_0^t X(s)dB(s)$ is defined for any $t \leq T$. We can then define the *Itô integral process* as

$$Y(t) = \int_0^t X(s)dB(s).$$

It is possible to show that there is a version of the Itô integral $Y(t)$ with continuous sample paths. We always assume the continuous version is taken. Moreover, we will see that the

Itô integral process has positive quadratic variation but infinite variation.

4.2.1 Martingale Property

Itô integrals of simple processes are clearly adapted and continuous. Since $Y(t)$ is a limit of such integrals, it is itself adapted.

Suppose now that $\int_0^T X^2(s)ds \overset{a.s.}{<} \infty$ and in addition $\int_0^T \mathbb{E}[\, X^2(s) \,]\, ds < \infty$. Then $Y(t)$ possesses the first two moments. It can be shown, first for simple processes and then in general, that for $s < t$,

$$\mathbb{E}\left[\int_s^t X(u)dB(u) \,\middle|\, \mathcal{F}_s \right] = 0.$$

Hence

$$\begin{aligned}
\mathbb{E}\,[\, Y(t) \mid \mathcal{F}_s \,] &:= \mathbb{E}\left[\int_0^t X(u)dB(u) \,\middle|\, \mathcal{F}_s \right] \\
&= \int_0^s X(u)dB(u) + \mathbb{E}\left[\int_s^t X(u)dB(u) \,\middle|\, \mathcal{F}_s \right] \\
&= \int_0^s X(u)dB(u) \\
&= Y(s).
\end{aligned}$$

Therefore $Y(t)$ is a martingale.

The second moments of $Y(t)$ are given by the isometry property. In particular,

$$\begin{aligned}
\sup_{t \le T} \mathbb{E}\left[\, Y^2(t) \,\right] &= \sup_{t \le T} \mathbb{E}\left[\left(\int_0^t X(s)dB(s) \right)^2 \right] \\
&= \int_0^T \mathbb{E}\left[\, X^2(s) \,\right] ds \\
&< \infty.
\end{aligned}$$

Definition 4.2.1 (Square Integrable)
A martingale is *square integrable* on $[0, T]$ if its second moments are bounded.

In summary of the discussion above, we have the following result.

Theorem 4.2.2
Let $X(t)$ be a regular adapted process such that $\int_0^T \mathbb{E}[\, X^2(s) \,]\, ds < \infty$. Then the Itô integral process $Y(t)$ is a continuous zero mean square integrable martingale.

We note that if $\int_0^t \mathbb{E}[\,X^2(s)\,]\,ds = \infty$, the Itô integral can fail to be a martingale, but it is always a *local martingale*, which we will see later.

We now have a way to construct martingales.

Corollary 4.2.3
For any bounded function f with discontinuities of the first kind on $\mathbb{R}$,

$$\int_0^t f(B(s))dB(s)$$

is a square integrable martingale.

Proof
$X(t) := f(B(t))$ is adapted and regular. Since $|f| \leq K \in \mathbb{R}_+$, $\int_0^T \mathbb{E}[\,f^2(B(s))\,]\,ds \leq KT$. The result follows by the previous theorem. $\square$

4.2.2 Quadratic Variation & Covariation of Itô Integrals

The Itô integral $Y(t)$ of $X(t)$ is a random function of t which is continuous and adapted. The quadratic variation of Y is thus defined

$$[Y,Y](t) := \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} [Y(t_{i+1}^n) - Y(t_i^n)]^2,$$

where the limit is taken in probability over all shrinking partitions.

Theorem 4.2.4
The quadratic variation of the Itô integral $\int_0^t X(s)dB(s)$ is given by

$$\left[\int_0^t X(s)dB(s), \int_0^t X(s)dB(s)\right](t) = \int_0^t X^2(s)ds.$$

This result can be proven first for simple processes and then in general by approximations for simple processes.

Example 4.2.5
The quadratic variation of the Itô integral for Brownian motion is given by

$$\left[\int_0^t B(s)dB(s)\right] = \int_0^t B^2(s)ds.$$

Corollary 4.2.6
If $\int_0^t X^2(s)ds \overset{a.s.}{>} 0$ for all $t \leq T$, then the Itô integral has infinite variation on $[0, t]$ for all $t \leq T$.

Proof
If its variation is finite, then its quadratic variation would be zero, which is a contradiction. □

Akin to Brownian motion, the Itô integral $Y(t)$ is a continuous but nowhere differentiable function of t. Suppose now that Y_1, Y_2 are Itô integrals of X_1, X_2, with respect to the same Brownian motion B. Then the process Y_1, Y_2 is an Itô integral of $X_1 + X + 2$ with respect to B.

Definition 4.2.7 (Quadratic Covariation)
The quadratic covariation of the Itô integral $Y_i(t) := \int_0^t X_i(s)dB(s)$ for $i = 1, 2$ is defined by

$$[Y_1, Y_2](t) := \frac{1}{2}\left([Y_1 + Y_2, Y_1 + Y_2] - [Y_1, Y_1](t) - [Y_2, Y_2](t)\right).$$

By our previous identity, it follows that

$$[Y_1, Y_2](t) = \int_0^t X_1(s)X_2(s)ds.$$

Remark that the quadratic covariation of Y_1, Y_2 is symmetric. Moreover, it can be shown that the quadratic covariation is the limit in probability of products of increments of the processes Y_1, Y_2 over shrinking partitions.

$$[Y_1, Y_2](t) = \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} (Y_1(t_{i+1}^n) - Y_1(t_i^n))(Y_2(t_{i+1}^n) - Y_2(t_i^n)).$$

4.3 Itô Integral & Gaussian Processes

We have seen that the Itô integral of simple deterministic processes is a normal random variable. It can be shown using moment generating functions that a limit in probability of Gaussians is again Gaussian. This implies the following result.

Theorem 4.3.1
If $X(t)$ is a deterministic function such that $\int_0^T X^2(s)ds < 0$, then its Itô integral $Y(t)$ is a Gaussian process with zero mean and covariance function

$$\operatorname{Cov}\left[\, Y(t), Y(t+u) \,\right] = \int_0^t X^2(s)ds.$$

Moreover, $Y(t)$ is a square integrable martingale.

Proof
Since the integrand is deterministic, we certainly have

$$\int_0^t \mathbb{E}\left[\, X^2(s) \,\right] ds = \int_0^t X^2(s)ds < \infty.$$

By the zero mean property, Y has zero mean. By computation,

$$\begin{aligned}
&\operatorname{Cov}\left[\, Y(t), Y(t+u) \,\right] \\
&= \mathbb{E}\left[\left(\int_0^t X(s)dB(s) \right)^2 \right] + \mathbb{E}\left[\int_0^t X(s)dB(s) \mathbb{E}\left[\int_t^{t+u} X(s)dB(s) \,\middle|\, \mathcal{F}_t \right] \right] \\
&= \int_0^t \mathbb{E}\left[\, X^2(s) \,\right] ds && \text{isometry} \\
&= \int_0^t X^2(s)ds. && \square
\end{aligned}$$

A proof of normality of integrals of non-random processes will be done later using Itô's formula.

Example 4.3.2
$J := \int_0^t s dB(s) \sim \mathcal{N}(0, t^3/3)$.

If $Y(t) = \int_0^t X(t,s)dB(s)$ where $X(t,s)$ can depend on the upper integration limit t, then $Y(t)$ need not be a martingale. However, it remains a Gaussian process.

Theorem 4.3.3
For any $t \le T$, Let $X(t,s)$ be a regular deterministic function with $\int_0^t X^2(t,s)ds < \infty$. Then the process $Y(t) := \int_0^t X(t,s)dB(s)$ is a Gaussian process with zero mean and covariance function

$$\operatorname{Cov}[\,Y(t), Y(t+u)\,] = \int_0^t X(t,s)X(t+u,s)ds.$$

Proof (Sketch)
We omit the proof that the process is Gaussian. It can be seen by approximating $X(t,s)$ by functions of the form $f(t)g(s)$.

Since the function is deterministic, the mean is zero and the covariance is computed as

$$\begin{aligned}
&\operatorname{Cov}[\,Y(t), Y(t+u)\,] \\
&= \mathbb{E}[\,Y(t)Y(t+u)\,] \\
&= \mathbb{E}\left[\int_0^t X(t,s)dB(s)\int_0^t X(t+u,s)dB(s)\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_0^t X(t,s)dB(s)\int_t^{t+u} X(t+u,s),dB(s)\,\middle|\,\mathcal{F}_t\right]\right] \\
&= \int_0^t X(t,s)X(t+u,s)ds.
\end{aligned}$$

□

4.4 Itô's Formula for Brownian Motion

Theorem 4.4.1
If g is a continuous function and $\{\,t_i^n\,\}$ forms a partition of $[0,t]$, then for any $\theta_i^n \in (B(t_i^n), B(t_{i+1}^n))$, the limit in probability

$$\lim_{\delta_n\to 0}\sum_{i=0}^{n-1} g(\theta_i^n)[B(t_{i+1}^n) - B(t_i^n)]^2 = \int_0^t g(B(s))ds.$$

Proof
We begin by assuming g has compact support. Intuitively, this suffices since Brownian motion is bounded with high probability.

Take first $\theta_i^n = B(t_i^n)$ to be the left end of the interval. By the continuity of $g(B(t))$ and the definition of the Riemann integral,

$$\sum_{i=0}^{n-1} g(B(t_i^n))[t_{i+1}^n - t_i^n] \to \int_0^t g(B(s))ds.$$

Define $\Delta B_i^n := B(t_{i+1}^n) - B(t_i^n)$ and $\Delta t_i^n := t_{i+1}^n - t_i^n$. We claim that

$$\sum_{i=0}^{n-1} g(B(t_i^n))\Delta B_i - \sum_{i=0}^{n-1} g(B(t_i^n))\Delta t_i \xrightarrow{2} 0$$

so that the sum in question converges to 0 in mean squared as desired.

By conditioning, it can be shown that

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{i=0}^{n-1} g(B(t_i^n))[(\Delta B_i^n)^2 - \Delta t_i^n]\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{n-1} g^2(B(t_i^n)) \cdot \mathbb{E}\left[\left((\Delta B_i^n)^2 - \Delta t_i^n\right)^2 \;\middle|\; \mathcal{F}_{t_i^n}\right]\right] \\
&= 2\mathbb{E}\left[\sum_{i=0}^{n-1} g^2(B(t_i^n))(\Delta t_i)^2\right] \\
&\leq \delta_n 2\mathbb{E}\left[\sum_{i=0}^{n-1} g^2(B(t_i^n))(\Delta t_i)\right] \\
&\to 0.
\end{aligned}$$

The second equality can be computed as follows.

$$\begin{aligned}
\mathbb{E}\left[\left((\Delta B_i^n)^2 - \Delta t_i\right)^2 \;\middle|\; \mathcal{F}_{t_i^n}\right] &= \mathbb{E}\left[(\Delta B_i^n)^4\right] - 2\Delta t_i \mathbb{E}\left[(\Delta B_i^n)^2\right] + (\Delta t_i)^2 \quad \Delta B_i^n \perp \mathcal{F}_{t_i} \\
&= 3(\Delta t_i^n)^2 - 2(\Delta t_i)^2 + (\Delta t_i)^2.
\end{aligned}$$

This proves the claim.

Now for any valid choice of θ_i^n, as $\theta_n \to 0$,

$$\begin{aligned}
&\sum_{i=0}^{n-1}[g(\theta_i^n) - g(B(t_i^n))][B(t_{i+1}^n) - B(t_i^n)]^2 \\
&\leq \max_i [g(\theta_i^n) - g(B(t_i^n))] \sum_{i=0}^{n-1}[B(t_{i+1}^n) - B(t_i^n)]^2 \\
&\to 0.
\end{aligned}$$

The first term in the product converges to 0 a.s. by (uniform) continuity and the second converges in mean squared to t, the quadratic variation of Brownian motion. Thus moving from the left endpoint to and arbitrary point in the partition does not change the limit in mean square.

Now for general g, take the stopping time

$$\tau(L) := \inf \{ s : |B_s| > L \} .$$

Our work above shows that the function $g(s)\mathbb{1}\{ s \leq \tau(L) \}$ satisfies

$$\begin{aligned}
&\lim_{\delta_n \to 0} \sum_{i=0}^{n-1} g(\theta_i^n)\mathbb{1}\{ \theta_i^n \leq \tau(L) \} [B(t_{i+1}^n) - B(t_i^n)]^2 \\
&= \int_0^t g(B(s))\mathbb{1}\{ s \leq \tau(L) \} ds.
\end{aligned}$$

For any $\varepsilon > 0$, there is some L_ε such that $\mathbb{P}\{ \tau(L_\varepsilon) < t \} < \varepsilon$. In such an event, we are done.

Thus the a.s. convergence we derived above for functions of compact support is weakened to convergence in probability for general continuous functions. □

Theorem 4.4.2
Let $B(t)$ be a Brownian motion on $[0, T]$ and $f \in C^2(\mathbb{R})$. For any $t \leq T$,

$$f(B(t)) = f(0) + \int_0^t f'(B(s))dB(s) + \frac{1}{2}\int_0^t f''(B(s))ds.$$

Proof
Note both integrals are well-defined since all functions in question are differentiable. Let $\{ t_i^n \}$ be a partition of $[0, t]$. We have

$$\begin{aligned}
&f(B(t)) \\
&= f(0) + \sum_{i=0}^{n-1} [f(B(t_{i+1}^n)) - f(B(t_i^n))] \\
&= f(0) + \sum_{i=0}^{n-1} f'(B(t_i^n))[B(t_{i+1}^n) - B(t_i^n)] \\
&\quad + \frac{1}{2}\sum_{i=0}^{n-1} f''(\theta_i^n)[B(t_{i+1}^n) - B(t_i^n)]^2 \qquad \theta_i^n \in (B(t_i^n), B(t_{i+1}^n)).
\end{aligned}$$

Taking limits as $\delta_n \to 0$, the first sum converges in probability to the Itô integral

$\int_0^t f'(B(s))dB(s)$. By the previous theorem, the second converges in probability to $\int_0^t f''(B(s))ds$.□

Example 4.4.3
Let $f(x) := x^m, m \geq 2$. We have

$$B^m(t) = m\int_0^t B^{m-1}(s)dB(s) + \frac{m(m-1)}{2}\int_0^t B^{m-2}(s)ds.$$

For the specific case of $m = 2$, we have

$$B^2(t) = 2\int_0^t B(s)dB(s) + t.$$

Rearranging recovers the stochastic integral of Brownian motion $\int_0^t B(s)dB(s)$.

Example 4.4.4
Let $f(x) = \exp(x)$. We have

$$e^{B(t)} = 1 + \int_0^t e^{B(s)}dB(s) + \frac{1}{2}\int_0^t e^{B(s)}ds.$$

4.5 Itô Processes & Stochastic Differentials

4.5.1 Itô Processes

Definition 4.5.1 (Itô Process)
An Itô process has the form

$$Y(t) = Y(0) + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dB(s)$$

for $t \in [0, T]$. $Y(0)$ is $\mathcal{F}_0$-measurable and processes $\mu(t), \sigma(t)$ are $\mathcal{F}_t$-adapted such that

$$\int_0^T |\mu(t)|dt, \int_0^T \sigma^2(t)dt \overset{a.s.}{<} \infty,$$

It is said that an Itô process $Y(t)$ has the *stochastic differential*

$$dY(t) = \mu(t)dt + \sigma(t)dB(t)$$

for $t \in [0, T]$.

Remark 4.5.2 The stochastic differential only has meaning by way of integrals and no other.

Note that the processes μ, σ may depend on $Y(s), B(s), s \leq t$. For exampple, it can depend on the maximum of Brownian motion.

> **Example 4.5.3**
> With $Y(t) = B^2(t)$, $\mu(s) = 1$ and $\sigma(s) = 2B(s)$, we can write
>
> $$\begin{aligned} Y(t) &= B^2(t) \\ &= t + 2\int_0^t B(s)dB(s) \\ &= \int_0^t \mu(s)ds + \int_0^t \sigma(s)dB(s). \end{aligned}$$
>
> In other words,
> $$d[B^2(t)] = 2B(t)dB(t) + dt.$$
> Again, this only has meaning by the integral relation.

Remark 4.5.4 (Itô's Formula for Stochastic Differentials) In differential notation, Itô's formula states that for any $f \in C^2(\mathbb{R})$,

$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt.$$

> **Example 4.5.5**
> The following relations hold.
>
> (a) $de^{B(t)} = e^{B(t)}dB(t) + \frac{1}{2}e^{B(t)}dt$
>
> (b) $d\sin(B(t)) = \cos(B(t))dB(t) - \frac{1}{2}\sin(B(t))dt$
>
> (c) $d\cos(B(t)) = -\sin(B(t))dB(t) - \frac{1}{2}\cos(B(t))dt$
>
> (d) $de^{iB(t)} = ie^{iB(t)}dB(t) - \frac{1}{2}e^{iB(t)}dt$
>
> The application of Itô's formula to complex-valued functions formally means its application to the real and complex parts of the function. At times, we can guess the result by treating i as a constant.

4.5.2 Quadratic Variation of the Itô Process

Let $Y(t)$ be an Itô process

$$Y(t) = Y(0) + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dB(s),$$

where it is assumed that μ, σ are such that the integrals in question are defined. The following hold.

1. $Y(t)$ is continuous a.s.
2. $\int_0^t \mu(s)ds$ is a continuous function of t. Moreover, any Riemann integrable function is continuous a.e. Thus the integral is differentiable a.e. and is therefore of finite variation.
3. $\int_0^t \sigma(s)dB(s)$ is continuous.
4. $Y(t)$ is of finite variation if and only if $\int_0^t \sigma(s)dB(s)$ is of finite variation.

Recall that the quadratic variation of Y defined by

$$[Y](t) = [Y, Y](t) = \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} [Y(t_{i+1}^n) - Y(t_i^n)]^2.$$

Here the limit is taken in probability over shrinking partitions. By expanding the covariation $[Y, Y](t)$, we have

$$\begin{aligned}
&[Y](t) \\
&= \left[\int_0^t \mu(s)ds\right](t) + 2\left[\int_0^t \mu(s)ds, \int_0^t \sigma(s)dB(s)\right](t) + \left[\int_0^t \sigma(s)dB(s)\right](t) \\
&= \left[\int_0^t \sigma(s)dB(s)\right](t) \\
&= \int_0^t \sigma^2(s)ds.
\end{aligned}$$

The second equality follows from the fact that the covariation between a continuous function and a function of finite variation is 0. The last equality follows from a previous result on the quadratic variation of Itô integrals.

If $Y(t), X(t)$ have stochastic differentials with respect to the same Brownian motion, then $Y(t) + X(t)$ also has a stochastic differential with respect to the same Brownian motion. We can then define the covariation of X, Y on $[0, t]$, similar to before by

$$\begin{aligned}
&[X, Y](t) \\
&= \frac{1}{2}[X + Y, X + Y](t) - \frac{1}{2}[X, X](t) - \frac{1}{2}[Y, Y](t) \\
&= \int_0^t \frac{1}{2}(\sigma_X(s) + \sigma_Y(s))^2 - \frac{1}{2}\sigma_X^2(s) - \frac{1}{2}\sigma_Y^2(s)ds \\
&= \int_0^t \sigma_X(s)\sigma_Y(s)ds \\
&= \left[\int_0^t \sigma_X(s)dB(s), \int_0^t \sigma_Y(s)dB(s)\right](t).
\end{aligned}$$

Theorem 4.5.6
If X, Y are Itô processes and X is of finite variation, then

$$[X, Y](t) = 0.$$

Example 4.5.7
Let $X(t) := \exp(t)$ and $Y(t) = B(t)$. Then

$$[X, Y](t) = 0.$$

We introduce a notation that allows formal manipulation with stochastic differentials.

$$\begin{aligned} dY(t)dX(t) &:= d[X, Y](t) \\ (dY(t))^2 &:= d[Y, Y](t). \end{aligned}$$

Again, recall that stochastic differentials only have meaning from the integral interpretation,

$$[X, Y](t) = \int_0^t d[X, Y](t) = \int_0^t dX(t)dY(t).$$

Example 4.5.8
Since $X(t) := t$ is a continuous function of finite variation and $Y(t) := B(t)$ is continuous with quadratic variation t, the following hold.

$$\begin{aligned} dB(t)dt &= 0 \\ (dt)^2 &= 0 \\ (dB(t))^2 &= d[B, B](t) \\ &= dt. \end{aligned}$$

4.5.3 Integrals with respect to Itô Processes

We have defined stochastic integrals with respect to Brownian motion. It is necessary to extend integration with respect to processes obtained from Brownian motion. Let the Itô process $Y(t) := \int_0^t X(s)dB(s)$ be defined for all $t \leq T$, where $X(t)$ is an adapted process such that $\int_0^T X^2(s)ds \overset{a.s.}{<} \infty$.

Consider now an adapted process $H(t)$ such that $\int_0^t H^2(s)X^2(s)ds \overset{a.s.}{<} \infty$. The Itô integral $Z(t) := \int_0^t H(s)X(s)dB(s)$ is also defined for all $t \leq T$. By identifying $dY(t)$ and $X(t)dB(t)$, we can formally define

$$\int_0^t H(s)dY(s) := \int_0^t H(s)X(s)dB(s).$$

We note that integrals with respect to $Y(t)$ can be defined in a direct way but the result agrees with the definition above.

Moreover generally, we have the following definition.

