Notes on Measure Theory, Probability, and Stochastic Calculus from "Real Analysis for Graduate Students" and "Stochastic Calculus: An Introduction with Applications"

Notes from material by Richard Bass and Gregory Lawler.

LaTeX by Agustin Esteva

July 19, 2024

Contents

Chapter 1

Preliminaries

1.1 Notation and Terminology

The following notation will be established:

1. $A_i \uparrow$ is $A_1 \subseteq A_2 \subseteq \dots$ and $A_i \uparrow A$ if $A = \bigcup\limits_{i}^{\infty} A$. $A_i \downarrow$ is notation for supersets.

2. $x \wedge y = \min(x, y)$ and $w \vee y = \max(x, y)$. Therefore, we can write any $x \in \mathbb{R}$ as a sum of its positive parts, $x^+ = x \vee 0$ and its negative parts $x^- = x \wedge 0$.

3. $\overline{z}$ is the complex conjugate of z.

4. A function is monotone if f is either increasing or decreasing.

5.
$$\lim_{n \to \infty} \sup(a_n) = \inf_n \sup_{m \geq n} (a_m)$$
$$\lim_{n \to \infty} \inf(a_n) = \sup_n \inf_{m \geq n} (a_m)$$
For example, given $a_n = \begin{cases} 1 & n = 2k \\ \frac{1}{n} & n = 2k+1 \end{cases}$, then:
$$\lim_{n \to \infty} \sup(a_n) = 1 \qquad \lim_{n \to \infty} \inf(a_n) = 0$$
Moreover, we say that the limit of a sequence exists if those two limits exist and equal each other.

1.2 Undergraduate Mathematics

Definition 1.2.1: Metric Space (Hausdorff Space)

A set X is a *metric space* if there exists a function $d : X \times X \to \mathbb{R}$, called a *metric*, such that:

1. $d(x, y) = d(y, x)$ for all $x, y \in X$.
2. $d(x, y) \geq 0$ for all $x, y \in X$ and $d(x, y) = 0$ if $x = y$.
3. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

In probability, we most often use $X = \Omega$, or a set of points ω, where Ω consists of all the possible results or outcomes, ω, of an experiment.

Definition 1.2.2: Event and Sample points

We say that a subset of Ω is an *event* and an element $\omega \in \Omega$ is a *sample point*.

Definition 1.2.3: Open Ball

Given a metric space X, we define the *open ball* of radius r centered at x to be:

$$B(x, r) = \{y \in X | d(x, y) < r\}$$

Definition 1.2.4: Interiors and Closures

The *interior* of A is the set $A^o \subset X$, such that for all $x \in X$, there exists an $r_x > 0$ such that $B(x, r_x) \subset A$. The *closure* of A is the set $\overline{A}$ such that every open ball centered at x contains at least one point of A.

Definition 1.2.5: Continuity

$f : X \to \mathbb{R}$ is continuous at some $x \in X$ if for all $\epsilon > 0$, there exists a $\delta > 0$ such that if $d(x, y) < \delta$, then $|f(x) - f(y)| < \epsilon$. f is continuos if it's continuous for all $x \in X$.

Remark.

> This is equivalent to stating that for every open set, U, $f^{-1}(U)$ is open in X.

Definition 1.2.6: Convergence and Completeness

We say that a series, a_n, *converges* to some a, if for all $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that if $n \geq N$, then $d(a_n, a) < \epsilon$. We say that a series is *Cauchy Convergent* if for all $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that if $n, m \geq N$, then $d(a_n, a_m) < \epsilon$. X is *complete* if every Cauchy sequence in X converges to a point in X.

Definition 1.2.7: Covers and Compactness

An open cover, $\mathcal{G}$, of X, is a collection of open sets, $\mathcal{G} = \{G_\beta\}_{\beta \in V}$, such that $X \subset \bigcup_V G_\beta$. We say that X is compact if every open cover has a finite open subcover, i.e, there exist $G_1, G_2, \ldots, G_n \in \mathcal{G}$ such that $X \subset \bigcup_{i=0}^n G_i$.

Proposition 1.2.8

If X is compact and $F \subset X$ is closed, then F is compact.