Definition 4.5.9
If Y is an Itô process satisfying

$$dY(t) = \mu(t)dt + \sigma(t)dB(t)$$

and H is adapted and satisfies $\int_0^t H^2(s)\sigma^2(s)ds \overset{a.s.}{<} \infty$ as well as $\int_0^t |H(s)\mu(s)|ds \overset{a.s.}{<} \infty$, Then $Z(t) = \int_0^t H(s)dY(s)$ is defined as

$$\int_0^t H(s)dY(s) := \int_0^t H(s)\mu(s)ds + \int_0^t H(s)\sigma(s)dB(s).$$

4.6 Itô's Formula for Itô Processes

Theorem 4.6.1
Let $X(t)$ have a stochastic differential

$$dX(t) = \mu(t)dt + \sigma(t)dB(t)$$

for $0 \le t \le T$. Suppose $f \in C^2(\mathbb{R})$. Then $Y(T) := f(X(t))$ has stochastic differential

$$\begin{aligned}
& df(X(t)) \\
& = f'(X(t))dX(t) + \frac{1}{2}f''(X(t))d[X,X](t) \\
& = f'(X(t))dX(t) + \frac{1}{2}f''(X(t))\sigma^2(t)dt \\
& = \left(f'(X(t))\mu(t) + \frac{1}{2}f''(X(t))\sigma^2(t)\right)dt + f'(X(t))\sigma(t)dB(t).
\end{aligned}$$

In integral notation, the above means

$$f(X(t)) = f(X(0)) + \int_0^t f'(X(s))dX(s) + \frac{1}{2}\int_0^t f''(X(s))\sigma^2(s)ds.$$

The proof is similar to that of Itô's formula for stochastic integrals and is omitted.

Example 4.6.2
Let $X(t)$ have stochastic differential

$$dX(t) = X(t)dB(t) + \frac{1}{2}X(t)dt.$$

Let us find a positive process X satisfying the above.

By Itô's formula for $\ln X(t)$,

$$\begin{aligned} d\ln X(t) &= \frac{1}{X(t)}dX(t) - \frac{1}{2X(t)^2}\sigma_X^2(t)dt \\ &= \frac{1}{X(t)}\left(X(t)dB(t) + \frac{1}{2}X(t)dt\right) - \frac{1}{2X^2(t)}X^2(t)dt \\ &= dB(t). \end{aligned}$$

It follows that

$$\begin{aligned} \ln X(t) &= \ln X(0) + B(t) \\ X(t) &= X(0)e^{B(t)}. \end{aligned}$$

We can verify that this choice of $X(t)$ indeed satisfies the desired stochastic differential through Itô's formula.

4.6.1 Integration by Parts

We give a representation of the quadratic covariation $[X, Y](t)$ of two Itô processes $X(t), Y(t)$ in terms of Itô integrals. This representation gives rise to the integration by parts formula. Note that the following section is not rigorous but can be made so by making the arguments more precise, by using Itô's formual for the function xy of two variables, or by approximations by simple processes.

Quadratic covariation is a limit in probability over decreasing partitions of $[0, t]$,

$$[X, Y](t) = \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} [X(t_{i+1}^n - X(t_i^n))][Y(t_{i+1}^n) - Y(t_i^n)].$$

The RHS sum can be rewritten as follows.

$$\begin{aligned}
&\sum_{i=0}^{n-1}[X(t_{i+1}^n)Y(t_{i+1}^n) - X(t_i^n)Y(t_i^n)] \\
&\quad - \sum_{i=0}^{n-1} X(t_i^n)[Y(t_{i+1}^n) - Y(t_i^n)] - \sum_{i=0}^{n-1} Y(t_i^n)[X(t_{i+1}^n) - X(t_i^n)] \\
&= X(t)Y(t) - X(0)Y(0) \\
&\quad - \sum_{i=0}^{n-1} X(t_i^n)[Y(t_{i+1}^n) - Y(t_i^n)] - \sum_{i=0}^{n-1} Y(t_i^n)[X(t_{i+1}^n) - X(t_i^n)].
\end{aligned}$$

The last two sums converge in probability to the Itô integrals $\int_0^t X(s)dY(s)$ and $\int_0^t Y(s)dX(s)$, respectively. This yields the integration by parts (stochastic product rule)

$$X(t)Y(t) - X(0)Y(0) = \int_0^t X(s)dY(s) + \int_0^t Y(s)dX(s) + [X,Y](t).$$

In differential notation, this reads

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + d[X,Y](t).$$

Remark 4.6.3 This provides yet another representation for quadratic variation

$$[X,X](t) = X^2(t) - X^2(0) - 2\int_0^t X(s)dX(s).$$

Quadratic variation is non-decreasing in t and is thus of finite variation. By the formula above, it must also be continuous. By the polarization identity, covariation is also continuous and of finite variation.

If

$$\begin{aligned}
dX(t) &= \mu_X(t)dt + \sigma_X(t)dB(t) \\
dY(t) &= \mu_Y(t)dt + \sigma_Y(t)dB(t),
\end{aligned}$$

then their covariation can be formally obtained by multiplying dX, dY,

$$\begin{aligned}
d[X,Y](t) &= dX(t)dY(t) \\
&= \sigma_X(t)\sigma_Y(t)(dB(t))^2 \\
&= \sigma_X(t)\sigma_Y(t)dt.
\end{aligned}$$

This leads to the formula

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + \sigma_X(t)\sigma_Y(t)dt.$$

Note that if one of the processes is of finite variation, then the covariation term is zero. Thus for such processes, the stochastic product rule is the same as usual.

Example 4.6.4 (Stochastic Quotient Rule)
Let us compute the stochastic differential

$$\begin{aligned}
&d(X(t)/Y(t)) \\
&= X(t)d\frac{1}{Y(t)} + \frac{1}{Y(t)}dX(t) + d[X, Y^{-1}(t)](t) \\
&= X(t)\left(-Y^{-2}(t)dY(t) + Y^{-3}(t)d[Y,Y](t)\right) \\
&\quad + \frac{1}{Y(t)}dX(t) \\
&\quad + dX(t)\left(-Y^{-2}(t)dY(t) + Y^{-3}(t)d[Y,Y](t)\right) \\
&= -\frac{X(t)}{Y^2(t)}dY(t) + \frac{X(t)}{Y^3(t)}d[Y,Y](t) + \frac{1}{Y(t)}dX(t) - \frac{1}{Y^2(t)}d[X,Y](t) \\
&= \frac{X(t)}{Y(t)}\left(\frac{1}{X(t)}dX(t) - \frac{1}{Y(t)}dY(t) - \frac{1}{X(t)Y(t)}d[X,Y](t) + \frac{1}{Y^2(t)}d[Y,Y](t)\right).
\end{aligned}$$

Example 4.6.5
Suppose $X(t)$ has stochastic differential

$$dX(t) = B(t)dt + tdB(t)$$

with the initial condition $X(0) = 0$.

Consider $X(t) = tB(t)$. This satisfies the above equation by the integration by parts formula. Thus $X(t)$ is Gaussian, with mean zero, and covariance function

$$\begin{aligned}
\gamma(t,s) &= \operatorname{Cov}[\,X(t), X(s)\,] \\
&= \mathbb{E}[\,X(t)X(s)\,] \\
&= ts\mathbb{E}[\,B(t)B(s)\,] \\
&= ts\operatorname{Cov}[\,B(t)B(s)\,] \\
&= ts\min(t,s).
\end{aligned}$$

Example 4.6.6
Let $Y(t)$ have stochastic differential

$$dY(t) = \frac{1}{2}Y(t)dt + Y(t)dB(t)$$

subject to $Y(0) = 1$. We have shown before that $Y(t) = e^{B(t)}$ satisfies the stochastic differential above.

Let $X(t) := tB(t)$. We find $d(X(t)Y(t))$. In order to apply the product rule, we need to determine $d[X,Y](t)$.

$$\begin{aligned}
d[X,Y](t) &= dX(t)dY(t) \\
&= (B(t)dt + tdB(t))\left(\frac{1}{2}Y(t)dt + Y(t)dB(t)\right) \\
&= \frac{1}{2}B(t)Y(t)(dt)^2 + \left(B(t)Y(t) + \frac{1}{2}tY(t)\right)dB(t)dt + tY(t)(dB(t))^2 \\
&= tY(t)dt.
\end{aligned}$$

The last equality follows since the covariation between a continuous function and a function of bounded variation is zero.

Thus we can apply the product rule

$$\begin{aligned}
d(X(t)Y(t)) &= X(t)dY(t) + Y(t)dX(t) + d[X,Y](t) \\
&= X(t)dY(t) + Y(t)dX(t) + tY(t)dt.
\end{aligned}$$

By substituting the expressions for X and Y, the solution is obtained.

Example 4.6.7
Let $f \in C^2(\mathbb{R})$ and $B(t)$ a Brownian motion. We find the quadratic covariation $[f(B), B](t)$.

By heuristics,

$$d[f(B), B](t) = df(B(t))dB(t).$$

By Itô's formula,

$$df(B(t)) = df'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt.$$

It follows that

$$\begin{aligned}
d[f(B), B](t) &= df(B(t))dB(t) \\
&= f'(B(t))(dB(t))^2 + \frac{1}{2}f''(B(t))dB(t)dt \\
&= f'(B(t))dt.
\end{aligned}$$

This holds since $(dB)^2 = dt$ and $dBdt = 0$. Thus

$$[f(B), B](t) = \int_0^t f'(B(s))ds.$$

In a more intuitive way, from the definition of the covariation, taking limits over shrinking partitions yields

$$\begin{aligned}
[f(B), B](t) &= \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} [f(B(t_{i+1}^n)) - f(B(t_i^n))][B(t_{i+1}^n) - B(t_i^n)] \\
&= \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} \left[\frac{f(B(t_{i+1}^n)) - f(B(t_i^n))}{B(t_{i+1}^n) - B(t_i^n)} \right] [B(t_{i+1}^n) - B(t_i^n)]^2 \\
&\approx \sum_{i=0}^{n-1} f'(B(t_i^n))[B(t_{i+1}^n) - B(t_i^n)]^2 \\
&= \int_0^t f'(B(s)) ds.
\end{aligned}$$

The last equality follows from a previous theorem.

Example 4.6.8
Let $f(t)$ be an increasing differentiable function and let $X(t) = B(f(t))$. We compute $[X, X](t)$.

By taking limits over shrinking partitions,

$$\begin{aligned}
[X, X](t) &= \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} [B(f(t_{i+1}^n)) - B(f(t_i^n))]^2 \\
&= \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} [f(t_{i+1}^n) - f(t_i^n)] \left[\frac{B(f(t_{i+1}^n)) - B(f(t_i^n))}{\sqrt{f(t_{i+1}^n) - f(t_i^n)}} \right]^2 \\
&= \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} [f(t_{i+1}^n) - f(t_i^n)] Z_i^2 \\
&=: \lim_{\delta_n \to 0} T_n.
\end{aligned}$$

Here $Z_i \sim_{iid} \mathcal{N}(0, 1)$.

For any n,

$$\begin{aligned}
\mathbb{E}[\, T_n \,] &= \sum_{i=0}^{n-1} [f(t_{i+1}^n) - f(t_i^n)] \\
&= f(t) \\
\mathrm{Var}[T_n] &= \mathrm{Var}\left[\sum_{i=0}^{n-1} [f(t_{i+1}^n) - f(t_i^n)] Z_i^2 \right] \\
&= 3 \sum_{i=0}^{n-1} [f(t_{i+1}^n) - f(t_i^n)]^2.
\end{aligned}$$

Here we used the independence of the Z_i's and the fact that $\operatorname{Var} Z^2 = 3$. The last sum converges to the quadratic variation of f, which is zero as f is of finite variation and continuous. This implies that

$$T_n \xrightarrow{2} f(t)$$

as desired.

4.6.2 Itô's Formula for Functions of Two Variables

If two processes X, Y both posses a stochastic differential with respect to $B(t)$ and $f(x, y) \in C^2(\mathbb{R}^2, \mathbb{R})$, then we claim $f(X(t), Y(t))$ also posseses a stochastic differential.

Indeed, consider the Taylor expansion of order two,

$$\begin{aligned} df(x,y) &= \frac{\partial f(x,y)}{\partial x}dx + \frac{\partial f(x,y)}{\partial y}dy \\ &\quad + \frac{1}{2}\left(\frac{\partial^2 f(x,y)}{(\partial x)^2}(dx)^2 + \frac{\partial^2 f(x,y)}{(\partial y)^2}(dy)^2 + 2\frac{\partial^2 f(x,y)}{\partial x \partial y}dxdy\right). \end{aligned}$$

We can then guess at the formula by using substituting $dx \leftarrow dX(t)$ and $dy \leftarrow dY(t)$.

Theorem 4.6.9
Let $f \in C^2(\mathbb{R}^2, \mathbb{R})$ and X, Y be Itô processes. Then

$$\begin{aligned} &df(X(t), Y(t)) \\ &= \frac{\partial f}{\partial x}(X(t), Y(t))dX(t) + \frac{\partial f}{\partial y}(X(t), Y(t))dY(t) \\ &\quad + \frac{1}{2}\frac{\partial^2 f}{(\partial x)^2}\left(X(t), Y(t)\right)\sigma_X^2(t)dt + \frac{1}{2}\frac{\partial^2 f}{(\partial y)^2}\left(X(t), Y(t)\right)\sigma_Y^2(t)dt \\ &\quad + \frac{\partial^2 f}{\partial x \partial y}\left(X(t), Y(t)\right)\sigma_X(t)\sigma_Y(t)dt. \end{aligned}$$

Example 4.6.10
If $f(x, y) = xy$, applying the theorem above yields the integration by parts formula.

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t) + \sigma_X(t)\sigma_Y(t)dt.$$

An important case of Itô's formula is for functions of the form $f(X(t), t)$.

Theorem 4.6.11
Let $f(x,t)$ be twice continuously differentiable in x and continuous differentiable in t. Suppose $X(t)$ is an Itô process. Then

$$df(X(t),t) = \frac{\partial f}{\partial x}(X(t),t)dX(t) + \frac{\partial f}{\partial t}(X(t),t)dt + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X(t),t)\sigma_X^2(t)dt.$$

Example 4.6.12
We wish to find a stochastic differential of

$$X(t) = \exp\left(B(t) - \frac{t}{2}\right).$$

We apply Itô's formula with $f(B(t),t)$ where $f(x,t) = \exp(x - {}^t/_2)$.

$$\begin{aligned} dX(t) &= df(B(t),t) \\ &= \frac{\partial f}{\partial x}dB(t) + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial^2 x}dt \\ &= f(B(t),t)dB(t) - \frac{1}{2}f(B(t),t)dt + \frac{1}{2}f(B(t),t)dt \\ &= f(B(t),t)dB(t) \\ &= X(t)dB(t). \end{aligned}$$

4.7 Itô Processes in Higher Dimension

Let $\bar{B}(t) = (B_1(t),\ldots,B_d(t))$ be Brownian motion in $\mathbb{R}^d$, ie all coordinates $B_i(t)$ are independent one-dimensional Brownian motions. Let $\mathcal{F}_t$ be the σ-field generated by $\bar{B}(s), s \le t$. Define $\bar{H}(t)$ to be a *regular adapted* d-dimensional vector process, ie each coordinate is regular and adapted. If for each j, $\int_0^T H_j^2(t)dt \overset{a.s.}{<} \infty$, then the Itô integrals $\int_0^T H_j(t)dB_j(t)$ are defined. If we require this condition to hold over all $j \in [d]$, an equivalent condition is

$$\int_0^T \|H(t)\|^2 dt \overset{a.s.}{<} \infty.$$

It is customary to use a scalar product notation

$$\bar{H}(t)\cdot dB(t) := \sum_{j=1}^d H_j(t)dB_j(t)$$
$$\int_0^T \bar{H}(t)\cdot dB(t) := \sum_{j=1}^d \int_0^T H_j(t)dB_j(t).$$

If $b(t)$ is an integrable function, then the process given by the stochastic differential

$$dX(t) = b(t)dt + \sum_{j=1}^{d} H_j(t)dB_j(t)$$

is well-defined. Indeed, it is a scalar Itô process driven by a d-dimensional Brownian motion.

More generally, we can have any number n of processes driven by a d-dimensional Brownian motion,

$$dX_i(t) = b_i(t)dt + \sum_{j=1}^{d} \sigma_{ij}(t)dB_j(t)$$

for $i \in [n]$. Here σ is a matrix-valued function, $\bar{B}$ is a d-dimensional Brownian motion, and $\bar{x}, \bar{b}$ are n-dimensional vector-valued functions. $\bar{X}$ is referred to as an *Itô process*. In vector form, we write

$$d\bar{X}(t) = \bar{b}(t)dt + \sigma(t)d\bar{B}(t).$$

The dependence of $\bar{b}(t), \sigma(t)$ on time t can be via the whole path of the process as well as the path of $\bar{B}(s)$, both up to time t. The only restrictions are the following:

(i) For any $i \in [n]$, $b_i(t)$ is adapted and $\int_0^T |b_i(t)|dt \overset{a.s.}{<} \infty$.

(ii) For any $i \in [n], j \in [d]$, $\sigma_{ij}(t)$ is adapted and $\int_0^T \sigma_{ij}^2(t)dt \overset{a.s.}{<} \infty$.

Example 4.7.1 (Diffusion Process)
An important case of the dependence is of the form $b(t) = b(\bar{X}(t), t)$ and $\sigma(t) = \sigma(\bar{X}(t), t)$. In this case, the stochastic differential is written as

$$d\bar{X}(t) = \bar{b}(\bar{X}(t), t)dt + \sigma(\bar{X}(t), t)d\bar{B}(t)$$

and $\bar{X}(t)$ is then a *diffusion process*.

Unsurprisingly, Itô's formula extends to this setting. We need the quadratic variation of a multi-dimensional Itô process. First, we check that that the quadratic covariation of two independent Brownian motions is zero.

Theorem 4.7.2
Let $B_1(t), B_2(t)$ be independent Brownian motions. Their covariation process exists and is identically zero.

Proof
Let $\{ t_i^n \}$ be a partition of $[0, t]$ and consider

$$T_n := \sum_{i=0}^{n-1} [B_1(t_{i+1}^n) - B_1(t_i^n)][B_2(t_{i+1}^n) - B(t_i^n)].$$

By the independence of B_1, B_2, $\mathbb{E}[\, T_n \,] = 0$. Moreover, the variance of the sum is simply the sum of variances

$$\begin{aligned}
\operatorname{Var} T_n &= \sum_{i=0}^{n-1} \mathbb{E}\left[\left(B_1(t_{i+1}^n) - B_1(t_i^n) \right)^2 \right] \cdot \mathbb{E}\left[\, (B_2(t_{i+1}) - B_2(t_i^n))^2 \, \right] \\
&= \sum_{i=0}^{n-1} (t_{i+1}^n - t_i^n)^2 \\
&\leq \max_i (t_{i+1}^n - t_i^n) t \\
&\to 0. && \delta_n \to 0
\end{aligned}$$

Thus $T_n \xrightarrow{2} 0$ as $\delta_n \to 0$. $\square$

It follows from the definition

$$dX_i(t) = b_i(t)dt + \sum_{j=1}^{d} \sigma_{ij}(t) dB_j(t)$$

that

$$d[X_i, X_j](t) = dX_i(t) dX_j(t) = a_{ij} dt$$

for $i, j \in [n]$. Here $a = \sigma\sigma^T$ is the *diffusion matrix.*

4.7.1 Itô's Formula for Functions of Several Variables

If $\bar{X}(t)$ is a vector Itô process and $f \in C^2(\mathbb{R}^n, \mathbb{R})$, then $f(\bar{X}(t))$ is also an Itô process. Moreover, its stochastic differential can be shown to be

$$df(\bar{X}(t)) = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\bar{X}(t)) dX_i(t) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} f(\bar{X}(t)) d[X_i, X_j](t).$$

When there is only one Brownian motion, ie $B_i = B_j$, this formula is a generalization of Itô's formula for a function of two variables.

Remark 4.7.3 (Integration by Parts) Let $X(t), Y(t)$ be two Itô processes that are adapted to independent Brownian motions B_1, B_2. Take $f(x, y) = xy$ and note that only one of the second partial derivatives, $\frac{\partial^2 xy}{\partial x \partial y}$, can be non-zero, but the corresponding differential term $d[B_1, B_2](t) = 0$ by independence. Thus the covariation of $X(t), Y(t)$ is zero and we obtain the equality

$$d(X(t)Y(t)) = X(t)dY(t) + Y(t)dX(t).$$

This is the usual integration by parts formula.

Chapter 5

Stochastic Differential Equations

Differential equations are used to describe the evolution of a system. Stochastic differential equations (SDEs) arise when random noise is introduced into ODEs. We introduce two concepts of solutions of SDEs: the strong solution and weak solution, but focus on the more basic strong solution. We also give a connection between SDEs and random ODEs, solutions to linear linear SDEs, stochastic exponential and logarithm, methods of solutions to some SDEs, and results on existence and uniqueness of solutions.

5.1 Definitions

5.1.1 Ordinary Differential Equations (ODEs)

If $x : \mathbb{R}_+ \to \mathbb{R}$ is a differentiable function of $t \geq 0$, $\mu(x,t) : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is a function of x, t, and the following relation is satisfied for all $0 \leq t \leq T$,

$$\frac{dx(t)}{dt} = x'(t) = \mu(x(t), t)$$

subject to initial conditions $x(0) = x_0$, then $x(t)$ is said to be a solution of the ODE μ with initial condition x_0.