Proof. Let $\mathcal{G}$ be an open cover of F. Consider some $\mathcal{G}' = \mathcal{G} \cup (\mathbb{R} \setminus F)$. For all $x \in X$:

1. If $x \in F$, then because $\mathcal{G}$ is a cover of F, $F \subset \bigcup_{G \in \mathcal{G}'} G$.

2. If $x \in (X \setminus F)$, then because $x \in (\mathbb{R} \setminus F)$, then $X \subset \bigcup_{G \in \mathcal{G}'} G$

Therefore, $\mathcal{G}'$ is a cover of X. Because F is closed, then $\mathbb{R} \setminus F$ and $\mathcal{G}$ are open, then $\mathcal{G}'$ is an open cover of X. Because X is compact, then there exists some finite open cover, $\mathcal{G}'' \subset \mathcal{G}'$ of X. Consider the set $\mathcal{G}''' = \mathcal{G}'' \setminus \{\mathbb{R} \setminus F\}$. For all $y \in F$, because $F \subset X \subset \bigcup_{G \in G''} G \cup \{\mathbb{R} \setminus F\}$, then $y \in G''$ for some $G'' \in \mathcal{G}''$. Because $y \notin \{\mathbb{R} \setminus F\}$, then $G'' \neq \{\mathbb{R} \setminus F\}$. Thus, $G'' \in \mathcal{G}'''$. Therefore, $\mathcal{G}'''$ is an open finite cover of F, and therefore F is compact. □

Proposition 1.2.9

(EVT): If X is compact and $f : X \to \mathbb{R}$ is continuous, then there exist $x_1, x_2 \in X$ such that

$$f(x_1) = \inf_{x \in X} f(x) \qquad f(x_2) = \sup_{x \in X} f(x)$$

Definition 1.2.10: Normed Linear Space

We say that X is a *normed linear space* if there exists a map $x \to ||x||$ such that:

1. $||x|| \geq 0$ for all $x \in X$ and $||x|| = 0$ iff $x = 0$.

2. $||cx|| = |c|||x||$ for all $c \in \mathbb{F}$, $x \in X$.

3. $||x + y|| \leq ||x|| + ||y||$ for all $x, y \in X$.

Remark.

> Given X, a normed linear space, one can set $||x - y|| = d(x, y)$ to turn X into a metric space.

Definition 1.2.11: Partial Order

We say that a set X has a *partial order*, " $\leq$ " if:

1. $x \leq x$ for all $x \in X$.
2. If $x \leq y$ and $y \leq x$, then $x = y$.
3. If $x \leq y$ and $y \leq z$, then $x \leq z$.

Remark.

> Note that this does not imply that for all $x, y \in X$, either $x \leq y$ or $y \leq x$.

Proposition 1.2.12

If X is closed and bounded, then X is compact.

Proposition 1.2.13

If $U \subset \mathbb{R}$ is open, then U can be written as the countable union of disjoint open intervals.

Proposition 1.2.14

If $f : \mathbb{R} \to \mathbb{R}$ is an increasing function, then both the left and right hand limits exists for every x. Moreover, the set of discontinuities is countable.

Chapter 2

Families of Sets

2.1 Algebras and $\sigma-$algebras

Definition 2.1.1: $\sigma-$algebra

An *algebra* is a collection, $\mathcal{A}$, of subsets of X such that:

1. $\emptyset, X \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $X \setminus A \in \mathcal{A}$.
3. If $A_1, A_2, \ldots, A_n \in \mathcal{A}$, then $\bigcap\limits_{i=0}^{n} A_i$ and $\bigcup\limits_{i=0}^{n} A_i$ are in $\mathcal{A}$.

Moreover, a $\sigma-$*algebra* is defined to be an algebra with the added condition that:

4. If $A_1, A_2, \dots \in \mathcal{A}$, (COUNTABLE) then $\bigcap\limits_{i=0}^{\infty} A_i$ and $\bigcup\limits_{i=0}^{\infty} A_i$ are in $\mathcal{A}$.

Definition 2.1.2: Probability Fields

Consider a nonempty space Ω, a class, $\mathcal{F}$ of subsets of Ω is a *field* if:

1. $\Omega \in \mathcal{F}$;
2. $A \in \mathcal{F}$ and $A^c \in \mathcal{F}$;
3. $A, B \in \mathcal{F}$ imply $A \cup B \in \mathcal{F}$.

with $\sigma-$field constructed the same as above.

Remark.

> (3) ensures that $\mathcal{F}$ is closed under finite intersections, as by DeMorgan's Law:
>
> $$A \cup B = (A^c \cap B^c)^c \qquad A \cap B = (A^c \cup B^c)^c$$

Remark.

If $A \in \mathcal{F}$, we say that A is *measurable* $\mathcal{F}$ or an $\mathcal{F}-$set.

Example.

Let $\mathcal{F}$ consist of the finite and the cofinite sets (A is cofinite if A^c is finite). Then $\mathcal{F}$ is a field. If Ω is finite, then $\mathcal{F}$ contains all subsets of Ω, and is thus a $\sigma-$field. If it is infinite, then it is simply a field

---

Definition 2.1.3: Measurable Space

A *measurable space* is denoted by the pair $(X, \mathcal{A})$. Note that a set A is measurable only if $A \in \mathcal{A}$

Remark.

By definition, $\emptyset, X$ are measurable

Example.

1. Let $X = \mathbb{R}$ and $\mathcal{A}$ be the collection of all the sets in $\mathbb{R}$, then $\mathcal{A}$ is a $\sigma-$algebra.
   (a) 1 is trivial.
   (b) 2 comes with the definition of a power set.
   (c) 3 comes because $\mathbb{R}$ is infinite.
2. Let $X = \mathbb{R}$ and $\mathcal{A} = \{A \subset \mathbb{R} | A \text{is countable or} A^C \text{is countable}\}$
   (a) 1 is true because $\emptyset$ is countable.
   (b) 2 is trivial.
   (c) Let $A_1, A_2, \dots \in \mathcal{A}$ be countable, then $\bigcup_i^\infty A_i$ is countable. Similar for complement, but using that the complement of unions is the intersection of complements which is contained in a complement.
3. Let $X = [0, 1]$ and $\mathcal{A} = \{\emptyset, X, [0, \frac{1}{2}), (\frac{1}{2}, 1]\}$ is a $\sigma-$algebra.
   (a) Duh.
   (b) Duh.
   (c) Duh.
4. Let $X = [0, 3]$ and $\mathcal{A} = \{\emptyset, X, \{1\}, \{2, 3\}\}$ is a $\sigma-$algebra.

Lemma 2.1.4: Intersection of $\sigma-$algebras

If $\mathcal{A}_\alpha$ is a $\sigma-$algebra for each α in a collection of $\alpha \subset I$, where I is a nonempty index set, then $\bigcap_{\alpha \in I} \mathcal{A}_\alpha$is a $\sigma-$algebra

---

Remark.

> $calF$ is closed under any countable set-theoretic operations

---

Definition 2.1.5: σ-algebra generated by a collection of sets

If we have a collection, $\mathcal{C}$ of subsets of X, we define

$$\sigma(\mathcal{C} = \cap\{\mathcal{A}_\alpha | \mathcal{A}_\alpha \text{is a } \sigma-\text{algebra}, \mathcal{C} \subset \mathcal{A}_\alpha\})$$

Remark.

> By Lemma 2.1.3, $\sigma(\mathcal{C})$ is a $\sigma-$algebra. Note also that this is not an intersection of empty classes, as there always exists $\mathcal{A}_X$, since any collection of subsets of X will be a subset to this collection.

Remark.

> If $\mathcal{C}_1 \subset \mathcal{C}_2$, then $\sigma(\mathcal{C}_1) \subset \sigma(\mathcal{C}_2)$ and $\sigma(\sigma(\mathcal{C})) = \sigma(\mathbf{C})$.

Consider some small class, $\mathcal{A}$. Set operations usually lead to sets outside of $\mathcal{A}$, and thus we need to consider the class of sets such that:

1. $\mathcal{A}$ is contained in this new class.
2. This new class is a σ field.
3. This new class is as small as possible

We define this to be the intersection of all the $\sigma-$fields containing $\mathcal{A}$, and is denoted $\sigma(\mathcal{A})$. Note that $\sigma(\mathcal{A})$ has the following properties:

1. $\mathcal{A} \subset \sigma(\mathcal{A})$.
2. $\sigma(\mathcal{A})$ is a $\sigma-$field.
3. If $\mathcal{A} \subset \mathcal{G}$ and $\mathcal{G}$ is a $\sigma-$field, then $\sigma(\mathcal{A}) \subset \mathcal{G}$.

Definition 2.1.6: Borel σ Algebras and Borel Sets

If X is a metric space and $\mathcal{G}$ is the collection of open sets of X, then $\mathcal{B} = \sigma(\mathcal{G})$ is a *Borel $\sigma-$algebra* on X. Elements of $\mathcal{B}$ are *Borel sets* and are said to be Borel measurable.

Let $\mathcal{G}$ be the class of subintervals of $\Omega = (0,1]$, and definite $\mathcal{B} = \sigma(\mathcal{G})$. The elements of $\mathcal{B}$ are the *Borel sets* on the unit interval.