Typically, we also require that $x'(t)$ is continuous so we can interpret the differential equation as an integral equation as well. The above equation can be written in other forms:

$$dx(t) = \mu(x(t), t)dt$$

$$x(t) = x(0) + \int_0^t \mu(x(s), s)ds. \qquad \text{assuming } x \in C^1$$

Before we rigoriously define SDEs, we first show how they arise as randomly perturbed ODEs and give a physical interpretation.

5.1.2 White Noise & SDEs

Intuively, we imagine a white noise process $\xi(t)$ as the "derivative" of Brownian motion

$$\xi(t) = \frac{dB(t)}{dt} = B'(t).$$

Of course, such a notion does not exist rigorously as a function of t in the standard calculus sense, as Browian motion is nowhere differentiable a.s.

If $\sigma(x,t)$ is the "intensity" of the noise at a point x and time t, it is agreed that

$$\int_0^T \sigma(X(t),t)\xi(t)dt := \int_0^T \sigma(X(t),t)dB(t),$$

where the integral is the Itô integral.

SDEs can be obtained when the coefficients of ODEs are perturbed by white noise.

> **Example 5.1.1 (Black-Scholes-Mertons)**
> The Black-Scholes-Mertons model is designed for growth with uncertain rate of return. Suppose $x(t)$ is the value of \$1 after time t, invested in a savings account, and r is the interest rate. According to the definition of compound interest, $x(t)$ satisfies the ODE
>
> $$x'(t) = rx(t).$$
>
> If there is uncertainty in the interest rate, we can model the uncertainty as perturbations by noise, $r + \sigma\xi(t)$, and the following SDE is obtained
>
> $$\begin{aligned}\frac{dX(t)}{dt} &= (r + \sigma\xi(t))X(t) \\ dX(t) &= rX(t)dt + \sigma X(t)dB(t).\end{aligned}$$
>
> The first equation is the substitution into the ODE and the second equation is the interpretation of the equation.
>
> We have already seen solutions to the SDE above, which is given by a geometric Brownian motion
>
> $$X(t) = \exp\left([r - \sigma^2/2]t + \sigma B(t)\right).$$
>
> This solution can be verified by Itô's formula.

> **Example 5.1.2 (Population Growth)**
> If $x(t)$ denotes the population density, then it can be described by the ODE
>
> $$\frac{dx(t)}{dt} = ax(t)(1 - x(t)).$$

The growth is exponential with birth rate a when this density is small, and slows down when the density increases. Random perturbation of the birth rate results in the SDE

$$\frac{dX(t)}{dt} = [a + \sigma\xi(t)]X(t)[1 - X(t)]$$
$$dX(t) = aX(t)[1 - X(t)]dt + \sigma X(t)[1 - X(t)]dB(t).$$

5.1.3 A Physical Model of Diffusion & SDEs

Brownian motion was inspired by the movement of particles suspended in a fluid. Molecules of the fluid move with various velocities and collide with the particle from every possible direction. As a result of these collisions, the particle exhibits erratic movements. This movement intensifies with an increase in the temperature of the fluid.

Heuristically, let $X(t)$ denote the displacement of the particle in one dimension from its initial position at time t. If $\sigma(x,t)$ measures the effect of temperature at point x and time t, the displacement due to colliding molecules in a small time interval $[t, t+\Delta]$ is modelled as $\sigma(x,t)[B(t+\Delta) - (t)]$. If the velocity of the fluid at point x and time t is $\mu(x,t)$, then the displacement of the particle due to the movement of the fluid during $[t, t+\Delta]$ is $\mu(x,t)\Delta$. Thus the total displacement from its position x at time t is given by

$$X(t+\Delta) - x \approx \mu(x,t)\Delta + \sigma(x,t)[B(t+\Delta) - B(t)].$$

From this approximation, we can guess that the mean displacement from x during a short time Δ is given by

$$\mathbb{E}\left[\, X(t+\Delta) - X(t) \mid X(t) = x \,\right] \approx \mu(x,t)\Delta.$$

The second moment of the displacement from x during time Δ is given by

$$\mathbb{E}\left[\, (X(t+\Delta) - X(t))^2 \,\middle|\, X(t) = x \,\right] \approx \sigma^2(x,t)\Delta.$$

The above relations show that for small intervals of time, bot the mean and second moment of the displacement of a diffusing particle at position x and time t are proportional to the length of the interval, with coefficients $\mu(x,t)$ and $\sigma^2(x,t)$, respectively.

It can be shown that by taking $\Delta \to 0$, these two requirements characterize *diffusion processes*. Indeed, assuming $\mu(x,t)$ and $\sigma(x,t)$ are smooth functions, the heuristic displacement equation above also indicates that for small intervals of time Δ, diffusions are approximately Gaussian processes. Given $X(t) = x$. That is, $X(t+\Delta) - X(t)$ is approximately Normally distributed, $\mathcal{N}(\mu(x,t)\Delta, \sigma^2(x,t)\Delta)$. Note this is only reasonable for "small" intervals. A stochastic differential equation is obtained heuristically from the heuristic equation above by replacing Δ by dt, $\Delta B = B(t+\Delta) - B(t)$ by $dB(t)$, and $X(t+\Delta) - X(t)$ by $dX(t)$.

$$dX(t) = \mu(x,t)dt + \sigma(x,t)dB(t).$$

5.1.4 Stochastic Differential Equations

Definition 5.1.3 ((Diffusion) SDE)
Let $B(t)$ be a Brownian motion. An equation of the form

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dB(t)$$

where $\mu(x, t), \sigma(x, t)$ are given and $X(t)$ is the unknown process, is called a *(diffusion-type) SDE driven by Brownian motion.* The functions $\mu(x, t), \sigma(x, t)$ are known as the *drift* and *diffusion* coefficients, respectively.

Definition 5.1.4 (Strong Solution)
A process $X(t)$ is called a *strong solution* of an SDE if for all $t > 0$, the integrals $\int_0^t \mu(X(s), s)ds$ and $\int_0^t \sigma(X(s), s)dB(s)$ exist and

$$X(t) = X(0) + \int_0^t \mu(X(s), s)ds + \int_0^t \sigma(X(s), s)dB(s).$$

Note that a strong solution is a function (functional) $F(\{ B(s) : s \leq t \}, t)$ of the given Brownian motion. More general SDEs have the form

$$dX(t) = \mu(\{ X(s) : s \leq t \}, t)dt + \sigma(\{ X(s), s \leq t \}, t)dB(t).$$

The only restriction on μ, σ is that they must be adapted processes, with respective integrals defined. We will focus on diffusion-type SDEs.

Example 5.1.5
As a review, consider the SDE

$$dX(t) = X(t)dB(t).$$

By Itô's formula,

$$\begin{aligned} d\ln X(t) &= \frac{1}{X(t)}dX(t) - \frac{1}{2X^2(t)}\sigma_X^2(t)dt \\ &= \frac{1}{X(t)}[X(t)dB(t)] - \frac{1}{2X^2(t)}X^2(t)dt \\ &= dB(t) - \frac{1}{2}dt. \end{aligned}$$

Hence a solution to the SDE above is given by

$$\begin{aligned}\ln X(t) &= \ln X(0) + B(t) - \frac{1}{2}t \\ X(t) &= X(0)\exp\left[B(t) - \frac{1}{2}t\right].\end{aligned}$$

Example 5.1.6
Let $a(t)$ be a deterministic and differentiable function. Consider the SDE

$$dX(t) = a(t)dB(t).$$

Clearly $X(t) = X(0) + \int_0^t a(s)dB(s)$.

We can explicitly represent this as a function of the Brownian motion through integration by parts.

$$\begin{aligned}d(a(t)B(t)) &= a(t)dB(t) + B(t)da(t) \\ &= a(t)dB(t) + B(t)a'(t)dt \\ &= dX(t) + B(t)a'(t)dt \\ X(t) &= X(0) + a(t)B(t) - \int_0^t B(s)a'(s)ds.\end{aligned}$$

Example 5.1.7
Consider the SDE

$$dX(t) = \mu X(t)dt + \sigma X(t)dB(t), X(0) = 1.$$

Again, by applying Itô's formula to $\ln X(t)$, we can show that one solution is given by

$$X(t) = X(0)\exp\left[(\mu - {}^{\sigma^2}\!/_2)t + \sigma B(t)\right].$$

Example 5.1.8 (Langevin Equation & Ornstein-Uhlenbeck Process)
Consider the SDE

$$dX(t) = -\alpha X(t)dt + \sigma dB(t)$$

for some $\alpha, \sigma > 0$. Note we cannot apply the previous example since the drift coefficient is not constant.

Consider the process $Y(t) = X(t)\exp(\alpha t)$. Note that the covariation of $\exp(\alpha t)$ with $X(t)$ is zero since it is differentiable. By the product rule,

$$\begin{aligned}dY(t) &= \exp(\alpha t)dX(t) + \alpha\exp(\alpha t)X(t)dt \\ &= \exp(\alpha t)(-\alpha X(t)dt + \sigma dB(t)) + \alpha\exp(\alpha t)X(t)dt \\ &= \sigma\exp(\alpha t)dB(t).\end{aligned}$$

This yields the solutions

$$Y(t) = Y(0) + \int_0^t \sigma \exp(\alpha s) dB(t)$$
$$X(t) = \exp(-\alpha t) \left[X(0) + \int_0^t \sigma \exp(\alpha s) dB(s) \right].$$

The process $X(t)$ is known as the *Ornstein-Uhlenbeck (OU)* process.

We can find the explicit functional dependence of the solution on the Brownian motion path by performing integration by parts on $Y(t)$.

$$X(t) = \exp(-\alpha t)X(0) + \sigma B(t) - \sigma\alpha \int_0^t \exp(-\alpha(t-s))B(s)ds.$$

More generally, consider the SDE

$$dX(t) = [\beta - \alpha X(t)]dt + \sigma dB(t),$$

with the solution

$$X(t) = \frac{\beta}{\alpha} + \exp(-\alpha t) \left(X(0) - \frac{\beta}{\alpha} + \int_0^t \sigma \exp(\alpha s) dB(s) \right).$$

This can be verified with Itô's formula.

Example 5.1.9
Consider the SDE

$$dX(t) = B(t)dB(t).$$

We have $X(t) = X(0) + \int_0^t B(s)dB(s)$ which can be explicitly computed by Itô's formula

$$X(t) = X(0) + \frac{1}{2}[B^2(t) - t].$$

5.1.5 SDEs & Random ODEs

Consider an SDE with a unit diffusion coefficient

$$X(t) = X(0) + \int_0^t \mu(X(s), s)ds + B(t).$$

Let $Y(t) := X(t) - B(t)$. Then $Y(t)$ satisfies the following equation

$$Y(t) = Y(0) + \int_0^t \mu(Y(s) + B(s), s)ds.$$

This is simply an ODE, albeit involving Brownian motion!

$$\frac{dY(t)}{dt} = \mu(Y(t) + B(t), t).$$

Thus if $Y(t)$ solves the above random ODE, then $X(t) = Y(t) + B(t)$ solves the SDE. Some SDEs with more general diffusion coefficients $\sigma(x)$ can be transformed into the equation with unit diffusion coefficient and the above method can then be applied. This transformation is related to the so-called *Doss-Sussman method*.

Remark 5.1.10 Only some classes of SDEs (ie linear SDEs) admit a closed form solution. When a closed form solution is difficult to find, existence and uniqueness results are important as otherwise, it si unclear what exactly the equation means. When a solution exists and is unique, then numerical methods can be used to approximate it.

5.2 Stochastic Exponential & Logarithm

5.2.1 Stochastic Exponential

Definition 5.2.1 (Stochastic Exponential)
Let X have a stochastic differential and U satisfy

$$dU(t) = U(t)dX(t) \qquad\qquad U(0) = 1$$
$$U(t) = 1 + \int_0^t U(s)dX(s).$$

Then U is the *stochastic exponential of* X and is denoted by $\mathcal{E}(X)$.

If X is of finite variation, we will see that $\mathcal{E}(X) = \exp(X(t) - X(0))$.

Theorem 5.2.2
If X has a stochastic differential, then its stochastic exponential is uniquely given by

$$\mathcal{E}(X)(t) := \exp\left(X(t) - X(0) - \frac{1}{2}[X, X](t)\right).$$

Proof
The proof of existence consists of applying Itô's formula to verify. Write $U(t) = \exp(V(t))$, where

$$V(t) = X(t) - X(0) - \frac{1}{2}[X, X](t).$$

By Itô's formula,

$$dU(t) = d(\exp(V(t))) = \exp(V(t))dV(t) + \frac{1}{2}\exp(V(t))d[V,V](t).$$

Now, the quadratic covariation is invariant under scalar multiplication. Moreover,

$$[X,[X,X]](t) = 0$$

since $[X,X](t)$ is of finite variation (increasing) and $X(t)$ is continuous. It follows that

$$[V,V](t) = [X,X](t).$$

By substituting the definition of V along with the computation for $[V,V]$ into the expression above, we arrive at the following conclusion.

$$\begin{aligned} dU(t) &= \exp(V(t))\left(dX(t) - \frac{1}{2}d[X,X](t)\right) + \frac{1}{2}\exp(V(t))d[X,X](t) \\ &= U(t)dX(t). \end{aligned}$$

Hence $\mathcal{E}(X) = U$ by definition.

To see uniqueness, consider another process U_1 satisfying the definition of the stochastic exponential. We claim that

$$d\frac{U_1(t)}{U(t)} = 0.$$

This would conclude the proof since Itô integrals are unique up to a set of measure zero. Indeed, recall the stochastic quotient rule states that

$$\begin{aligned} &d(X(t)/Y(t)) \\ &= \frac{X(t)}{Y(t)}\left(\frac{1}{X(t)}dX(t) - \frac{1}{Y(t)}dY(t) - \frac{1}{X(t)Y(t)}d[X,Y](t) + \frac{1}{Y^2(t)}d[Y,Y](t)\right). \end{aligned}$$

Plugging in $X = U_1, Y = U_2$ and using the assumptions that

$$\begin{aligned} dU(t) &= U(t)dX(t) \\ dU_1(t) &= U_1(t)dX(t), \end{aligned}$$

we have that

$$\begin{aligned} &d(U_1(t)/U(t)) \\ &= \frac{U_1(t)}{U(t)}\left(\frac{1}{U_1(t)}U_1(t)dX(t) - \frac{1}{U(t)}U(t)dX(t)\right. \\ &\qquad \left. - \frac{1}{U_1(t)U(t)}U_1(t)U(t)d[X,X](t) + \frac{1}{U^2(t)}U^2(t)d[X,X](t)\right) \\ &= \frac{U_1(t)}{U(t)}\left(dX(t) - dX(t) - d[X,X](t) + d[X,X](t)\right) \\ &= 0. \end{aligned}$$

$\square$

Thus if U satisfies

$$dU(t) = U(t)dX(t)$$

for some arbitrary $U(0)$, then the solution is given by

$$U(t) = U(0)\mathcal{E}(X)(t).$$

Remark 5.2.3 Unlike the usual exponential $\exp(f)(t) := \exp(f(t))$, the stochastic exponential $\mathcal{E}(X)$ requires the knowledge of the all the values of the process upt to time t, as it involves the quadratic variation term $[X, X](t)$.

> **Example 5.2.4**
> The stochastic exponential of Brownian motion $B(t)$ is given by
>
> $$\mathcal{E}(B)(t) = \exp\left(B(t) - \frac{1}{2}t\right).$$

> **Example 5.2.5 (Stock Process & Return Process)**
> Let $S(t)$ denote the price of a stock and assume that it is an Itô process. The process of its return $R(t)$ is defined by the relation
>
> $$\begin{aligned} dR(t) &= \frac{dS(t)}{S(t)} \\ dS(t) &= S(t)dR(t). \end{aligned}$$
>
> Thus the stock price is the stochastic exponential of the return. Returns are typically easier to mdoel from first principles. For instance, the Black-Scholes model assumes that returns over non-overlapping time intervals are independent and have finite variance. This assumption leads to the model for the retur process
>
> $$R(t) = \mu t + \sigma B(t).$$
>
> The stock price is then given by
>
> $$\begin{aligned} \mathcal{E}(R)(t) &= S(0)\exp\left(R(t) - R(0) - \frac{1}{2}[R, R](t)\right) \\ &= S(0)\exp\left((\mu - {}^{\sigma^2}\!/_2)t + \sigma B(t)\right). \end{aligned}$$

5.2.2 Stochastic Logarithm

If $U = \mathcal{E}(X)$, the process X is called the *stochastic logarithm of* U, denoted $\mathcal{L}(U)$. For example, $B(t)$ is the stochastic logarithm of $\exp(B(t) - {}^{t}\!/_2)$.

Theorem 5.2.6
Let U have a stochastic differential be a.s. non-zero. Then the stochastic logarithm $X = \mathcal{L}(U)$ of U satisfies the SDE

$$dX(t) = \frac{1}{U(t)} dU(t), X(0) = 0.$$

Moreover,

$$\mathcal{L}(U)(t) = \ln\left(\frac{U(t)}{U(0)}\right) + \int_0^t \frac{1}{2U^2(s)} d[U,U](s).$$

Proof
By definition, $X = \mathcal{L}(U)$ means that $U = \mathcal{E}(X)$ and

$$\begin{aligned} dU(t) &= U(t)dX(t) & \qquad U(0) &= 1 \\ dX(t) &= \frac{1}{U(t)} dU(t) & \qquad X(0) &= 0 \end{aligned}$$

The initial condition $X(0) = 0$ enforces that $U(0) = 1$.

Consider the process

$$X(t) = \ln U(t) - \ln U(0) + \frac{1}{2}\int_0^t \frac{1}{U^2(s)} d[U,U](s).$$

By Itô's formula, it has a stochastic differential

$$\begin{aligned} dX(t) &= \frac{1}{U(t)} dU(t) - \frac{1}{2U^2(t)} d[U,U](t) + \frac{1}{2U^2(t)} d[U,U](t) \\ &= \frac{1}{U(t)} dU(t). \end{aligned}$$

□

Example 5.2.7
Let $U(t) := \exp(B(t))$. By computation,

$$\begin{aligned} d[U,U](t) &= \exp(2B(t)) d[B,B](t) \\ &= \exp(2B(t)) dt. \end{aligned}$$

We can apply our formula to conclude that

$$\begin{aligned}\mathcal{L}(U)(t) &= \ln U(t) - \ln(1) + \int_0^t \frac{\exp(2B(s))}{2\exp(2B(s))} ds \\ &= B(t) + \int_0^t \frac{1}{2} ds \\ &= B(t) + \frac{1}{2}t.\end{aligned}$$

5.3 Solutions to Linear SDEs

Linear SDEs form a class of SDEs that can be solved explicitly. Consider a general linear SDE in one dimension.

$$dX(t) = [\alpha(t) + \beta(t)X(t)]dt + [\gamma(t) + \delta(t)X(t)]dB(t).$$

Here $\alpha, \beta.\gamma, \delta$ are given adapted processes, and are continuous functions of t.

5.3.1 Stochastic Exponential SDEs

Consider the simpler case where $\alpha, \gamma \equiv 0$. The SDE can be expressed in the form

$$dU(t) = \beta(t)U(t)dt + \delta(t)U(t)dB(t)$$

$$\begin{aligned}dU(t) &= U(t)dY(t) \\ dY(t) &= \beta(t)dt + \gamma(t)dB(t).\end{aligned}$$

This is in the form of the stochastic exponential SDE and so

$$\begin{aligned}&U(t) \\ &= U(0)\mathcal{E}(Y)(t) \\ &= U(0)\exp\left(Y(t) - Y(0) - \frac{1}{2}[Y,Y](t)\right) \\ &= U(0)\exp\left(\int_0^t \beta(s)ds + \int_0^t \delta(s)dB(s) - \frac{1}{2}\int_0^t \delta^2(s)ds\right) \\ &= U(0)\exp\left(\int_0^t \left[\beta(s) - \frac{1}{2}\delta^2(s)\right] ds + \int_0^t \delta(s)dB(s)\right).\end{aligned}$$

5.3.2 General Linear SDEs

Intuitively, our first attemp at finding a solution for general linear SDEs should involve tweaking stochastic exponential SDEs to include the extra terms α, γ.

We look for a solution of the form

$$\begin{aligned} X(t) &= U(t)V(t) \\ dX(t) &= U(t)dV(t) + V(t)dU(t) + d[U,V](t), \end{aligned}$$

where

$$\begin{aligned} dU(t) &= \beta(t)U(t)dt + \delta(t)U(t)dB(t) \\ dV(t) &= a(t)dt + b(t)dB(t). \end{aligned}$$

Substituting this into the differential of the product yields

$$\begin{aligned} dX(t) &= U(t)\left(a(t)dt + b(t)dB(t)\right) \\ &\quad + V(t)\left(\beta(t)U(t)dt + \delta(t)U(t)dB(t)\right) \\ &\quad + \delta(t)U(t)b(t)dt \\ &= a(t)U(t)dt + b(t)U(t)dB(t) \\ &\quad + \beta(t)X(t)dt + \delta(t)X(t)dB(t) \\ &\quad + \delta(t)U(t)b(t)dt \end{aligned}$$

It is then clear that we can choose coefficients a, b such that the relation $X(t) = U(t)V(t)$ holds. Indeed, the desired coefficients must satisfy

$$\begin{aligned} b(t)U(t) &= \gamma(t) \\ a(t)U(t) &= \alpha(t) - \delta(t)\gamma(t). \end{aligned}$$

Without loss of generality, set $U(0) = 1, V(0) = X(0)$. We have already solved U and thus a, b are determined. We have

$$\begin{aligned} X(t) &= U(t)V(t) \\ &= U(t)\left(X(0) + \int_0^t \frac{\alpha(s) - \delta(s)\gamma(s)}{U(s)} ds + \int_0^t \frac{\gamma(s)}{U(s)} dB(s)\right). \end{aligned}$$

5.3.3 Langevin-Type SDE

Let $X(t)$ satisfy

$$dX(t) = a(t)X(t)dt + dB(t),$$

where $a(t)$ is a given adapted and continuous process. When $a(t) = -\alpha$, the equation reduces to the Langevin equation. We can solve this problem using general solution to linear SDEs.