Proposition 2.1.7

If $X = \mathbb{R}$, then the Borel $\sigma-$algebra on $\mathbb{R}$, $\mathcal{B}$, is generated by each of the following collections of sets:

1. $\mathcal{C}_1 = \{(a,b)|a,b \in \mathbb{R}\}$.
2. $\mathcal{C}_2 = \{[a,b]|a,b \in \mathbb{R}\}$.
3. $\mathcal{C}_3 = \{(a,b]|a,b \in \mathbb{R}\}$.
4. $\mathcal{C}_4 = \{(a,\infty)|a \in \mathbb{R}\}$.

Proof. Let $\mathcal{G}$ be the collection of open sets:

1. Since every element of $\mathcal{C}_1$ is open, then $\mathcal{C}_1 \subset \mathcal{G}$, and thus $\sigma(\mathcal{C}_1) \subset \sigma(\mathcal{G})$. If G is open, then G is a finite union of open sets. By definition, every finite open interval is in $\mathcal{C}_1$. Consider then (a,∞). But this is again just $\bigcup_{n=0}^{\infty}(a, a+n)$. Therefore, $G \in \sigma(\mathcal{C}_1)$, and so by the previous remark, $\sigma(\mathcal{G}) \subset (\sigma(\mathcal{C}_1))$. Therefore, $\mathcal{B} = \sigma(G) = \sigma(\mathcal{C}_1)$.
2. One can use the fact that

$$[a,b] = \bigcap_{n=0}^{\infty}(a - \frac{1}{n}, b + \frac{1}{n}) \in \sigma(\mathcal{G})$$

to show that if $[a,b] \in \mathcal{C}_2$, then $\sigma(\mathcal{C}_2) \subset \mathcal{B}$. Then note that $\mathcal{C}_1 \subset \mathcal{C}_2$ since $(a,b) = \bigcup_{n=n_0}^{\infty} [a + \frac{1}{n}, b - \frac{1}{n}] \in \sigma(\mathcal{C}_2)$.

□

2.2 The Monotone Class Theorem

Definition 2.2.1: Monotone Class

A *monotone class*, is a collection of subsets $\mathcal{M}$ of X such that:

1. If $A_i \uparrow A$ and each $A_i \in \mathcal{M}$, then $A \in \mathcal{M}$.
2. If $A_i \downarrow A$ and each $A_i \in \mathcal{M}$, then $A \in \mathcal{M}$.

Theorem 2.2.2: Monotone Class Theorem

Suppose $\mathcal{A}$ is an algebra, and $\mathcal{A}, \mathcal{M}$ are the smallest $\sigma-$algebra and monotone class containing $\mathcal{A}$ (respectively), then $\mathcal{M} = \mathcal{A}$.

Chapter 3

Measures

Definition 3.0.1: Measure

Let X be a set and $\mathcal{A}$ be a $\sigma-$algebra consisting of subsets of X, then a *measure* on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$, such that:

1. $\mu(\emptyset) = 0$.
2. If $A_i \in \mathcal{A}$ and $i = 1, 2, \ldots$ are pairwise (mutually) disjoint, then:

$$\mu(\bigcup_{n=0}^{\infty} A_i) = \sum_{n=0}^{\infty} \mu(A_i)$$

A set function P on a field $\mathcal{F}$ is a *probability measure* if it satisfies:

1. $0 \leq P(A) \leq 1$ for $A \in \mathcal{F}$
2. $P(\emptyset) = 0$, $P(\Omega) = 1$.
3. If $A_1, A_2, \ldots$ is a disjoint sequence of $\mathcal{F}$ sets and $\bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{F}$, then

$$P(\bigcup_{k=1}^{\infty})A_k = \sum_{k=1}^{\infty} P(A_k).$$

Remark.

> The second property is called *countable additivity*.

Remark.

> A *measure space* is the triple $(X, \mathcal{A}, \mu)$. A *probability measure space*, or simply a *probability space*, is the triple $(\Omega, \mathcal{F}, P)$.

Example.

> Let X be a set, $\mathcal{A}$ the collection of all subsets of X, and $\mu(A)$ be the number of elements in A. We call μ the *counting measure*.

Example.

> Let $X = \mathbb{R}$, $\mathcal{A}$ be the collection of all subsets of $\mathbb{R}$, $x_1, x_2, \cdots \in \mathbb{R}$ and $a_1, a_2, \cdots \geq 0$. Then, set
> $$\mu(A) = \sum_{i|x_i \in A} a_i.$$

Example.

> Let $\delta_x = 1$ if $x \in A$ and 0 otherwise. We call this measure the *point mass at x*.