The coefficients are $\beta(t) = a(t)$, $\gamma(t) = 1$, and $\alpha(t), \delta(t) = 0$. We first solve the stochastic exponential SDE

$$dU(t) = a(t)U(t)dt,$$

which yields the solution

$$U(t) = \exp\left(\int_0^t a(s)ds\right).$$

The solution to $X(t)$ is then given by

$$\begin{aligned} X(t) &= U(t)\left(X(0) + \int_0^t \frac{1}{U(s)}dB(s)\right) \\ &= \exp\left(\int_0^t a(s)ds\right)\left(X(0) + \int_0^t \exp\left(-\int_0^s a(u)du\right)dB(s)\right). \end{aligned}$$

Note that we can also directly solve this SDE through integration by parts similar to how we solved the Langevin equation.

5.3.4 Brownian Bridge

The *(pinned) Brownian Bridge* is a solution to the following SDE

$$dX(t) = \frac{b - X(t)}{T - t}dt + dB(t)$$

for $t \in [0, T]$ and $X(0) = a$. We remark that it is a lienar SDE with coefficients

$$\begin{aligned} \alpha(t) &= \frac{b}{T-t} \\ \beta(t) &= -\frac{1}{T-t} \\ \gamma(t) &= 1 \\ \delta(t) &= 0. \end{aligned}$$

Again, we first solve the stochastic exponential SDE

$$dU(t) = -\frac{1}{T-t}U(t)dt.$$

This is solved by a deterministic process

$$\begin{aligned}
U(t) &= \exp\left(\int_0^t -\frac{1}{T-s}ds\right) \\
&= \exp\left(\int_T^{T-t} \frac{1}{u}du\right) && u = T-s, du = -ds \\
&= \exp\left(\ln(T-t) - \ln(T)\right) \\
&= \frac{T-t}{T}.
\end{aligned}$$

We can then plug this into our general formula for $X(t)$.

$$\begin{aligned}
X(t) &= \left(\frac{T-t}{T}\right)\left(a + \int_0^t T\frac{b/T-s}{T-s}ds + \int_0^t \frac{T}{T-s}dB(s)\right) \\
&= \left(\frac{T-t}{T}\right)\left(a + bT\int_0^t \frac{1}{(T-s)^2}ds + T\int_0^t \frac{1}{T-s}dB(s)\right) \\
&= \left(\frac{T-t}{T}\right)\left(a + b\frac{t}{T-t} + T\int_0^t \frac{1}{T-s}dB(s)\right) \\
&= a\left(1-\frac{t}{T}\right) + b\frac{t}{T} + (T-t)\int_0^t \frac{1}{T-s}dB(s)
\end{aligned}$$

for $t \in [0,T]$. From the expression we derived above, we see that the Brownian bridge is a warped Brownian motion with fixed values at each end of the interval $[0,T]$ with $X(0) = a$ and $X(T) = b$.

Remark that for any $t < T$,

$$\int_0^t \frac{1}{(T-s)^2}ds < \infty.$$

Thus the process $\int_0^t 1/(T-s) dB(s)$ is a martingale. Moreover, the function under the Itô integral is deterministic. This means that $X(t)$ is a Gaussian process on $[0,T]$ with initial value $X(0) = a$. The final value $X(T) = b$ is determined by continuity, which we will see below. All in all, from our work before on Itô integral processes, a Brownian bridge is a continuous Gaussian processon $[0,T]$ with mean function $a(1 - t/T) + bt/T$, and covariance function

$$\begin{aligned}
\mathrm{Cov}\left[\, X(t), X(t+u) \,\right] &= \int_0^t \frac{T-t}{T-s}\cdot\frac{T-t-u}{T-s}ds \\
&= [T-t][T-(t+u)]\int_0^t \frac{1}{T-s}ds \\
&= [T-t][T-(t+u)]\cdot\frac{t}{T(T-t)} \\
&= t - \frac{t(t+u)}{T}.
\end{aligned}$$

Lemma 5.3.1
For any continuous function $g(s)$,

$$\lim_{t \uparrow T} (T-t) \int_0^t \frac{g(s)}{(T-s)^2} ds = g(T).$$

Proof
By u-substitution,

$$\begin{aligned} &\frac{1}{L} \int_0^{T - 1/L} \frac{g(s)}{(T-s)^2} ds \\ &\quad = \frac{1}{L} \int_{1/T}^{L} g(T - 1/u) du. \qquad\qquad u = \frac{1}{T-s}, du = \frac{1}{(T-s)^2} ds \end{aligned}$$

From elementary calculus,

$$g(T) = \frac{1}{L} \int_0^L g(T) ds.$$

We can thus write

$$\left| g(T) - \frac{1}{L} \int_{1/T}^{L} g(T - 1/u) du \right| \le \left| \frac{1}{LT} \cdot g(T) \right| + \left| \frac{1}{L} \int_{1/T}^{L} g(T) - g(T - 1/u) du \right|.$$

Fix $\varepsilon > 0$. For some $\delta > 0$, $|t - T| \le \delta$ implies that $|g(t) - g(T)| < \varepsilon$. Since continuous functions are bounded on compact (closed and bounded) sets, g is M-uniformly bounded on $[0, T]$ for some $M > 0$. It follows that

$$\begin{aligned} &\left| g(T) - \frac{1}{L} \int_{1/T}^{L} g(T - 1/u) du \right| \\ &\le \left| \frac{1}{LT} \cdot g(T) \right| + \left| \frac{1}{L} \int_{1/T}^{1/\delta} g(T) - g(T - 1/u) du \right| + \left| \frac{1}{L} \int_{1/\delta}^{L} g(T) - g(T - 1/u) du \right| \\ &\le \left| \frac{1}{LT} \cdot g(T) \right| + \frac{1}{L} \cdot 2M \left(\frac{1}{\delta} - \frac{1}{T} \right) + \frac{L - 1/\delta}{L} \cdot \varepsilon. \end{aligned}$$

Choosing L sufficiently large ensures the estimate above is at most a constant fraction of ε, concluding the proof. □

Proposition 5.3.2
The following holds.

$$\lim_{t \uparrow T} (T-t) \int_0^t \frac{1}{T-s} dB(s) \overset{a.s.}{=} 0.$$

Proof
We apply integration by parts, which is the same as the standard formula since the covariation between a deterministic term and Brownian motion is zero. For any $t < T$,

$$\int_0^t \frac{1}{T-s} dB(s) = \frac{1}{T-t} B(t) - \int_0^t \frac{1}{(T-s)^2} B(s) ds$$
$$(T-t) \int_0^t \frac{1}{T-s} dB(s) = B(t) - (T-t) \int_0^t \frac{1}{(T-s)^2} B(s) ds.$$

By the lemma above, the expression above tends to 0 as $t \uparrow T$. $\square$

5.4 Existence & Uniqueness of Strong Solutions

In this section, we consider the general diffusion-type SDE

$$dX(t) = \mu(X(t), t) + \sigma(X(t), t) dB(t).$$

We begin with some sufficient conditions to guarantee the existence and uniquess of strong solutions on $[0, T]$.

Theorem 5.4.1 (Existence & Uniqueness)
Suppose the following conditions hold.

(i) (Locally Lipschitz in x, uniformly over t) For every $T, L > 0$, there is a constant $K^*_{T,L} > 0$ depending only on T, L such that for all $x, y \in [-L, L]$ and $t \in [0, T]$,

$$|\mu(x,t) - \mu(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leq K^*_{T,L} |x - y|.$$

(ii) (Linear growth condition) For some $K_T > 0$ depending on T, the coefficients satisfy

$$|\mu(x,t)| + |\sigma(x,t)| \leq K_T(1 + |x|)$$

for all $t \in [0, T]$.

(iii) $X(0)$ is independent of $\{ B(t) : t \in [0, T] \}$.

(iv) $\mathbb{E}[X^2(0)] < \infty$.

Then there is a unique strong solution $X(t)$ of the SDE with continuous paths on $[0, T]$. Moreover, there is a constant $C_{T,K_T} > 0$ depending only on T, K_T such that

$$\mathbb{E}\left[\sup_{t \in [0,T]} X^2(t) \right] \leq C_{T,K_T} \left(1 + \mathbb{E}\left[X^2(0) \right]\right).$$

The weaker version of this theorem where the constant $K_{T,L} \leftarrow K_T$ depends only T can be proven similarly to Picard's theorem for the existence and uniqueness for solutions to ODEs, ie through Picard iterations

$$\xi_{m+1}(t) := X(0) + \int_0^t \mu(\xi_m(s), s)ds + \int_0^t \sigma(\xi_m(s), s)dB(s).$$

Remark 5.4.2 The local Lipschitz condition holds if we assume partial derivatives $\frac{\partial \mu}{\partial x}(x,t)$ and $\frac{\partial \sigma}{\partial x}(x,t)$ are bounded for all $x \in [-L, L]$ and $t \in [0,T]$. This is in turn true if the derivatives are continuous.

5.4.1 Less Stringent Conditions for Uniquness of Strong Solutions

The next result is specific for uniqueness of solutions for one-dimensional SDEs. We state if for the case of time-independent coefficients. A similar result holds for time-dependent coefficients.

Theorem 5.4.3 (Yamada-Watanabe)
Suppose $\mu(x)$ is Lipschitz and $\sigma(x)$ is Hölder continuous of order $\alpha \geq 1/2$, ie

$$|\sigma(x) - \sigma(y)| \leq K|x-y|^\alpha.$$

Then if a strong solution exists, it is unique.

Example 5.4.4 (Girsanov's SDE)
Consider the SDE

$$dX(t) = |X(t)|^r dB(t)$$

where $X(0) = 0$ and $r \in [1/2, 1]$. Note that $|x|^r$ is Hölder continuous but not Lipschitz.

By inspection, $X(t) \equiv 0$ is a strong solution. The previous theorem tells us that this is in fact the only solution.

5.5 Markov Property of Solutions

Recall that a process is Markov if the following property holds: If $\mathcal{F}_t$ is the σ-field generated by the process up to time t, then for any $0 \leq s < t$,

$$\mathbb{P}\{\, X(t) \leq y \mid \mathcal{F}_s \,\} \overset{a.s.}{=} \mathbb{P}\{\, X(t) \leq y \mid X(s) \,\}.$$

Suppose $X(t)$ is a solution to some SDE. Intuitively, for small Δ, given $X(t) = x$, $X(t+\Delta)$ depends on $B(t+\Delta) - B(t)$ which is independent of the past. We state but not prove this result.

Theorem 5.5.1
Let $X(t)$ be a solution to a diffusion-type SDE. Then $X(t)$ has the Markov property.

5.5.1 Transition Function

Markov processes are characterized by their transition probability function, denoted by

$$P(A, t, x, s) := \mathbb{P}\{ X(t) \in A \mid X(s) = x \}.$$

We may also write $P(y, t, x, s)$ in the case that $A = (-\infty, y]$. Note that P should be measurable function of x and for any fixed t, x, s, it should be a probability measure in A.

By the law of total probability, by conditioning on all possible values z of the process at time u for $u \in (s, t)$, we obtain that the transition probability function satisfies the *Chapman-Kolmogorov equation*

$$P(A, t, x, s) = \int_{-\infty}^{\infty} P(A, t, z, u) P(dz, u, x, s)$$

for any $u \in (s, t)$.

In the case that a *transition density* $p(y, t, x, s)$ exists, ie

$$P(A, t, x, s) = \int_A p(z, t, x, s) dz,$$

then the Chapman-Kolmogorov equation can be interpreted as

$$p(y, t, x, s) = \int_{-\infty}^{\infty} p(y, t, z, u) p(z, u, x, s) dz$$

for every $u \in (s, t)$.

Example 5.5.2
Suppose we have a transition density

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

which is the density function of $\mathcal{N}(x, t-s)$. Then the corresponding diffusion process is Brownian motion. Indeed, Brownian motion is Markov and the conditional distribution of $B(t) \mid B(s) = x$ is precisely $\mathcal{N}(x, t-s)$.

Example 5.5.3
Let $X(t)$ solve the SDE
$$dX(t) = \mu X(t)dt + \sigma X(t)dB(t)$$
for some $\mu, \sigma \in \mathbb{R}$. We have seen that

$$\begin{aligned} X(t) &= X(0)\exp\left([\mu - \sigma^2/2]t + \sigma B(t)\right) \\ X(t) &= X(s)\exp\left([\mu - \sigma^2/2][t-s] + \sigma[B(t) - B(s)]\right). \end{aligned}$$

Using the independence of $B(t) - B(s)$ and $X(s)$, we compute the transition probability function

$$\begin{aligned} P(y,t,x,s) &:= \mathbb{P}\{\, X(t) \le y \mid X(s) = x \,\} \\ &= \mathbb{P}\{\, x\exp\left([\mu - \sigma^2/2][t-s] + \sigma[B(t)-B(s)]\right) \le y \,\} \\ &= \mathbb{P}\{\, \exp\left([\mu - \sigma^2/2][t-s] + \sigma[B(t)-B(s)]\right) \le y/x \,\} \\ &= \Phi\left(\frac{\ln(y/x) - (\mu - \sigma^2/2)(t-s)}{\sigma\sqrt{t-s}}\right). \end{aligned}$$

We introduce a useful notation

$$X_s^x(t) := X(t) \mid X(s) = x.$$

This is the value of the process at time t when it takes on value x at time $s \le t$. For $0 \le s < t$, we have the identity
$$X_0^x(t) = X_s^{X_0^x(s)}(t).$$

The Markov property states that conditional on $X_s^{x_0}(t) = x$, the processes $X_s^{x_0}(u)$ for some $s \le u \le t$ and $X_t^x(v)$ for $v \ge t$ are independent.

Let τ be any finite $\{\, \mathcal{F}_t \,\}$-stopping time. Recall that a process has the strong Markov property if
$$\mathbb{P}\{\, X(\tau + t) \le y \mid \mathcal{F}_\tau \,\} \stackrel{a.s.}{=} \mathbb{P}\{\, X(\tau+t) \le y \mid X(\tau) \,\}$$

for every such τ. Solutions to SDEs also have the strong Markov property.

If an SDE has a strong solution $X(t)$, then $X(t)$ has a transition probability function which can be found as a solution to the *Kolmogorov forward-backward equations* as we will soon see.

A transition probability function may exist for SDEs without a strong solution which uniquely determines a Markov process (all finite-dimensional distributions). This process is known as a *weak solution* to an SDE. In this way, we will see that one can definie a solution for an SDE under less stringent conditions on its coefficients.

5.6 Weak Solutions to SDEs

The concept of weak solutions give meaning to an SDE when strong solutions do not exist. Weak solutions are solutions in distribution. They can be realized on some other probability space and exist under less stringent conditions on the coefficients of the SDE.

Definition 5.6.1 (Weak Solution)
Consider the SDE

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dB(t)$$

with initial distribution $X(0)$. Suppose there is some (possibly different) filtered probability space and a Brownian motion $\hat{B}(t)$, process $\hat{X}(t)$, both adapted to the filtration. If $\hat{X}(0) \overset{d}{=} X(0)$, the integrals below are defined for all t, and $\hat{X}(t)$ satisfies

$$\hat{X}(t) = \hat{X}(0) + \int_0^t \mu(\hat{X}(s), s)ds + \int_0^t \sigma(\hat{X}(s), s)d\hat{B}(s),$$

then $\hat{X}(t)$ is a *weak solution* of the SDE.

We say a weak solution is *unique* if any two solutions (on possibly different probability spaces) are equivalent in distribution. In other words, any two solutions have the same finite dimensional distributions.

By definition, a strong solution is also a weak solution. It can be shown that uniqueness of the strong solution (pathwise uniqueness) implies uniqueness of the weak solution.

Next, we see an SDE where no strong solution exists, but a weak solution exists and is unique.

Example 5.6.2 (Tanaka's SDE)
Consider the SDE

$$dX(t) = \operatorname{sign}(X(t))dB(t),$$

where

$$\operatorname{sign}(x) := \begin{cases} 1, & x \geq 0 \\ -1, & x < 0. \end{cases}$$

Since $\sigma := \operatorname{sign}$ is not continuous, it cannot be Lipschitz so the conditions we know for the existence of strong solutions fail. It can be shown that a strong solution to Tanaka's SDE does not exist. We show that Brownian motion is the unique weak solution to Tanaka's SDE.

Let $X(t)$ be some Brownian motion. Consider the process

$$Y(t) := \int_0^t \frac{1}{\operatorname{sign}(X(s))} dX(t) = \int_0^t \operatorname{sign}(X(s)) dX(t).$$

This is well-defined continuous martingale since $\operatorname{sign}(X(t))$ is adapted and

$$\int_0^T \operatorname{sign}^2(X(s)) ds = T < \infty.$$

Moreover,

$$[Y, Y](t) := \int_0^t \operatorname{sign}^2(X(s)) d[X, X](s) = \int_0^t ds = t.$$

By Lévy's characterization of Brownian motion, $Y(t)$ must be a Brownian motion.

Example 5.6.3 (Girsanov's SDE)
We have shown that the SDE

$$dX(t) = |X(t)|^r dB(t)$$

has a strong solution $X \equiv 0$ for $r > 0, t \geq 0$. If $r \geq 1/2$, this is the only strong solution by a previous theorem. In this case, there cannot be any non-zero weak solutions. It is interesting to consider the ODE version of Girsanov's SDE

$$dx(t) = 2|x(t)|^{\frac{1}{2}} dt,$$

which has two solutions $x(t) = 0, t^2$.

For $0 < r < 1/2$, it can be shown that the SDE has infinitely many solutions.

5.7 Construction of Weak Solutions

In this section, we state results on the existence and uniqueness of weak solutions to SDEs of the form

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dB(t).$$

Theorem 5.7.1
If $\mu(\cdot, t), \sigma(\cdot, t) \in C_b$ for every $t > 0$, then the SDE has at least one weak solution starting at any point x and time s.
Moreover, if $\mu(\cdot, t), \sigma(\cdot, t) \in C_b^2$, then the SDE has a unique weak solution starting at any point x and time s.

Even better conditions are available.

Theorem 5.7.2
Suppose $\sigma(x,t)$ is positive, continuous and for any $T > 0$, there is some $K_T > 0$ such that

$$|\mu(x,t)| + |\sigma(x,t)| \leq K_T(1 + |x|)$$

for any $x \in \mathbb{R}, t \in [0,T]$. Then there exists a unique weak solution to the SDE starting at any point x and time s. Moreover, the solution has the strong Markov property.

5.7.1 Canonical Space for Diffusions

Weak solutions to SDEs or diffusions can be realized on the probability space of continuous functions. We indicate

(a) how to define probabilities on this space through transition functions,

(b) how to find the transitions function from a given SDE, and

(c) how to verify that the constructed process indeed satisfies the given SDE.

5.7.2 Probability Space

Weak solutions can be constructed on the canonical space $\Omega = C[0,\infty)$. The Borel σ-field on Ω is the one generated by the open sets according to some metric. For instance, the distance between two continuous functions ω_1, ω_2 can be taken as

$$d(\omega_1,\omega_2) := \sum_{n\geq 1} \frac{1}{2^n} \frac{\sup_{t\in[0,n]}|\omega_1(t) - \omega_2(t)|}{1 + \sup_{t\in[0,n]}|\omega_1(t) - \omega_2(t)|}.$$

Convergence of the elements of Ω in this metric is the uniform convergence of functions on bounded closed intervals $[0,T]$. Diffusions on finite intervals can be realized with the infinity norm for simplicity.

The canonical process $X(t)$ is defined by $X(t,\omega) = \omega(t)$. It is known that the Borel σ-field $\mathcal{F}$ on $C[0,\infty)$ is given by $\sigma\{\, X(t) : t \in [0,\infty) \,\}$. The filtration is defined by the σ-fields $\mathcal{F}_t := \sigma\{\, X(s) : s \in [0,t] \,\}$.

5.7.3 Probability Measure

We now outline the construction of probability measures from a given transition function $P(y,t,x,s)$. For any fixed $x \in \mathbb{R}, s \geq 0$, a probability measure $\mathbb{P} = \mathbb{P}_{x,s}$ on $(\Omega,\mathcal{F})$ can be constructed by using the following properties.

(i) $\mathbb{P}\{\, X(u) = x : u \in [0,s] \,\} = 1$.

(ii) $\mathbb{P}\{\, X(t_2) \in B \mid \mathcal{F}_{t_1} \,\} = P(B, t_2, X(t_1), t_1)$.

The second property asserts that for any Borel sets $A, B \subseteq \mathbb{R}$,

$$\begin{aligned}\mathbb{P}_{t_1,t_2}(A \times B) &= \mathbb{P}\{\, X(t_1) \in A, X(t_2) \in B \,\} \\ &= \int_A \int_B P(dy_2, t_2, y_1, t_1) P_{t_1}(dy_1).\end{aligned}$$

Here $P_{t_1}(C) := \mathbb{P}\{\, X(t_1) \in C \,\}$. This extends to the n-dimensional cylinder sets $J_n \subseteq \mathbb{R}^n$ by

$$P_{t_1,\ldots,t_{n+1}}(J_{n+1}) = \int_{J_{n+1}} P(dy_{n+1}, t_{n+1}, y_n, t_n) P_{t_1,\ldots,t_n}(dy_1 \times \cdots \times dy_n).$$

These probabilities yield the finite dimensional distributions. Consistency of these probabilities is a consequence of the Chapman-Kolmogorov equation for the transition function. Consequently, Kolmogorov's extension theorem states that $\mathbb{P}$ can be uniquely extended to all $\mathcal{F}$. This probability measure $\mathbb{P} = \mathbb{P}_{x,s}$ corresponds to the Markov process started at point x and time s, denoted earlier by $X_s^x(t)$. Thus any transition function defines a probability so that the canonical process is a Markov process.

5.7.4 Transition Function

Under appropriate conditions on the coefficients $\mu(x,t), \sigma(x,t)$, we will see that $P(y,t,x,s)$ is determined from a PDE

$$\frac{\partial u}{\partial s}(x,s) + L_s u(x,s) = 0$$

known as the *Kologorov backward equation*. Here the second order differential operator L_s is given by

$$L_s f(x,s) := \frac{1}{2}\sigma^2(x,s)\frac{\partial^2 f}{\partial x^2}(x,s) + \mu(x,s)\frac{\partial f}{\partial x}(x,s).$$

5.7.5 Weak Solutions & the Martingale Problem

We now consider an equivalent definition of the a weak solution to the SDE

$$dX(t) = \mu(X(t),t)dt + \sigma(X(t),t)dB(t).$$

Definition 5.7.3 (Martingale Problem)
The martingale problem for the coefficients $\mu(x,t), \sigma(x,t)$, or the operator L_s, is as follows.
For each $x \in \mathbb{R}, s > 0$, find a probability measure $\mathbb{P}_{x,s}$ on $\Omega, \mathcal{F}$ such that

(a) $\mathbb{P}_{x,s}\{ X(u) = x, u \in [0,s] \} = 1$.

(b) For any $f \in C^2$ supported on a finite interval, the following process is a martingale under $\mathbb{P}_{x,s}$ with respect to $\mathcal{F}_t$:

$$f(X(t)) - \int_s^t (L_u f)(X(u))du.$$

In the case there is a unique solution to the martingale problem (in distribution), it is said that the martingale problem is *well-posed*.

Remark 5.7.4 If a function vanishes outside a finite interval K, its derivatives also vanish outside that interval. Thus for $f \in C_K^2$, $L_s f$ exists, is continuous, and vanishes outside the interval K. Thus ensures that

$$\mathbb{E}\left[f(X(t)) - \int_s^t (L_u f)(X(u))du \right]$$

exists and the martingale problem is well-defined.

If we consider $f \in C_b^2$, then $L_s f$ exists but may not be bounded and the expectation above may not exist. In this case, one seeks solutions to the *local martingale problem*, and any such solution makes the process above into a *local martingale*.

Theorem 5.7.5
The definition of a weak solutions is equivalent to the definition of a solution to a martingale problem.

Extra concepts of local martingales and their integrals are required to prove the claim rigorously.

Proof (Sketch[a])
Without loss of generality, suppose processes begin at time $s = 0$.

Let $X(t) \sim \mathbb{P}_{x,s}$ be a solution to the martingale problem. Then

$$f(X(t)) - \int_0^t (L_u f)(X(u))du$$

is a martingale for $f(x) = x$ and $f(x) = x^2$ (or approximations by C_K^2 functions on a finite interval). It follows that the following are martingales

$$\begin{aligned} &X(t) - \int_0^t \mu(X(u), u)du && =: Y(t) \\ &X^2(t) - \int_0^t [\sigma^2(X(u), u) + 2\mu(X(u), u)]du. \end{aligned}$$

Intuitively, $Y(t)$ corresponds to the Itô integral portion of the Itô process since we have subtracted the drift portion. We have

$$Y(t) = \int_0^t \sigma(X(u), u) \cdot \frac{dY(u)}{\sigma(X(u), u)}.$$

Here we avoid defining dY. Thus it suffices to show that

$$B(t) := \int_0^t \frac{dY(u)}{\sigma(X(u), u)}$$

is a Brownian motion. This can be accomplished by using the martingale relations above to compute the quadratic variation $[B, B](t) = t$ and invoking Lévy's characterization of Brownian motion.

Conversely, let $X(t)$ be a weak solution so that there is a space supporting a Brownian motion $B(t)$ such that

$$X(t) = \underbrace{X(s)}_{=x} + \int_s^t \mu(X(u), u)du + \int_s^t \sigma(X(u), u)dB(u).$$

Let $f \in C_K^2$. By Itô's formula,

$$\begin{aligned} f(X(t)) &= f(X(s)) + \int_s^t (L_u f)(X(u))du \\ &\quad + \int_s^t f'(X(u))\sigma(X(u), u)dB(u) \\ f(X(t)) - \int_s^t (L_u f)(X(u))du &= f(X(s)) + \int_s^t f'(X(u))\sigma(X(u), u)dB(u). \end{aligned}$$

By assumption, $f'(X(u))$ is bounded. From our work with Itô integrals, the integral on the RHS is a martingale in t for $t \geq s$. We conclude that the martingale problem has a solution. □

[a] `http://localwww.math.unipd.it/~fischer/Didattica/MarkovMP.pdf`

Example 5.7.6
Since $B(t) = \int_0^t 1 \cdot dB(s)$, the theorem above implies that

$$f(B(t)) - \int_0^t \frac{1}{2} f''(B(s)) ds$$

is a martingale for any $f \in C_K^2$. In other words, Brownian motion is a solution to the martingale problem for the Laplace operator $L = \frac{1}{2}\frac{d^2}{dx^2}$. Since Brownian motion exists and is uniquely determined by its distribution, the martingale problem for L is well-posed.

5.8 Backward & Forward Equations

In many applications such as physics, engineering, and finance, the importance of diffusions lies in their connection to PDEs. In these cases, diffusions are often specified by a PDE known as the *Fokker-Planck equation* (forward equation). Although PDEs are difficult to solve analytically, they can typically be solved numerically. In practice, it often suffices to check for existence and uniqueness of solutions and then the solution can be computed by a PDE solver.

This section outlines how to obtain the transition function that determines the weak solution to an SDE

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dB(t), t \geq 0$$

as well as the connection to PDEs

Recall the differential operator $L_s, s \in [0, T]$ given by

$$L_s f(x, s) = \frac{1}{2}\sigma^2(x, s)\frac{\partial^2 f}{\partial x^2}(x, s) + \mu(x, s)\frac{\partial f}{\partial x}(x, s).$$

Definition 5.8.1 (Fundamental Solution)
A *fundamental solution* of the (backward) PDE

$$\frac{\partial u}{\partial s}(x,s) + L_s u(x,s) = 0$$

is a non-negative function $p(y,t,x,s)$ with the following properties:

(i) $p(y,t,x,s)$ is continuous and twice continuously differentiable in x.

(ii) $p(y,t,\cdot,\cdot)$ satisfies the PDE above.

(iii) For any $g \in C_b, t > 0$,

$$u(x,s) := \int_{\mathbb{R}} g(y)p(y,t,x,s)dy$$

is bounded, satisfies the PDE above, and $\lim_{s\uparrow t} u(x,s) = g(x)$ for every $x \in \mathbb{R}$.

Theorem 5.8.2 (Forward-Backward Equations)
[Backward Equation] Suppose $\sigma(x,t), \mu(x,t) \in C_b$ satisfy

(A1) (Uniform Ellipticity) $\sigma^2(x,t) \geq c > 0$.

(A2) $\mu(x,t), \sigma^2(x,t)$ are Hölder continuous. Thus there is some $K > 0$ such that for all $x, y \in \mathbb{R}, s, t > 0$,

$$|\mu(y,t) - \mu(x,s)| + |\sigma^2(y,t) - \sigma^2(x,s)| \leq K(|y-x|^\alpha + |t-s|^\alpha).$$

Then the PDE

$$\frac{\partial u}{\partial s}(x,s) + L_s u(x,s) = 0$$

has a fundamental solution $p(y,t,x,s)$ which is unique and strictly positive.
[Forward Equation] If in addition $\mu(x,t), \sigma(x,t)$ have two partial derivatives with respect to x which are bounded and Hölder continuous with respect to x, then $q(y,t) = p(y,t,x,s)$ also satisfies the PDE

$$-\frac{\partial q}{\partial t}(y,t) + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(\sigma^2(y,t)q(y,t)\right) - \frac{\partial}{\partial y}\left(\mu(y,t)q(y,t)\right) = 0.$$

Note that the SDE from the definition of a fundamental solution is with respect to the backward variables x, s and hence we refer to it as the backward equation. On the other hand, the second PDE in the theorem is with respect to the forward variables y, t and hence its name.

Theorem 5.8.3
Suppose the coefficients of L_s satisfy conditions (A1), (A2) of the previous theorem. Then the backward PDE has a unique fundamental solution $p(y,t,x,s)$. This uniquely defines a transition probability

$$P(A,t,x,s) = \int_A p(y,t,x,s)dy$$

when interpreted as a transition density. Moreover, for any $f \in C_b^{2,1}(\mathbb{R} \times [0,t])$,

$$\mathbb{E}\left[\, f(X(t),t) - f(x,s) \mid X(s) = x \,\right] = \mathbb{E}\left[\int_s^t \left(\frac{\partial}{\partial u} + L_u\right) f(X(u),u)du \,\middle|\, X(s) = x \right]$$

for all $0 \le s < t, x \in \mathbb{R}$.

The transition density $p(y,t,x,s)$ in the above theorem uniquely defines a Markov process $X(t)$ called a *diffusion*. The differential operator L_s is called its *generator*. The

5.8.1 Deriving the Forward-Backward Equations

Intuitively, assuming the drift and diffusion coefficients are sufficiently well-behaved, the transition density of a weak solution to a diffusion-type SDE necessarily follows the forward-backward equations.

Let us sketch how one might derive the Kolmogorov forward-backward equations [ii]. Consider a diffusion-type SDE of the form

$$dX(t) = \sigma(X(t),t)dB(t) + \mu(X(t),t)dt.$$

Throughout, we will assume that μ, σ are sufficiently "nice" so that a weak solution with transition density exists.

Suppose $X(t)$ is a solution to the SDE above and $u(x,t) \in C^{2,1}$. By Itô's formula,

$$du(X(t),t) = (\partial_t + L_t)u(X(t),t)dt + \partial_x u(X(t),t)\sigma(X(t),t)dB(t).$$

By the definition of a stochastic integral, we can compute the value of $X(t)$ given $X(s) = x$.

$$\begin{aligned} &u(X(t),t) - u(x,s) \\ &= \int_s^t (\partial_v + L_v)u(X(v),v)dv + \text{martingale} \end{aligned}$$

[ii] https://cims.nyu.edu/~holmes/teaching/asa19/handout_Lecture10_2019.pdf

The conditional expectation of $X(t)$ is thus

$$\begin{aligned}&\mathbb{E}\left[\, u(X(t),t) - u(x,s) \mid X(s) = x \,\right] \\ &= \mathbb{E}\left[\, \int_s^t (\partial_v + L_v) u(X(v),t) dv \,\middle|\, X(s) = x \,\right].\end{aligned}$$

Note that taking the conditional expectation given $X(s) = x$ eliminates the zero-mean martingale.

Backward Equation

Consider an arbitrary test function $f \in C_K^2$ and define

$$u(x,s) := \mathbb{E}\left[\, f(X(t)) \mid X(s) = x \,\right] = \int_{\mathbb{R}} f(y) p(y,t,x,s) dy, s \leq t.$$

Using the previous identity and differentiating, we extract

$$\begin{aligned}0 &= \lim_{t \downarrow s} \frac{1}{t-s} \mathbb{E}\left[\, u(X(t),t) - u(x,s) \mid X(s) = x \,\right] \\ &= \lim_{t \downarrow s} \mathbb{E}\left[\, \frac{1}{t-s} \int_s^t (\partial_v + L_v) u(X(v),v) dv \,\middle|\, X(s) = x \,\right] \\ &= \partial_s u(x,s) + L_s u(x,s). && \text{mean-value theorem}\end{aligned}$$

Assuming the conditions for Leibniz integral rule holds, we expand the definition of u to see that

$$\begin{aligned}0 &= (\partial_s + L_s) u(x,s) \\ &= (\partial_s + L_s) \int_{\mathbb{R}} f(y) p(y,t,x,s) dy \\ &= \int_{\mathbb{R}} f(y) (\partial_s p(y,t,x,s) + L_s p(y,t,x,s)) dy.\end{aligned}$$

But the test function f was arbitrary, hence we conclude p follows the backward PDE in the variables x, s

$$\boxed{\partial_s p(y,t,x,s) + L_s p(y,t,x,s) = 0}.$$

If in addition, we enforce the initial condition $u(x,t) = f(x)$, the density also satisfies the condition $p(y,t,x,t) = \delta(x-y)$.

Foward Equation

Consider now the formal adjoint L_t^* of L_t defined by

$$\int (L_t f)(y,t) \cdot g(y,t) dy = \int f(y,t) \cdot (L_t^* g)(y,t) dy.$$

It can be shown that

$$(L_t^* g)(y,t) = \frac{1}{2} \frac{\partial^2}{\partial y^2} \left(\sigma^2(y,t) g(y,t) \right) - \frac{\partial}{\partial y} \left(\mu(y,t) g(y,t) \right).$$

Consider an arbitrary test function $f \in C_K^2$. By the identity we computed above,

$$\mathbb{E}\left[f(X(t)) - f(x) \mid X(s) = x \right] = \mathbb{E}\left[\int_s^t (L_v f)(X(v), v) dv \,\middle|\, X(s) = x \right].$$

Rewriting in terms of the transition density,

$$\begin{aligned} \int_{\mathbb{R}} f(y) p(y,t,x,s) dy - f(x) &= \int_s^t \int_{\mathbb{R}} (L_v f)(y,v) \cdot p(y,v,x,s) dy dv \\ &= \int_s^t \int_{\mathbb{R}} f(y) \cdot (L_v^* p)(y,v,x,s) dy dv. \end{aligned}$$

Again, assuming the conditions to apply Leibniz integral rule holds,

$$\int_{\mathbb{R}} f(y) \cdot \partial_t p(y,t,x,s) dy = \int_{\mathbb{R}} f(y) \cdot (L_v^* p)(y,v,x,s) dy.$$

By the arbitrary choice of f, we conclude that the forward (Fokker-Planck) equation holds

$$\boxed{\partial_t p(y,t,x,s) = (L_t^* p)(y,t,x,s)},$$

with the initial condition $p(y,s,x,s) = \delta(x-y)$.

If $X(t)$ has density $\rho(\cdot,t)$, then ρ also satisfies the forward Kolmogorov equation. Indeed,

$$\rho(y,t) = \int_{\mathbb{R}} p(y,t,x,s) \rho_0(x) dx$$

and one can integrate both sides of the forward Kolmogorov equation against ρ_0.

$$\begin{aligned} \partial_t \rho(y,t) &= \int_{\mathbb{R}} \partial_t p(y,t,x,s) \rho_0(x) dx \\ &= \int_{\mathbb{R}} (L_t^* p)(y,t,x,s) \rho_0(x) dx \\ &= L_t^* \rho(y,t). \end{aligned}$$

5.9 Stratonovich Stochastic Calculus

Stochastic integrals in aplication are often taken in the sense of Stratonovich calculus. This calculus is designed in such a way that its basic rules like the chain rule and integration by parts are the same as in standard calculus.

Although the rules of manipulations are the same, the calculi are still different. The processes need to be adapted, just as in Itô calculus. Since Stratonovich stochastic integrals can be reduced to Itô integrals, the standard SDE theory can be used for Stratonovich SDEs. Note also that the Stratonovich integral is more suited to generalizations of stochastic calculus on manifolds.

A direct definition of the Stratonovich integral, denoted $\int_0^t Y(s)\partial X(s)$ to distinguish from Itô integrals, is done as a limit in mean square of Stratonovich approximating sums

$$\sum_{i=0}^{n-1} \frac{1}{2}[Y(t_{i+1}^n) + Y(t_i^n)][X(t_{i+1}^n) - X(t_i^n)]$$

as partitions become finer. Note that we used the average value of Y on the interval $[t_i^n, t_{i+1}^n]$ for Stratonovich calculus as compared to the left most value of Y in the Itô integral.

Alternatively, we use the machinery we already developped for Itô integrals in order to define the Stratonovich integral.

Definition 5.9.1 (Stratonovich Integral)
Let X, Y be continuous adapted procsses such that the Itô integral $\int_0^t Y(s)dX(s)$ is defined. The *Stratonovich integral* is defined by

$$\int_0^t Y(s)\partial X(s) := \int_0^t Y(s)dX(s) + \frac{1}{2}[Y, X](t).$$

The *Stratonovich differential* is defined by

$$Y(t)\partial X(t) := Y(t)dX(t) + \frac{1}{2}d[Y, X](t).$$

5.9.1 Integration by Parts: Stratonovich Product Rule

Theorem 5.9.2
Provided all terms below are defined,

$$\begin{aligned} X(t)Y(t) - X(0)Y(0) &= \int_0^t X(s)\partial Y(s) + \int_0^t Y(s)\partial X(s), \\ \partial(X(t)Y(t)) &= X(t)\partial Y(t) + Y(t)\partial X(t). \end{aligned}$$

Proof
The proof is a direct application the stochastic product rule.

$$\begin{aligned} d(X(t)Y(t)) &= X(t)dY(t) + Y(t)dX(t) + d[X,Y](t) \\ &= X(t)\partial Y(t) + Y(t)\partial X(t). \end{aligned}$$

□

5.9.2 Change of Variables: Stratonovich Chain Rule

Theorem 5.9.3
Let X be continuous and $f \in C^3$. Then

$$\begin{aligned} f(X(t)) - f(X(0)) &= \int_0^t f'(X(s))\partial X(s), \\ \partial f(X(t)) &= f'(X(t))\partial X(t). \end{aligned}$$

Proof
We wish to show that

$$f'(X(t))\partial X(t) = df(X(t))$$

where the RHS is the Itô differential. By Itô's formula,

$$\begin{aligned} df(X(t)) &= f'(X(t))dX(t) + \frac{1}{2}f''(X(t))d[X,X](t) \\ df'(X(t)) &= f''(X(t))dX(t) + \frac{1}{2}f'''(X(t))d[X,X](t). \end{aligned}$$

Hence

$$\begin{aligned} d[f'(X),X](t) &= df'(X(t))dX(t) \\ &= f''(X(t))dX(t)dX(t) \\ &= f''(X(t))d[X,X](t) \end{aligned}$$

and

$$\begin{aligned}
f'(X(t))\partial X(t) &:= f'(X(t))dX(t) + \frac{1}{2}d[f'(X), X](t) \\
&= f'(X(t))dX(t) + \frac{1}{2}f''(X(t))d[X, X](t) \\
&= df(X(t))
\end{aligned}$$

as desired. □

Example 5.9.4
If $B(t)$ is a Brownian motion, then the Stratonovich and Itô differentials below are

$$\begin{aligned}
\partial B^2(t) &= 2B(t)\partial B(t) \\
dB^2(t) &= 2B(t)dB(t) + dt.
\end{aligned}$$

5.9.3 Conversion of Stratonovich SDEs into Itô SDEs

Theorem 5.9.5
Suppose that $X(t)$ satisfies the following SDE in the Stratonovich sense

$$dX(t) = \mu(X(t))dt + \sigma(X(t))\partial B(t),$$

where $\sigma \in C^2$. Then $X(t)$ satisfies the Itô SDE

$$dX(t) = \left(\mu(X(t)) + \frac{1}{2}\sigma'(X(t))\sigma(X(t))\right)dt + \sigma(X(t))dB(t).$$

Thus the diffusion coefficient remains the same in both types of calculi, but the Stratonovich drift coefficient of $\mu(x)$ is transformed to $\mu(x) + \frac{1}{2}\sigma'(x)\sigma(x)$.

Proof
By the definition of the Stratonovich integral,

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t) + \frac{1}{2}d[\sigma(X), B](t).$$

Consider the term $d[\sigma(X), B](t)$. By Itô's formula,

$$\begin{aligned} d\sigma(X(t)) &= \sigma'(X(t))dX(t) + \frac{1}{2}\sigma''(X(t))d[X, X](t) \\ d[\sigma(X), B](t) &= d\sigma(X(t))dB(t) \\ &= \sigma'(X(t))dX(t)dB(t) \\ &= \sigma'(X(t))d[X, B](t) \\ &= \sigma'(X(t))\sigma(X(t))dB(t)dB(t) \\ &= \sigma'(X(t))\sigma(X(t))dt. \end{aligned}$$

Substituting this value into the definition above concludes the proof. □

Chapter 6

Martingales

Martingales play a central role in the modern theory of stochastic processes and stochastic calculus. The following useful properties all hold under reasonable conditions: (i) Martingales arise from Itô integrals and from diffusions. (ii) Martingales have a constant expectation, which remains the same under random stopping. (iii) Moreover, martingales converge almost surely.

6.1 Definitions

Definition 6.1.1 (Martingale)
A stochastic process $M(t)$ where t is continuous ($t \in [0, T]$ or $t \in \mathbb{R}_+$) or discrete ($t = 0, 1, \ldots, T$, or $t \in \mathbb{N}$), adapted to a filtration $\mathcal{F} = \{\mathcal{F}_t\}$, is a *martingale* if

(i) $M(t) \in L^1$ for each t.

(ii) $\mathbb{E}[\, M(t) \mid \mathcal{F}_s \,] \overset{a.s.}{=} M(s)$ for every $0 \leq s < t$.

Similarly, we can define super and submartingales.

Definition 6.1.2 (Supermartingale)
A stochastic process $M(t)$ adapted to a filtration $\mathcal{F}$ is a *supermartingale* if

(i) $M(t) \in L^1$ for each t.

(ii) $\mathbb{E}[\, M(t) \mid \mathcal{F}_s \,] \overset{a.s.}{\leq} M(s)$ for every $0 \leq s < t$.

$M(t)$ is said to be a *submartingale* if $-M(t)$ is a supermartingale.

The mean of a supermartingale is non-increasing with t and vice versa for submartingales. We can test is a supermartingale is a true martingale by checking if the mean remains constant.

Theorem 6.1.3
A supermartingale $M(t), t \in [0,T]$ is a martingale if and only if $\mathbb{E}[\, M(T)\,] = \mathbb{E}[\, M(0)\,]$.

Proof
The condition is certainly necessary. To see that it is sufficient, assume towards a contradiction that for some $s < t$ we have

$$\mathbb{E}\left[\, M(t) \mid \mathcal{F}_s \,\right] < M(s)$$

on a set of positive probability. By taking expectation, we obtain that

$$\mathbb{E}\left[\, M(t)\,\right] < \mathbb{E}\left[\, M(s)\,\right].$$

Since the expectation of a supermartingale is non-increasing, we see that

$$\mathbb{E}\left[\, M(T)\,\right] \le \mathbb{E}\left[\, M(t)\,\right] < \mathbb{E}\left[\, M(s)\,\right] \le \mathbb{E}\left[\, M(0)\,\right].$$

This contradicts the condition, concluding the proof. □

From hereonforth, we assume that we work with CADLAG versions of supermartingales. Recall these versions always exist, roughly speaking if we assume the mean function is right-continuous.

6.1.1 Square Integrable Martingales

Square integrable martingales play a role in the theory of integration.

Definition 6.1.4 (Square Integrable Process)
A process $X(t)$ for $t \in [0,T]$ or $t \in \mathbb{R}_+$ is said to be *square integrable* if

$$\sup_t \mathbb{E}\left[\, X^2(t)\,\right] < \infty.$$

Example 6.1.5
Brownian motion $B(t)$ on a finite interval $t \in [0,T]$ is a square integrable martingale, since $\mathbb{E}[\, B^2(t)\,] = t < T < \infty$. Similarly, $B^2(t) - t$ is a square integrable martingale.

However, neither processes are square integrable over $t \in \mathbb{R}_+$.

Example 6.1.6
Suppose $f \in C_b(\mathbb{R})$. Then Itô integrals of the form $\int_0^t f(B(s))dB(s)$ and $\int_0^t f(s)dB(s)$ are square integrable martingales on any finite time interval $t \in [0,T]$. Indeed, suppose $|f| \leq K$.

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t f(B(s))dB(s)\right)^2\right] &= \mathbb{E}\left[\int_0^t f^2(B(s))ds\right] \\ &\leq K^2 t \\ &\leq K^2 T \\ &< \infty.\end{aligned}$$

Moreover, if $\int_0^\infty f^2(s)ds < \infty$, then $\int_0^t f(s)dB(s)$ is a square integrable martingale on $t \in \mathbb{R}_+$.

6.2 Uniform Integrability

In order to appreciate the definition of uniform integrability of a process, let us first recall that a random variable X is integrable if $\mathbb{E}[\,|X|\,] < \infty$.

Proposition 6.2.1
A random variable X is integrable if and only if

$$\lim_{n\to\infty} \mathbb{E}\left[\,|X| \cdot \mathbb{1}\{\,|X| > n\,\}\,\right] = 0.$$

Proof
If X is integrable, then the statement of the theorem holds by the dominated convergence theorem. Indeed, we have the pointwise limit

$$\lim_{n\to\infty} |X| \cdot \mathbb{1}\{\,|X| > n\,\} = 0,$$

as well as the inequality $|X| \cdot \mathbb{1}\{\,|X| > n\} \leq |X|$.

Conversely, let N be sufficiently large so that

$$\mathbb{E}\left[\,|X| \cdot \mathbb{1}\{\,|X| > n\,\}\,\right] < \infty.$$

Then
$$\mathbb{E}\left[\,|X|\,\right] = \mathbb{E}\left[\,|X| \cdot \mathbb{1}\{\,|X| > N\,\}\,\right] + \mathbb{E}\left[\,|X| \cdot \mathbb{1}\{\,|X| \leq N\,\}\,\right] < \infty$$
as desired. □

We now state the definition of uniform integrability for a stochastic process.

Definition 6.2.2 (Uniform Integrability)
A process $X(t)$ is *uniformly integrable* if

$$\lim_{n\to\infty} \sup_t \mathbb{E}\left[\, |X(t)| \cdot \mathbb{1}\left\{\, |X(t)| > n \,\right\} \right] = 0.$$

Here the supremum is taken over $[0, T]$ in finite time and $\mathbb{R}_+$ if the process is considered on $\mathbb{R}_+$.

Proposition 6.2.3
If $X(t)$ is uniformly integrable, then

$$\sup_t \mathbb{E}\left[\, |X(t)| \,\right] < \infty.$$

The proof is by an identical argument to the case of random variables.

The remainder consists of sufficient conditions of uniform integrability. In particular, we will show that martingales uniformly integrable on any finite time interval.

Theorem 6.2.4
If the process X is dominated by an integrable random variable, say $|X(t)| \leq Y$ for every t such that $\mathbb{E}[\,Y\,] < \infty$, then it is uniformly integrable. In particular, if $\mathbb{E}\left[\, \sup_t |X(t)| \,\right] < \infty$, then it is uniformly integrable.

Proof
The proof is as one might expect.

$$\sup_t \mathbb{E}\left[\, |X(t)| \cdot \mathbb{1}\left\{\, |X(t)| > n \,\right\} \right] \leq \mathbb{E}\left[\, |Y| \cdot \mathbb{1}\left\{\, |Y| > n \,\right\} \right] \to 0$$

as $n \to \infty$. □

There are uniformly integrable processes which are not dominated by an integrable random variable, thus the theorem above is sufficient but not necessary. Another sufficient condition for uniform integrability is provided below.

Theorem 6.2.5 (de la Vallée Poussin)
$X(t)$ is uniformly integrable if and only if for some positive, increasing, convex function $G : (0, \infty) \to \mathbb{R}$ such that $\lim_{x\to\infty} {G(x)}/{x} = \infty$, we have

$$\sup_t \mathbb{E}\left[\, G(|X(t)|) \,\right] < \infty.$$

Proof (Sketch[a])
First suppose such function G exists. Fix $\varepsilon > 0$. By assumption, there is some $x_\varepsilon > 0$ such that for every $x \geq x_\varepsilon$,

$$\frac{G(x)}{x} \geq \frac{1}{\varepsilon} \iff x \leq \varepsilon G(x).$$

It follows that

$$\begin{aligned}
&\sup_t \mathbb{E}\left[\, |X(t)| \cdot \mathbb{1}\{\, |X(t)| > n \,\} \,\right] \\
&\leq \sup_t \varepsilon \mathbb{E}\left[\, G(|X(t)|) \cdot \mathbb{1}\{\, |X(t)| > n \,\} \,\right] && n \geq x_\varepsilon \\
&\leq \varepsilon \sup_t \mathbb{E}\left[\, G(|X(t)|) \,\right] \\
&\to O(\varepsilon). && n \to \infty
\end{aligned}$$

By the arbitrary choice of $\varepsilon > 0$, we conclude that $X(t)$ is indeed uniformly integrable.

Now suppose that X is uniformly integrable. Then for every $m \geq 1$, there is some $x_m \geq \max(x_{m-1}, m)$ such that

$$\sup_t \mathbb{E}\left[\, |X(t)| \cdot \mathbb{1}\{\, |X(t)| > x_m \,\} \,\right] \leq 2^{-m}.$$

Note we can replace 2^{-m} by any summable sequence in m. Consider the function

$$G(x) := \sum_{m\geq 1} (x - x_m)_+$$

where $(\cdot)_+ := \max(0, \cdot)$. This function is positive, increasing, and convex. Furthermore,

$$\begin{aligned}
\lim_{x\to\infty} \frac{G(x)}{x} &= \lim_{x\to\infty} \sum_{m\geq 1} \left(1 - \frac{x_m}{x}\right)_+ \\
&\to \sum_{m\geq 1} 1 \\
&= \infty.
\end{aligned}$$

Finally, for any t,

$$\begin{aligned}
&\mathbb{E}\left[\, G(|X(t)|)\, \right] \\
&= \mathbb{E}\left[\sum_{m\geq 1} (|X(t)| - x_m)_+ \right] \\
&\leq \mathbb{E}\left[\sum_{m\geq 1} |X(t)| \cdot \mathbb{1}\{\, X(t)\,\} > x_m \right] \\
&= \sum_{m\geq 1} \mathbb{E}\left[\, |X(t)| \cdot \mathbb{1}\{\, X(t)\,\} > x_m \,\right] && \text{Monotone Convergence Theorem} \\
&\leq \sum_{m\geq 1} 2^{-m} \\
&= 1.
\end{aligned}$$

□

[a] https://djalil.chafai.net/blog/2014/03/09/de-la-vallee-poussin-on-uniform-integrability/

In practice, this result above is applied with $G(x) = x^p$ for some $p > 1$ and uniform integrability is checked by using moments. For second moments $r = 2$, this yields the following result.

Corollary 6.2.6
If $X(t)$ is uniformly square integrable, ie $\sup_t \mathbb{E}[\, X^2(t)\,] < \infty$, then it is uniformly integrable.

The following result provides a construction of uniformly integrable martingales.

Theorem 6.2.7 (Doob-Lévy Martingale)
Let Y be an integrable random variable, ie $\mathbb{E}[\,|Y|\,] < \infty$, and define

$$M(t) := \mathbb{E}\left[\, Y \mid \mathcal{F}_t \,\right].$$

Then $M(t)$ is a uniformly integrable martingale.

Recall that if τ is a stopping time, then $\mathcal{F}_\tau$ is the set of events observed before or at time T.

$$\mathcal{F}_\tau = \{\, A \in \mathcal{F} : \forall t, A \cap \{\, \tau \leq t \,\} \in \mathcal{F}_t \,\}.$$

Proof
The martingale property can be checked using properties of conditional expectation. We demonstrate uniform integrability for the case of $Y \geq 0$. The general case follows by considering Y^+ and Y^-.

We claim that $M^* := \sup_t M(t) \overset{a.s.}{<} \infty$. Indeed, consider the sequence of stopping times

$$\tau_n := \inf\{\, t \geq 0 : M(t) > n \,\}$$

where the infimum of an empty set is taken to be ∞. Observe that

$$\{\, M^* = \infty \,\} \subseteq \bigcap_n \{\, \tau_n < \infty \,\}.$$

Now, if $\mathbb{P}\{\, M^* = \infty \,\} > 0$, then we have

$$\mathbb{E}[\, M(\tau_n) \,] \geq n \cdot \mathbb{P}\{\, \tau_n < \infty \,\} \geq n\mathbb{P}\{\, M^* = \infty \,\}.$$

Thus $\mathbb{E}[\, M(\tau_n) \,]$ can be make arbitrarily large by varying n. However,

$$\mathbb{E}[\, M(\tau_n) \,] = \mathbb{E}[\, \mathbb{E}[\, Y \mid \mathcal{F}_{\tau_n} \,] \,] = \mathbb{E}[\, Y \,] \leq \mathbb{E}[\, |Y| \,] < \infty$$

which is a contradiction.

Having established the claim, we proceed by computation.

$$\begin{aligned}
\mathbb{E}[\, M(t) \cdot \mathbb{1}\{\, M(t) > n \,\} \,] &:= \mathbb{E}[\, \mathbb{E}[\, Y \mid \mathcal{F}_t \,] \cdot \mathbb{1}\{\, M(t) > n \,\} \,] \\
&= \mathbb{E}[\, Y \cdot \mathbb{1}\{\, M(t) > n \,\} \,] \\
&\leq \mathbb{E}[\, Y \cdot \mathbb{1}\{\, M^* > n \,\} \,].
\end{aligned}$$

The last expression does not depend on t and tends to 0 as $n \to \infty$ since Y is integrable and $M^* \overset{a.s.}{<} 0$. □

Recall that Doob's martingale is *closed* by Y. An immediate corollary follows.

Corollary 6.2.8
Any martingale $M(t)$ on a finite time interval $t \in [0, T]$ is uniformly integrable and is closed by $M(T)$.

We will soon see that a uniformly integrable martingale on $\mathbb{R}_+$ is also closed by some random variable denoted $M(\infty)$, such that the martingale property holds for all $0 \leq s < t \leq \infty$.

6.3 Martingale Convergence

In this section, we consider martingales on the infinite time interval $\mathbb{R}_+$.

Theorem 6.3.1 (Martingale Convergence)
If $M(t), t \in \mathbb{R}_+$ is an integrable martingale (supermartingale or submartingale), ie $\sup_t \mathbb{E}[\,|M(t)|\,] < \infty$, then there is some integrable random variable Y such that

$$\lim_{t\to\infty} M(t) \overset{a.s.}{=} Y.$$

The proof of this theorem is non-trivial and omitted.

Remark 6.3.2 If $M(t)$ is a martingale, then the integrability condition

$$\sup_t \mathbb{E}\left[\,|M(t)|\,\right] < \infty$$

is equivalent to any of the following conditions:

1. $\lim_{t\to\infty} \mathbb{E}[\,|M(t)|\,] < \infty$. This holds as $|x|$ is a convex function, which we have seen implies $|M(t)|$ is a submartingale. But the expectation of a submartingale is an increasing function of t, and hence the supremum is the same as the limit.
2. $\lim_{t\to\infty} \mathbb{E}[\,M^+(t)\,] < \infty$. Since martingales have constant expectation, say $\mathbb{E}[\,M(t)\,] = \mathbb{E}[\,M^+(t)\,] - \mathbb{E}[\,M^-(t)\,] = c \in \mathbb{R}$, we can conclude that $\mathbb{E}[\,|M(t)|\,] = 2\,\mathbb{E}[\,M^+(t)\,] - c$ also has a finite limit in t.
3. $\lim_{t\to\infty} \mathbb{E}[\,M^-(t)\,] < \infty$. Similar.

Now, if $M(t)$ is a submartingale, it suffices to ask that $\sup_t \mathbb{E}[\,M^-(t)\,] < \infty$, and in the case of supermartingales that $\sup -t\mathbb{E}\left[\,M^-(t)\,\right] < \infty$.

Corollary 6.3.3
The following hold.

(a) Uniformly integrable martingales converge a.s.

(b) Square integrable martingales converge a.s.

(c) Positive martingales converge a.s.

(d) Submartingales bounded from above (negative) converge a.s.

(e) Supermartingales bounded from below (positive) converge a.s.

The expectation $\mathbb{E}[\,M(t)\,]$ may or may not converge to the expectation of the limit $\mathbb{E}[\,Y\,]$. In fact, the above holds if and only if $M(t)$ is uniformly integrable.

Theorem 6.3.4
If $M(t)$ is a uniformly integrable martingale, then it converges to a random variable Y a.s. and in L^1. Conversely, if $M(t)$ is a martingale that converges in L^1 to a random variable Y, then $M(t)$ is uniformly integrable and converges a.s. to Y. In any case,

$$M(t) = \mathbb{E}\,[\, Y \mid \mathcal{F}_t \,]\,.$$

The following example illustrates the theorem above.

Example 6.3.5 (Exponential Martingale of Brownian Motion)
Consider $M(t) := e^{B(t) - t/2}$. We have shown that $M(t), t \in \mathbb{R}_+$ is a martingale. Since it is positive, it converges a.s. to some limit Y. By the law of large numbers, $B(t)/t$ converges a.s. to zero. Thus it must be that

$$M(t) = e^{t(B(t)/t - 1/2)} \xrightarrow{a.s.} 0 = Y$$

as $t \to 0$. This is only possible if $M(t)$ is not uniformly integrable, as $\mathbb{E}[\, Y\,] = 0 \neq 1 = \mathbb{E}[\, M(t)\,]$.

Example 6.3.6
Let $f(s)$ be deterministic such that $\int_0^\infty f^2 < \infty$. We show that $M(t) := \int_0^t f(s)dB(s)$ is a uniformly integrable martingale, and find a representation for the closing random variable.

Consider

$$Y := \int_0^\infty f(s)dB(s).$$

Since $\sup_t \mathbb{E}[\, M^2(t)\,] < \infty$ by the isometry property and assumption, $M(t)$ is uniformly integrable and converges a.s. to some limit Y. Convergence also holds in L^1 since the mean is also zero. It follows that Y must be the closing variable. Indeed, from the martingale property of Brownian motion,

$$\begin{aligned}
\mathbb{E}\,[\, Y \mid \mathcal{F}_t \,] &= \mathbb{E}\left[\int_0^\infty f(s)dB(s) \,\middle|\, \mathcal{F}_t \right] \\
&= \int_0^t f(s)dB(s) + 0 \\
&= M(t).
\end{aligned}$$

Example 6.3.7
Let f, Y be as in the previous example. Then consider the bounded positive martingale

$$\begin{aligned} N(t) &:= \mathbb{E}\left[\mathbb{1}\{ Y > 0 \} \mid \mathcal{F}_t \right] \\ &= \mathbb{P}\{ Y > 0 \mid \mathcal{F}_t \} \\ &= \mathbb{P}\left\{ \int_t^\infty f(s)dB(s) + \int_0^t f(s)dB(s) > 0 \;\middle|\; \mathcal{F}_t \right\} \\ &= \Phi\left(\frac{\int_0^t f(s)dB(s)}{\sqrt{\int_t^\infty f^2(d)ds}} \right). \end{aligned}$$

The last equality is due to the normality of the Itô integral for a deterministic function. In particular, the conditioning makes reduces the randomness to the integral on $[0, \infty)$.

By taking $f(s) := \mathbb{1}_{[0,T]}(s)$, we obtain that

$$\Phi\left(\frac{B(t)}{\sqrt{T-t}} \right)$$

is a positive bounded martingale on $[0, T]$.

6.4 Optional Stopping

Recall that a random time τ is a stopping time if for any $t > 0$, the sets $\{ \tau \leq t \} \in \mathcal{F}_t$. In this section, we consider results on stopping martingales at random times.

Theorem 6.4.1 (Basic Stopping Equation)
Let $M(t)$ be a martingale and τ a stopping time. The *stopped process* $M(\tau \wedge t)$ is also a $\{ \mathcal{F}_t \}$-martingale. Moreover,

$$\mathbb{E}\left[M(\tau \wedge t) \right] = \mathbb{E}\left[M(0) \right].$$

We refer the equation of expectations in the theorem above as the *basic stopping equation.*

Note that the process $M(\tau \wedge t)$ is also an $\{ \mathcal{F}_{\tau \wedge t} \}$-martingale.

Example 6.4.2 (Exit Time of Brownian Motion)
For a Brownian motion $B(t)$, let

$$\tau := \inf\{ t : B(t) = a, b \}.$$

By the basic stopping equation,

$$\mathbb{E}\left[B(t \wedge \tau) \right] = B(0) = x.$$

Since $|B(t \wedge \tau)| \leq \max(|a|, |b|)$, dominated convergence applies and by taking $t \to \infty$, we observe that $\mathbb{E}[\, B(\tau) \,] = x$. However, $B(\tau)$ only takes on values a, b, say with probability $1-p, p$. From these equations we deduce that

$$p = \frac{x-a}{b-a}.$$

In the example above, we have $\mathbb{E}[\, M(\tau) \,] = \mathbb{E}[\, M(0) \,]$. However, it is unrealistic to expect this to hold for all stopping times. We now give some sufficient conditions for *optional stopping* to hold.

Theorem 6.4.3 (Optional Stopping)
Let $M(t)$ be a martingale and τ a stopping time such that any of the below hold.

(i) τ is a bounded stopping time, say $\tau \leq K < \infty$.

(ii) $M(t)$ is uniformly integrable.

Then $\mathbb{E}[\, M(\tau) \,] = \mathbb{E}[\, M(0) \,]$.

Interpreting the result above in the context of gambling, there is no loss or gain on average when betting on a martingale, even if a clever stopping rule is used, provided optional stopping holds.

A similar result holds for finite stopping times.

Theorem 6.4.4
Let $M(t)$ be a martingale and τ a finite stopping time. If $\mathbb{E}[\, |M(\tau)| \,] < \infty$ and

$$\lim_{t \to \infty} \mathbb{E}\left[\, M(t)\mathbb{1}\left\{\, \tau > t \,\right\} \,\right] = 0,$$

then $\mathbb{E}[\, M(\tau) \,] = \mathbb{E}[\, M(0) \,]$.

Proof
Write

$$M(\tau \wedge t) = M(t)\mathbb{1}\left\{\, \tau > t \,\right\} + M(\tau)\mathbb{1}\left\{\, \tau \leq t \,\right\}.$$

By the basic stopping equation, $\mathbb{E}[\, M(\tau \wedge t) \,] = \mathbb{E}[\, M(0) \,]$. It follows that

$$\begin{aligned}
\mathbb{E}[\, M(0) \,] &= \mathbb{E}[\, M(\tau \wedge t) \,] \\
&= \mathbb{E}\left[\, M(t)\mathbb{1}\left\{\, \tau > t \,\right\} \,\right] + \mathbb{E}\left[\, M(\tau)\mathbb{1}\left\{\, \tau \leq t \,\right\} \,\right] \\
&\to 0 + \mathbb{E}\left[\, M(\tau) \,\right]. && t \to \infty
\end{aligned}$$

The limit is justified by dominated convergence since

$$|M(\tau)|\mathbb{1}\left\{\, \tau \leq t \,\right\} \leq |M(\tau)|. \qquad \square$$

Recall that for a moment generating function $G_X(u)$ of a random variable X,

$$G_X(0) = \int_{\mathbb{R}} e^{0\cdot x} d\mathbb{P}_X(x) = \mathbb{P}\{\, X \in \mathbb{R} \,\}.$$

Example 6.4.5 (Hitting Time of Brownian Motion)
We derive the Laplace transform of hitting times, which also shows that they are finite. For a Brownian motion starting at 0, consider

$$T_b := \inf\{\, t \geq 0 : B(t) = b \,\}$$

for some $b > 0$.

The exponential martingale $M_u(t)$ of Brownian motion stopped at T_b is also a martingale. Hence for $u > 0$, the basic stopping equation yields

$$\mathbb{E}[\, M_u(t \wedge T_b) \,] = \mathbb{E}\left[\exp\left(uB(t \wedge T_b) - (t \wedge T_b)\frac{u^2}{2} \right) \right] = 1.$$

Now, $|M_u(t\wedge\tau)|$ is bounded above by e^{ub}. Assume for now that $T_b \overset{a.s.}{<} \infty$. Then dominated convergence applies and

$$\mathbb{E}\left[e^{ub - T_b u^2/2} \right] = 1.$$

Setting $u \leftarrow \sqrt{2u}$ and rearranging yields the Laplace transform

$$\psi(u) := \mathbb{E}\left[e^{-uT_b} \right] = e^{-b\sqrt{2u}}.$$

It remains to show the finiteness of T_b, which turns out to be no easier than computing $\psi(u)$ itself. Indeed, it suffices to compute $\lim_{u\downarrow 0} \psi(u)$. We have

$$\begin{aligned}
&1 \\
&= \mathbb{E}[\, M_u(t \wedge T_b) \,] \\
&= \mathbb{E}\left[e^{ub - T_b u^2/2} \mathbb{1}\{\, T_b \leq t \,\} \right] + \mathbb{E}\left[e^{uB(t) - tu^2/2} \mathbb{1}\{\, T_b > t \,\} \right] \\
&\to \mathbb{E}\left[e^{ub - T_b u^2/2} \mathbb{1}\{\, T_b < \infty \,\} \right] + 0 && t \to \infty \\
&= \mathbb{E}\left[e^{ub - T_b u^2/2} \right] && \mathbb{E}\left[e^{ub - T_b u^2/2} \mathbb{1}\{\, T_b = \infty \,\} \right] = 0
\end{aligned}$$

It follows that $\mathbb{E}[\, e^{-T_b u^2/2} \,] = e^{-ub}$ and taking the limit as $u \downarrow 0$ concludes the proof.

Note that we have previously considered the distribution of T_b.

Theorem 6.4.6
Let $X(t), t \geq 0$ be such that for any bounded stopping time τ, $X(\tau)$ is integrable and $\mathbb{E}[X(\tau)] = \mathbb{E}[X(0)]$. Then $X(t), r \geq 0$ is a martingale.

Proof
The proof consists of applying the assumption with appropriate stopping times. Firstly, $X(t)$ is integrable since $\tau = t$ is a stopping time. Without loss of generality, suppose that $X(0) = 0$. In order to check the martingale property, it suffices to show for any $s < t$ and $B \in \mathcal{F}_s$,

$$\mathbb{E}\left[\, X(t)\mathbb{1}_B \,\right] = \mathbb{E}\left[\, X(s)\mathbb{1}_B \,\right].$$

Consider the stopping time

$$\tau := s\mathbb{1}_B + t\mathbb{1}_{B^c}.$$

We have by assumption that

$$\mathbb{E}\left[\, X(\tau) \,\right] = \mathbb{E}\left[\, X(s)\mathbb{1}_B \,\right] + \mathbb{E}\left[\, X(t)\mathbb{1}_{B^c} \,\right].$$

But $\mathbb{E}\left[\, X(\tau) \,\right] = 0$ so

$$\mathbb{E}\left[\, X(s)\mathbb{1}_B \,\right] = -\mathbb{E}\left[\, X(t)\mathbb{1}_{B^c} \,\right].$$

The RHS is the same for all $s \leq t$, so the desired property holds. □

The following result is sometimes known as the optional sampling theorem.

Theorem 6.4.7 (Optional Sampling)
Let $M(t)$ be a uniformly integrable martingale and $\tau_1 \leq \tau_2 \leq \infty$ be two stopping times. Then

$$\mathbb{E}\left[\, M(\tau_2) \mid \mathcal{F}_{\tau_1} \,\right] \overset{a.s.}{=} M(\tau_1).$$

6.4.1 Discrete Time Martingales

A useful application of optional stopping is discrete time martingales such as ones arising from a random walk.

Consider the classic example of Gambler's ruin. You and an opponent begin with x, b dollars respectively and flip a coin each round. The game is zero sum and you win \$1 if the coin yields heads and lose the same amount on tails. Let S_n denote the amount of money after the n-th round. Then S_n is a random walk.

Consider the first case of the fair coin. Then S_n is a martingale. Let τ be the time when the game stops, ie either $S_n = 0, x + b$. Let u denote the probability $S_\tau = 0$. Assuming optional stopping holds, we would have

$$\mathbb{E}\left[\, S_\tau \,\right] = S_0 = x.$$

It follows that

$$x = (x+b)(1-u) + 0 \cdot u \iff u = \frac{b}{x+b}.$$

We now justify the application of optional stopping. S_n is a martingale and τ is a stopping time. Thus the stopped process $S_{n\wedge\tau}$ is also a martingale. It is non-negative and bounded above by $x+b$, hence $S_{n\wedge\tau}$ is a uniformly integrable martingale. Thus it has an L^1 (and a.s.) limit $Y = S_\tau$ with $\mathbb{E}[\,|Y|\,] = x < \infty$. If $\tau \overset{a.s.}{<} \infty$, then we also satisfy the condition

$$\mathbb{E}\left[\, M(t)\mathbb{1}\left\{\, \tau > t \,\right\} \,\right] \to 0$$

since $M(t)$ is bounded in the event $\tau > t$.

It remains now only to check that τ is finite. We know that $S_n^2 - n$ is a martingale, and so is $S_{n\wedge\tau}^2 - n \wedge \tau$. By the basic stopping equation,

$$\mathbb{E}\left[\, S_{n\wedge\tau}^2 \,\right] = \mathbb{E}\left[\, n \wedge \tau \,\right] + \mathbb{E}\left[\, S_0^2 \,\right].$$

By dominated convergence, the LHS has a finite limit Y^2. But then $\mathbb{E}\left[\, n \wedge \tau \,\right] \geq n\mathbb{P}\left\{\, \tau > n \,\right\}$ must also have a finite limit. This implies that $\mathbb{P}\left\{\, \tau > n \,\right\} \to 0$ and $\mathbb{P}\left\{\, \tau < \infty \,\right\} = 1$.

Note that the standard proof of finiteness of τ is done using the theory of Markov chains, ie recurrence states in a random walk.

In the case when the random walk is biased, ie walks left with probability p and right with probability q for some $p \neq q$. In this case, the exponential martingale of the random walk $M_n = (q/p)^{S_n}$ is used. Stopping this martingale, we obtain the ruin probability

$$\frac{(q/p)^{b+x} - (q/p)^x}{(q/p)^{b+x} - 1}.$$

Justification of the equation $\mathbb{E}\left[\, M_\tau \,\right] = M_0$ is similar to the previous case.

Proposition 6.4.8
Let $M(t)$ be a discrete time martingale and τ a stopping time such that $\mathbb{E}[\,|M(\tau)|\,] < \infty$. The following hold.

1. If $\mathbb{E}[\,\tau\,] < \infty$ and $|M(t+1) - M(t)| \leq K$ for some constant $K > 0$, then $\mathbb{E}[\,M(\tau)\,] = \mathbb{E}[\,M(0)\,]$.
2. If $\mathbb{E}[\,\tau\,] < \infty$ and $\mathbb{E}[\,|M(t+1) - M(t)| \mid \mathcal{F}_t\,] \leq K$ for some $K > 0$, then $\mathbb{E}[\,M(\tau)\,] = \mathbb{E}[\,M(0)\,]$.

Proof
We show the first statement. By assumption,

$$\begin{aligned} M(t) &= M(0) + \sum_{i=0}^{t-1} M(i+1) - M(i) \\ &\leq |M(0)| + \sum_{i=0}^{t-1} |M(i+1) - M(i)| \\ &\leq |M(0)| + Kt. \end{aligned}$$

For simplicity, assume that $M(0)$ is deterministic. Then

$$\begin{aligned} &\mathbb{E}\left[\, M(t)\mathbb{1}\{\, \tau > t \,\}\,\right] \\ &\leq |M(0)|\mathbb{P}\{\, \tau > t \,\} + Kt\mathbb{P}\{\, \tau > t \,\} \\ &\leq |M(0)|\mathbb{P}\{\, \tau > t \,\} + K\mathbb{E}\left[\, \tau\mathbb{1}\{\, \tau > t \,\}\,\right] \\ &\to 0. && t \to \infty \end{aligned}$$

The limit is justified by dominated convergence. Thus the condition of optional stopping is satisfied and the result follows.

The second statement can be proved similarly. □

6.5 Localization & Local Martingales

As we know, Itô integrals $\int_0^t X(s)dB(s)$ are martingales under the condition that $\mathbb{E}[\int_0^t X^2(s)ds\,] < \infty$. In general, stochastic integrals with respect to martingales are only *local martingales*. This invites the introduction of local martingales.

A property of a stochastic process $X(t)$ is said to hold *locally* if there is a sequence of stopping times τ_n, called a *localization sequence*, such that $\tau_n \overset{a.s.}{\uparrow} \infty$ as $n \to \infty$, and for each n, the stopped process $X(t \wedge \tau_n)$ has this property.

Example 6.5.1 (Locally Uniformly Integrable)
Recall that a martingale which converges in L^1 is uniformly integrable. Let $M(t)$ be any martingale and consider the localization sequence $\tau_n := n$. Then $M(t \wedge \tau_n) \xrightarrow{1} M(n)$ as $t \to \infty$ and so the uniform integrability property holds for all martingales locally.

local martingales are defined by localizing the martingale property.

Definition 6.5.2 (Local martingale)
An adapted proces $M(t)$ is a *local martingale* if there is a localization sequence τ_n such that $M(t \wedge \tau_n)$ is a uniformly integrable martingale in t for every $n \geq 1$.

Any martingale is a local martingale. However, the converse does not hold in general.

Example 6.5.3 (Itô Integrals)
Consider

$$M(t) := \int_0^t e^{B^2(s)} dB(s), t > \frac{1}{4}.$$

Here $B(t)$ is a one-dimensional Brownian motion with $B(0) = 0$. Define

$$\tau_n := \inf \left\{ t > 0 : e^{B^2(t)} = n \right\}.$$

Then for $t \leq \tau_n$, the integrand is bounded above by n and $M(t \wedge \tau_n)$ is a martingale in t for any n. Moreover, it converges in L^1 to the constant n which implies uniform integrability. Finally, $\tau_n \overset{a.s.}{\uparrow} \infty$ since $B^2(t) \overset{a.s.}{\uparrow} \infty$ and so $M(t)$ is a local martingale.

To see that $M(t)$ is not a martingale, note that for $t > 1/4$, $\mathbb{E}[\, e^{2B^2(t)} \,] = \infty$ so $M(t)$ cannot be a martingale.

Remark 6.5.4 It is not sufficient for a local martingale to be integrable if order to be a true martingale. For instance, positive local martingales are integrable, but in general they are not martingales and only supermartingales. Even uniformly integrable local martingales may not be martingales. However, if a local martingale is dominated by an integrable random variable, then it is a martingale.

Theorem 6.5.5
Let $M(t), t \in \mathbb{R}_+$ be a local martingale such that $|M(t)| \leq Y$ for some $\mathbb{E}[\, Y \,] < \infty$. Then M is a uniformly integrable martingale.

Proof
To see that $M(t)$ is a martingale, fix a localization sequence τ_n. Then for any n and $s < t$,

$$\mathbb{E}\left[\, M(t \wedge \tau_n) \mid \mathcal{F}_s \,\right] = M(s \wedge \tau_n).$$

M is certainly integrable, since $\mathbb{E}[\, |M(t)| \,] \leq \mathbb{E}[\, Y \,] < \infty$. By dominated convergence of conditional expectations, we have

$$\lim_{n \to \infty} \mathbb{E}\left[\, M(t \wedge \tau_n) \mid \mathcal{F}_s \,\right] = \mathbb{E}\left[\, M(t) \mid \mathcal{F}_s \,\right].$$

Thus the martingale property holds by considering the local martingale property and taking limits in n.

To see uniform integrability, recall that if a martingale is dominated by an integrable random variable, then it is uniformly integrable. □

Corollary 6.5.6
Let $M(t), t \in \mathbb{R}_+$ be a local martingale such that for all t,

$$\mathbb{E}\left[\sup_{s\le t}|M(s)|\right] < \infty.$$

Then $M(t)$ is a martingale and as such, is uniformly integrable on any finite interval $[0, T]$. If in addition,

$$\mathbb{E}\left[\sup_{t\ge 0}|M(t)|\right] < \infty,$$

then $M(t)$ is uniformly integrable on $\mathbb{R}_+$.

Positive local martingales occur in financial applications.

Theorem 6.5.7
A non-negative local martingale $M(t), t \in [0, T]$ is a supermartingale.

Proof
Let τ_n be a localization sequence. Then since $M(t \wedge \tau_n) \ge 0$, Fatou's lemma states that

$$\mathbb{E}\left[\liminf_{n\to\infty} M(t\wedge\tau_n)\right] \le \liminf_{n\to\infty} \mathbb{E}\left[\, M(t\wedge\tau_n)\,\right].$$

Since the limit exists by the definition of a localization sequence, we have

$$\begin{aligned}\mathbb{E}\left[\, M(t)\,\right] &= \mathbb{E}\left[\liminf_{n\to\infty} M(t\wedge\tau_n)\right]\\ &\le \liminf_{n\to\infty} \mathbb{E}\left[\, M(t\wedge\tau_n)\,\right]\\ &= \liminf_{n\to\infty} \mathbb{E}\left[\, M(0\wedge\tau_n)\,\right]\\ &= \mathbb{E}\left[\, M(0)\,\right].\end{aligned}$$

Thus $M(t)$ is integrable.

To see the supermartingale property, we follows similar steps. By Fatou's lemma for conditional expectations,

$$\begin{aligned}\mathbb{E}\left[\liminf_{n\to\infty} M(t\wedge\tau_n) \,\middle|\, \mathcal{F}_s\right] &\le \liminf_{n\to\infty} \mathbb{E}\left[\, M(t\wedge\tau_n) \mid \mathcal{F}_s\,\right]\\ &= \liminf_{n\to\infty} M(s\wedge\tau_n)\end{aligned}$$

and we obtain

$$\mathbb{E}\left[\, M(t) \mid \mathcal{F}_s\,\right] \overset{a.s.}{\le} M(s).$$

□

From the theorem above and elementary martingale theory, we obtain the following result.

Theorem 6.5.8
A non-negative local martingale $M(t), t \in [0, T]$ is a martingale if and only if

$$\mathbb{E}\,[\,M(T)\,] = M(0).$$

6.5.1 Dirichlet Class (D)

For a general local martingale, a necessary and sufficient condition to be a uniformly integrable martingale is described in terms of the property of Dirichlet class (D). This class of processes also arises in other areas of calculus and is given in the following section.

Definition 6.5.9 (Dirichlet Class)
A process X is of Dirichlet class (D) if the family

$$\{\, X(\tau) : \tau \text{ is a finite stopping time} \,\}$$

is uniformly integrable.

Any uniformly integrable martingale M is of class (D). Indeed, any such martingale is closed by some $Y = M(\infty)$ and $M(\tau) = \mathbb{E}[\,Y \mid \mathcal{F}_\tau\,]$. Thus the family is indeed uniformly integrable. The converse can be shown using localization, leading to the following result.

Theorem 6.5.10
A local martingale M is a uniformly integrable martingale if and only if it is of class (D).

Proof
Suppose M is a local martingale of class (D). Then there is a localization sequence τ_n. such that $M(t \wedge \tau_n)$ is uniformly integrable martingale in t. For any $s < t$,

$$M(s \wedge \tau_n) = \mathbb{E}\,[\,M(t \wedge \tau_n) \mid \mathcal{F}_s\,]\,.$$

Since $\tau_n \to \infty$, $M(s \wedge \tau_n) \xrightarrow{a.s.} M(s)$. By construction, $s \wedge \tau_n$ is a finite stoppping time. Since M is in (D), the sequence of random variables

$$\{\, M(s \wedge \tau_n) \,\}_n$$

is uniformly integrable. It follows that $M(s \wedge \tau_n) \xrightarrow{1} M(s)$. But then by the properties of conditional expectation,

$$
\begin{aligned}
&\mathbb{E}\left[\,|\mathbb{E}[\,M(t\wedge\tau_n)\mid\mathcal{F}_s\,]-\mathbb{E}[\,M(t)\mid\mathcal{F}_s\,]|\,\right] \\
&=\mathbb{E}\left[\,|\mathbb{E}[\,M(t\wedge\tau_n)\,]-\mathbb{E}[\,M(t)\mid\mathcal{F}_s\,]|\,\right] \\
&\leq\mathbb{E}\left[\,\mathbb{E}[\,|M(t\wedge\tau_n)-M(t)|\mid\mathcal{F}_s\,]\,\right] \\
&=\mathbb{E}\left[\,|M(t\wedge\tau_n)-M(t)|\,\right] \\
&\to 0 && n\to\infty
\end{aligned}
$$

The martingale property thus follows by considering the local martingale property and taking limits in n.

To see the uniform integrability property, apply the definition of (D) with all deterministic stopping times n.

We have already shown the converse, concluding the proof. $\square$

6.6 Quadratic Variation of Martingales

Recall the quadratic variation of a process $X(t)$ is defined as a limit in probability

$$[X,X](t) := \lim \sum_{i=1}^{n}[X(t_i^n)-X(t_{i-1}^n)]^2$$

where the limit is taken over shrinking partitions.

Recall that if $M(t)$ is a martingale and $f(M(t))$ is integrable for some convex function f, then $f(M(t))$ is a submartingale. In particular, $M^2(t)$ is a submartingale so its mean is non-decreasing. By compensating $M^2(t)$ by some non-decreasing process, it is possible to make it into a martingale. It turns out that the quadratic variation process of M is precisely the compensating process. It can be shown that quadratic variation of martingales always exist and is characterized by the above property.

Theorem 6.6.1
Let $M(t)$ be a martingale with finite seconds moments for all t. Then its quadratic variation process $[M,M](t)$ exists. Moreover, $M^2(t)-[M,M](t)$ is a martingale.

Proof (Sketch)
First we observe that

$$
\begin{aligned}
\mathbb{E}[\,M(t)M(s)\,] &= \mathbb{E}[\,\mathbb{E}[\,M(t)M(s)\mid\mathcal{F}_s\,]\,] \\
&= \mathbb{E}[\,M(s)\mathbb{E}[\,M(t)\mid\mathcal{F}_s\,]\,] \\
&= \mathbb{E}\left[\,M^2(s)\,\right].
\end{aligned}
$$

Using this property, we obtain the equality

$$\mathbb{E}\left[\,(M(t)-M(s))^2\,\right] = \mathbb{E}\left[\,M^2(t)\,\right] - \mathbb{E}\left[\,M^2(s)\,\right].$$

This enables us to write

$$\begin{aligned}\mathbb{E}\left[\,M^2(t)-M^2(s)\;\middle|\;\mathcal{F}_s\,\right] &= \mathbb{E}\left[\,(M(t)-M(s))^2\;\middle|\;\mathcal{F}_s\,\right]\\ &= \mathbb{E}\left[\,\sum_{i=0}^{n-1}(M(t_{i+1})-M(t_i))^2\;\middle|\;\mathcal{F}_s\,\right],\end{aligned}$$

where $\{\,t_i\,\}$ is a partition of $[s,t]$. Taking the limit over shrinking partitions, it is possible to show that

$$\mathbb{E}\left[\,M^2(t)-M^2(s)\;\middle|\;\mathcal{F}_s\,\right] = \mathbb{E}\left[\,[M,M](t)-[M,M](s)\mid\mathcal{F}_s\,\right].$$

Rearranging yields the martingale property of $M^2(t)-[M,M](t)$. □

Recall local martingales are locally square integrable. Thus the following result follows.

Theorem 6.6.2
If $M(t)$ is a local maritngale, then $[M,M](t)$ exists. Moreover, $M^2(t)-[M,M](t)$ is a local martingale.

Remark 6.6.3 As part of the proof, we showed that

$$\mathbb{E}\left[\,(M(t)-M(0))^2\,\right] = \mathbb{E}\left[\,M^2(t)\,\right] - \mathbb{E}\left[\,M^2(0)\,\right].$$

Recall that $\mathbb{E}\left[\,M^2(t)\,\right] \geq \mathbb{E}\left[\,M^2(0)\,\right]$ Thus M^2 cannot be a martingale on $[0,t]$ unless $M(t) = M(0)$, ie M is constant on $[0,t]$.

Theorem 6.6.4
Let M be a martingale with $M(0)=0$. Then $[M,M](t)=0$ if and only if $M(s)\overset{a.s.}{=}0$ for all $s\leq t$. The result also holds for local martingales.

The proof follows by the fact that the quadratic variation compensates the submartingale $M^2(t)$. It also follows that $M,[M,M]$ have the same intervals of constancy. Remarkably, we can now show that a non-constant continuous martingale has infinite vriation on any interval.

Theorem 6.6.5
Let M be a continuous local martingale and fix t. If $M(t)$ is not identically equal to $M(0)$, then M has infinite variation over $[0,t]$.

Proof
$M(t) - M(0)$ is a martingale, takes on zero at $t = 0$, and with value at time t that is not identically zero. By the previous theorem, M has positive quadratic variation on $[0, t]$. By then a continuous process of finite variation on $[0, t]$ has zero quadratic over this interval. It follows that M must have infinite variation over $[0, t]$. □

Corollary 6.6.6
If a continuous local martingale has finite variation over an interval, then it must be a constant over that interval.

Remark 6.6.7 Note that there are martingales with finite variation, but by the previous result, they cannot be continuous. An example of such a martingale is the Poisson process martingale $N(t) - t$.

6.7 Martingale Inequalities

Let $M(t)$ be a martingale or local martingale on the $[0, T]$ or $\mathbb{R}_+$.

Theorem 6.7.1
If $M(t)$ is a martingale or positive submartingale, then for $p \geq 1$,

$$\mathbb{P}\left\{\sup_{s\leq t}|M(s)| \geq a\right\} \leq a^{-p}\sup_{s\leq t}\mathbb{E}\left[\,|M(s)|^p\,\right].$$

If $p > 1$,

$$\mathbb{E}\left[\sup_{s\leq t}|M(s)|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[\,|M(t)|^p\,\right].$$

The case of $p = 2$ is known as *Doob's inequality* for martingales:

$$\mathbb{E}\left[\sup_{s\leq T} M^2(s)\right] \leq 4\mathbb{E}\left[\,M^2(T)\,\right].$$

Consequently, if for some $p > 1$,

$$\sup_{t\leq T}\mathbb{E}\left[\,|M(t)|^p\,\right] < \infty,$$

then $M(t)$ is uniformly integrable.

Theorem 6.7.2
If $M(t)$ is a locally square integrable martingale with $M(0) = 0$,

$$\mathbb{P}\left\{ \sup_{t\le T}|M(t)| > a \right\} \le a^{-2}\mathbb{E}\left[\, [M,M](T)\,\right].$$

Theorem 6.7.3 (Davis Inequality)
There are constants $c > 0$ and $C < \infty$ such that for any local martingale $M(t)$ where $M(0) = 0$,

$$c\mathbb{E}\left[\, \sqrt{[M,M](T)}\,\right] \le \mathbb{E}\left[\sup_{t\le T}|M(t)|\right] \le C\mathbb{E}\left[\, \sqrt{[M,M](T)}\,\right].$$

Theorem 6.7.4 (Burkholder-Gundy Inequality)
Suppose $1 \le p < \infty$. There are constants c_p, C_p depending only on p, such that for any local martingale $M(t)$ where $M(0) = 0$,

$$c_p\mathbb{E}\left[\, [M,M]^{p/2}(T)\,\right] \le \mathbb{E}\left[\left(\sup_{t\le T}|M(t)|\right)^p\right] \le C_p\mathbb{E}\left[\, [M,M]^{p/2}(T)\,\right].$$

Moreover, if $M(t)$ is continuous, then the result also holds for $0 < p < 1$.

The above inequalities also hold when T is a stopping time. Proofs of these inequalities involve concepts of stochastic calculus for general processes. We use these inequalities to give sufficient conditions for a local martingale to be a true martingale.

Theorem 6.7.5
Let $M(t)$ be a local martingale where $M(0) = 0$.

(a) If $\mathbb{E}[\,\sqrt{[M,M](t)}\,] < \infty$ for all t. Then $M(t)$ is a uniformly integrable martingale on $[0,T]$ for any finite T.

(b) If $\mathbb{E}[\,[M,M](t)\,] < \infty$ for all t, then $M(t)$ is a martingale where

$$\mathbb{E}\left[\, M^2(t)\,\right] = \mathbb{E}\left[\, [M,M](t)\,\right] < \infty.$$

(c) If $\sup_{t<\infty}\mathbb{E}[\,[M,M](t)\,] < \infty$, then $M(t)$ is a square integrable martingale.

Proof
(a) By Davis' inequality, $\sup_{t \le T}|M(t)|$ is an integrable random variable since

$$\mathbb{E}\left[\sup_{t\le T}|M(t)|\right] \le C\mathbb{E}\left[\sqrt{[M,M](T)}\right] < \infty.$$

Thus $M(t)$ is dominated by an integrabl random variable on any finite time interval. We have see that this implies it is a uniformly integrable martingale.

(b) The condition that $\mathbb{E}[\,[M,M](t)\,] < \infty$ implies the previous condition $\mathbb{E}\left[\sqrt{[M,M](t)}\right] < \infty$. This is because for $X \ge 0$,

$$\mathbb{E}[\,X\,] \ge \left(\mathbb{E}\left[\sqrt{X}\right]\right)^2,$$

as $\mathrm{Var}[\sqrt{X}] \ge 0$.

Alternatively, an application of the Burkholder-Gundy inequality for $p = 2$ shows that $M(t)$ is square integrable for each t. We have previously seen that if $M(t)$ is a martingale with $\mathbb{E}[\,M^2(t)\,] < \infty$, then $M^2(t) - [M,M](t)$ is a martingale. In particular, for any finite t,

$$\mathbb{E}\left[\,M^2(t)\,\right] = \mathbb{E}[\,[M,M](t)\,]$$

as desired.

(c) Notice that both sides of the equation above is non-decreasing and as such have a limit. But then by assumption,

$$\sup_t \mathbb{E}\left[\,M^2(t)\,\right] = \sup_t \mathbb{E}[\,[M,M](t)\,] < \infty$$

and so $M(t)$ is a square integrable martingale by definition. □

6.7.1 Application to Itô Integrals

Let $X(t) = \int_0^t H(s)dB(s)$. Being an Itô integral, X is a local martingale and its quadratic variation is given by

$$[X,X](t) = \int_0^t H^2(s)ds.$$

The Burkholder-Gundy inequality with $p = 2$ yields

$$\mathbb{E}\left[\sup_{t\le T} X^2(t)\right] \le C\mathbb{E}[\,[X,X](T)\,] = \mathbb{E}\left[\int_0^T H^2(s)ds\right].$$

If $\mathbb{E}[\int_0^T H^2(s)ds] < \infty$, then $X(t)$ is a square integrable martingale. Thus from the previous theorem, we recover the fact that

$$\mathbb{E}\left[X^2(t)\right] = \mathbb{E}\left[\int_0^t H^2(s)ds\right].$$

The Davis inequality further gives

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t H(s)dB(s)\right|\right] \leq C\mathbb{E}\left[\sqrt{\int_0^T H^2(s)ds}\right].$$

Thus the condition

$$\mathbb{E}\left[\sqrt{\int_0^T H^2(s)ds}\right] < \infty$$

is a sufficient condition for the Itô integral to be a martingale and in particular, have zero mean. Note that this condition does not however assure second moments exist.

6.8 Continuous Martingales - Change of Time

It can be shown that Brownian motion is the basic continuous martingale from which all continuous martingales can be constructed in one of two ways. The first way is through stochastic integration which we will soon see. This section explores the second way: a random change of time.

6.8.1 Lévy's Characterization of Brownian Motion

Theorem 6.8.1 (Lévy)
A process M with $M(0) = 0$ is a Brownian motion if and only if it is a continuous local martingale with quadratic variation $[M, M](t) = t$.

Proof (Sketch)
If M is a Brownian motion, then the statement holds.

Suppose now that M is a continuous local martingale with $M(0) = 0$ and $[M, M](t) = t$. It follows that $[uM, uM](t) = u^2t$. Using the general theory of stochastic exponential martingales, it can be shown that

$$U(t) := e^{uM(t)-u^2t/2} = e^{uM(t)-[uM,uM](t)/2}$$

is a martingale. It follows that

$$\mathbb{E}\left[e^{uM(t)-u^2t/2} \;\middle|\; \mathcal{F}_s \right] = e^{uM(s)-u^2s/2}$$
$$\mathbb{E}\left[e^{u[M(t)-M(s)]} \;\middle|\; \mathcal{F}_s \right] = e^{u^2(t-s)/2}.$$

Taking expectation of the above yields

$$\mathbb{E}\left[e^{u[M(t)-M(s)]} \right] = e^{u^2(t-s)/2}.$$

This is the moment generating function of a normal distribution with mean zero and variance $t-s$. To see independent increments, we compute

$$\begin{aligned}
\mathbb{E}\left[e^{u[M(t)-M(s)]}e^{vM(s)} \right] &= \mathbb{E}\left[\mathbb{E}\left[e^{u[M(t)-M(s)]}e^{vM(s)} \;\middle|\; \mathcal{F}_s \right] \right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{u[M(t)-M(s)]} \;\middle|\; \mathcal{F}_s \right] e^{vM(s)} \right] \\
&= \mathbb{E}\left[e^{u^2(t-s)/2+vM(s)} \right] \\
&= \mathbb{E}\left[e^{u[M(t)-M(s)]} \right] \mathbb{E}\left[e^{vM(s)} \right].
\end{aligned}$$

But then the moment generating function of the product is merely the product of moment generating functions, implying the independence of increments. □

Example 6.8.2 (Tanaka's SDE)
Any weak solution of Tanaka's SDE

$$dX(t) = \operatorname{sign}(X(t))dB(t), X(0) = 0$$

is a Brownian motion. Indeed, $X(t) = \int_0^t \operatorname{sign}(X(s))dB(s)$ is an Itô integral and thus a (local) martingale. It is continuous and its quadratic variation is given by

$$[X,X](t) = \int_0^t \operatorname{sign}^2(X(s))ds = t.$$

Thus is is a Brownian motion.

6.8.2 Change of Time for Martingales

The main result below states that a continuous martingale M is a Brownian motion with a change of time, where time is measured by the quadratic variation $[M,M](t)$. Namely, there is a Brownian motion $B(t)$ such that $M(t) = B([M,M](t))$. This $B(t)$ is constructed from $M(t)$: Define

$$\tau_t := \inf\left\{ s : [M,M](s) > t \right\}.$$

Note that if $[M,M](t)$ is strictly increasing, then τ_t is its inverse.

Theorem 6.8.3 (Dambis-Dubins-Schwarz)
Let $M(t), M(0) = 0$ be a continuous martingale such that $[M, M](t) \uparrow \infty$ and

$$\tau_t := \inf\{ s : [M, M](s) > t \}.$$

Then the process $B(t) := M(\tau_t)$ is a Brownian motion with respect to the filtration $\mathcal{F}_{\tau_t}$. Moreover, $[M, M](t)$ is a stopping time with respect to this filtration and the martingale M can be obtained from the Brownian motion B by the change of time $M(t) = B([M, M](t))$.
This result also holds when M is a continuous local martingale.

Proof (Sketch)
Let $M(t)$ be a local martingale. Then each τ_t is a finite stopping time since $[M, M](t) \uparrow \infty$. Thus each $\mathcal{F}_{\tau_t}$ is well-defined.

Note that
$$\{ [M, M](s) \leq t \} = \{ \tau_t \geq s \}.$$
This implies that $[M, M](t)$ are stopping times for $\mathcal{F}_{\tau_t}$.

Since $[M, M](s)$ is continuous, we must have $[M, M](\tau_t) = t$. We claim that $X(t) := M(\tau_t)$ is a continuous local martingale. Indeed, the map $t \mapsto \tau_t$ can only have jump discontinuities when $[M, M](t)$ takes on constant values over an interval. However, we have shown that $M, [M, M]$ has the same intervals of constancy so that the composition $t \mapsto M(\tau_t)$ is continuous. The local martingale property can also be verified.

By the characterizing property of quadratic variation of (local) martingales, ie $X^2(t) - [X, X](t)$ is a local martingale, we obtain

$$\mathbb{E}\left[X^2(t) \right] = \mathbb{E}[[X, X](t)] = \mathbb{E}[[M, M](\tau_t)] = t.$$

Hence X is a Brownian motion by Lévy's characterization theorem.

The converse can be proven using the observation that

$$X([M, M](t)) = M(\tau_{[M,M](t)}) = M(t).$$

This once again relies on $M, [M, M]$ having the same intervals of constancy. $\square$

Example 6.8.4
Let $M(t) = \int_0^t f(s) dB(s)$, with f continuous and deterministic. Then M is a Gaussian martingale and its quadratic variation is given by

$$[M, M](t) = \int_0^t f^2(s) ds.$$

For example with $f(s) = s$, we have $M(t) = \int_0^t s dB(s)$ and

$$[M, M](t) = \int_0^t s^2 ds = \frac{t^3}{3}.$$

for this particular example, $[M, M](t)$ is deterministic and increasing, therefore yielding a an inverse

$$\tau_t = (3t)^{1/3}.$$

Let

$$X(t) := M(\tau_t) = \int_0^{\sqrt[3]{3t}} s dB(s).$$

It is clear that $X(t)$ is continuous as it is a composition of continuous functions. It is also a martingale with quadratic variation $\tau_t^3/3 = t$. Hence by Lévy's theorem, it is indeed a Brownian motion. The previous theorem also guarantees that

$$M(t) = X(t^3/3).$$

Example 6.8.5
If

$$M(t) = \int_0^t H(s) dB(s)$$

is an Itô integral, then it is a local martingale with quadratic variation

$$[M, M](t) = \int_0^t H^2(s) ds.$$

If $\int_0^\infty H^2(s) ds = \infty$, then

$$M(t) = \hat{B}\left(\int_0^t H^2(s) ds\right),$$

where $\hat{B}(t)$ is a Brownian motion that can be recovered from $M(t)$ with an appropriate change of time.

Example 6.8.6 (Brownian Bridge)
The solution for Brownian bridge can be written as

$$X(t) = a\left(1 - \frac{t}{T}\right) + b\frac{t}{T} + (T - t)\int_0^t \frac{1}{T - s} dB(s).$$

Let $Y(t)$ be the result of the Itô integral term. Since for any $t < T$, Y is a continuous martingale with quadratic variation

$$[Y, Y](t) = \int_0^t \frac{1}{(T - s)^2} ds = \frac{t}{T(T - t)},$$

it follows by the DDS theorem that

$$Y(t) = \hat{B}\left(\frac{t}{T(T-t)}\right)$$

for some Brownian motion $\hat{B}$. Therefore the SDE above has the following representation:

$$X(t) = a\left(1 - \frac{t}{T}\right) + b\frac{t}{T} + (T-t)\hat{B}\left(\frac{t}{T(T-t)}\right)$$

for $t \in [0, T]$.

In this representation, $t = T$ is allowed and understood by continuity since the limit of $tB(1/t) \to 0$ as $t \to 0$ by the law of large numbers of Brownian motion.

6.8.3 Change of Time in SDEs

We use the DDS theorem for constructing weak solutions of some SDEs. Let

$$X(t) = \int_0^t \sqrt{f'(s)}dB(s)$$

where $f(t), f(0) = 0$ is an adapted, positive, increasing, and differentiable process. $X(t)$ is a local martingale with quadratic variation

$$[X, X](t) = \int_0^t f'(s)ds = f(t).$$

Thus $\tau_t = f^{-1}(t)$, the inverse of f, and according to the DDS theorem, the process $X(f^{-1}(t)) =: \hat{B}(t)$ is a Brownian motion with respect to $\mathcal{F}_{\tau_t}$ and $X(t) = \hat{B}(f(t))$. This is summarize in the following result.

Theorem 6.8.7
Let $f(t)$ be an adapted, positive, increasing, differentiable process, and consider

$$dX(t) = \sqrt{f'(t)}dB(t).$$

Then the process $X(f^{-1}(t)) =: \hat{B}(f(t))$ is a weak solution.

We can express the conclusion of the theorem above as

$$d\hat{B}(f(t)) = \sqrt{f'(t)}dB(t).$$

In the case of non-random change of time in Brownian motion $B(f(t))$, it can be directly verified that $M(t) = B(f(t))$ is a martingale with respect to the filtration $\mathcal{F}_{f(t)}$. The quadratic variation of $B(f(t))$ is

$$[M, M](t) = [B(f), B(f)](t) = f(t)$$

and can be calculated directly.

Example 6.8.8 (Ornstein-Uhlenbeck)
Let

$$f(t) := \frac{\sigma^2(e^{2\alpha t}-1)}{2\alpha}.$$

The process $B(\sigma^2 e^{2\alpha t}-1/2\alpha)$ is a weak solution to the SDE

$$dX(t) = \sigma e^{\alpha t} dB(t).$$

Next we construct a weak solution to SDEs of the form

$$dX(t) = \sigma(X(t))dB(t)$$

for some $\sigma(x) > 0$ such that

$$G(t) := \int_0^t \frac{1}{\sigma^2(B(s))} ds$$

is finite for any $t \geq 0$ but $G(t) \xrightarrow{a.s.} \infty$ as $t \to \infty$. Then $G(t)$ is adapted, continuous, and strictly increasing to $G(\infty) := \infty$. Thus it has an inverse

$$\tau_t := G^{-1}(t).$$

Note that for each fixed t, τ_t is a stopping time as it is the first time the process $G(s)$ hits t. Moroever, τ_t is increasing.

Theorem 6.8.9
Let $\sigma(x) > 0$ be such that

$$G(t) := \int_0^t \frac{1}{\sigma^2(B(s))} ds$$

is finite for any $t \geq 0$ but $G(t) \xrightarrow{a.s.} \infty$ as $t \to \infty$. Define

$$\tau_t := G^{-1}(t).$$

The process $X(t) := B(\tau_t)$ is a weak solution to the SDE

$$dX(t) = \sigma(X(t))dB(t).$$

Proof
We have

$$X(t) = B(\tau_t) = B(G^{-1}(t)).$$

By the previous theorem with $f = G^{-1}$,

$$dB(G^{-1}(t)) = \sqrt{(G^{-1})'(t)} d\hat{B}(t)$$

for some Brownian motion $\hat{B}(t)$.

By the inverse function theorem,

$$
\begin{aligned}
(G^{-1})'(t) &= \frac{1}{G'(G^{-1}(t))} \\
&= \left(\frac{1}{\sigma^{-2}(B(G^{-1}(t)))}\right)^{-1} \\
&= \sigma^2(B(\tau_t)).
\end{aligned}
$$

It follows that

$$dX(t) = dB(\tau_t) = \sigma(B(\tau_t))d\hat{B}(t) = \sigma(X(t))d\hat{B}(t)$$

as desired. □

An application of the DDS theorem yields a result on the uniqueness of the solution. This result is weaker than that of Engelbert-Schmidt.

Theorem 6.8.10
Let $\sigma(x) \geq \delta$ for some fixed $\delta > 0$. Then the SDE

$$dX(t) = \sigma(X(t))dB(t)$$

has a unique weak solution.

Proof (Sketch)
Let $X(t)$ be such a weak solution. Then $X(t)$ is a local martingale and there is a Brownian motion $\beta(t)$ such that $X(t) = \beta([X, X](t))$. Now,

$$[X, X](t) = \int_0^t \sigma^2(X(s))ds = \int_0^t \sigma^2(\beta([X, X](s)))ds.$$

Thus $[X, X](t)$ is a solution to the ODE

$$da(t) = \sigma^2(\beta(a(t)))dt.$$

It can be shown that the solution to this ODE is unique, the solution to the SDE in question is also unique. □

Let us now consider a more general change of time for diffusion-type SDEs.

Theorem 6.8.11
Consider the following SDE

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t).$$

Let $g(x)$ be a positive function for which

$$G(t) := \int_0^t g(X(s))ds$$

is finite for all $t \geq 0$ but $G(t) \xrightarrow{a.s.} \infty$ as $t \to \infty$. Furthermore, define

$$\tau_t := G^{-1}(t).$$

Let $X(t)$ be a weak solution to the SDE above and define

$$Y(t) := X(\tau_t).$$

Then $Y(t)$ is a weak solution to the SDE

$$dY(t) = \frac{\mu(Y(t))}{g(Y(t))}dt + \frac{\sigma(Y(t))}{\sqrt{g(Y(t))}}dB(t), Y(0) = X(0).$$

Example 6.8.12 (Lamperti's Change of Time)
Let $X(t)$ satisfy Feller's branching diffusion SDE

$$dX(t) = \mu X(t)dt + \sigma\sqrt{X(t)}dB(t), X(0) = x > 0.$$

Here $\mu, \sigma > 0$ are some positive constants. Lamperti's change of time is given by

$$G(t) = \int_0^t X(s)ds,$$

so with $g(x) = x$ as the identity function. Then $Y(t) := X(\tau_t)$ satisfies the SDE

$$\begin{aligned} dY(t) &= \frac{\mu Y(t)}{Y(t)}dt + \frac{\sigma\sqrt{Y(t)}}{\sqrt{Y(t)}}dB(t) \\ &= \mu dt + \sigma .dB(t) \qquad\qquad Y(0) = x \end{aligned}$$

It follows that

$$Y(t) = x + \mu t + \sigma B(t).$$

Thus with a random change of time, the branching diffusion is a Brownina motion with drift. Note the other direction is also true: A banching diffusion can be obtained from a Brownian motion with drift.

We have now seen two main methods for solving SDEs: change of variables (Itô's formula) and change of time. There is another method known as the change of measure.

Part III

Applications