Definition 3.0.2: Discrete Probability Space

Let $\mathcal{F}$ be the $\sigma-$field of all subsets of a countable space Ω, and let $p(\omega)$ be a non-negative function on Ω. Suppose that $\sum_{\omega \in \Omega} p(\omega) = 1$, and define $P(A) = \sum_{\omega \in A} p(\omega)$.

Suppose $A - \bigcup_{i=1}^{\infty} A_i$, where A_i are disjoint and $\omega_{i1}, \omega_{i2}, \cdots \in A_i$ Therefore:

$$P(A) = \sum_{ij} p(wij) = \sum_i \sum_j p(\omega_{ij}) = \sum_i P(A_i).$$

Let $A, B \in \mathcal{F}$ with $A \subset B$, then we have that $P(A) + P(A - B) = P(B)$, and thus:

$$P(A) \leq P(B)$$

as a special case of the above,

$$P(A^c) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap C)$$

If $A_k \in \mathcal{A}$ not necessarily disjoint, then *Boole's* inequality yields:

$$P(\bigcup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} P(A_k)$$

Proposition 3.0.3

The following statements hold:

1. If $A, B \in \mathcal{A}$ and $A \subset B$, then $\mu(A) \leq \mu(B)$.
2. If $A_i \in \mathcal{A}$ and $A = \bigcup_{i=1}^{\infty} A_i$, then $\mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i)$.
3. Suppose $A_i \in \mathcal{A}$ and $A_i \uparrow A$. Then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$
4. Suppose $A_i \in \mathcal{A}$ with $A_i \downarrow A$. If $\mu(A_1) < \infty$, then we have that $\mu(A) = \lim_{n\to\infty} (A_n)$.

Proof. :

1. Consider that $B = A \cup (B \backslash (A))$, and thus $\mu(B) = \mu(A \cup (B \backslash (A))) = \mu(A) + \mu(B \backslash (A)) \geq \mu(A)$.
2. Let $B_i = A_i \setminus (\bigcup_{j=1}^{i-1} A_i)$, then all B_i are pairwise disjoint and $\sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^{\infty} \mu(B_i)$. Therefore, $\mu(A) = \sum_{i=1}^{\infty} \mu(B_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$, where the last inequality holds because $B_i \subset A_i$ for any i.
3. With B_i defined as before, we get that:
$$\mu(A) = \mu(\bigcup_{i=1}^{\infty} A_i) = \mu(\bigcup_{i=1}^{\infty} B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \mu(\lim_{n\to\infty} \bigcup_{i=1}^{n} B_i) = \mu(\lim_{n\to\infty} \bigcup_{i=1}^{n} A_i).$$
4. Apply (3) to the sets $A_1 \setminus A_i$.

□

Example.

> To see why the $A_1 < \infty$ condition is necessary in 3.0.2-(4), consider $X = \mathbb{N}$, $A_i = \{i, i+1, \dots\}$ and μ the counting measure. Then $\mu(A_i) = \infty$ for all i, but $\mu(\bigcap A_i) = \mu(\emptyset) = 0$.

Definition 3.0.4: Finite measures

A measure, μ, is called a *finite measure* if $\mu(X) < \infty$. A measure is $\sigma-$*finite* if there exists sets $E_1, E_2, \dots \in \mathcal{A}$ such that $\mu(E_i) < \infty$ for all i and also $X = \bigcup_{i=1}^{\infty} E_i$.

If μ is a finite measure, then $(X, \mathcal{A}, \mu)$ is a finite measurable space, and if μ is $\sigma-$finite, then $(x, \mathcal{A}, \mu)$ is a$\sigma-$finite measurable space.

Definition 3.0.5: Null Sets, Complete Measure Sets, Completion

Let $(X, \mathcal{A}, \mu)$ be a measurable space. We say that $B \in \mathcal{A}$ is a *null set* if there exists an $A \subset B$ (A is not necessarily in $\mathcal{A}$) and $\mu(B) = 0$.

If $\mathcal{A}$ contains all the null sets, we say that $(X, \mathcal{A}, \mu)$ is a *complete measurable set.*

The *completion* of $\mathcal{A}$ is the smallest $\sigma-$algebra, $\overline{\mathcal{A}}$ containing $\mathcal{A}$, such that $(X, \overline{\mathcal{A}}, \overline{\mu}$ is a complete measurable space, where $\overline{\mu}$ is an extension of μ such that $\overline{\mu}(B) = \mu(B)$ if $B \in \mathcal{A}$.

Definition 3.0.6: Probability Measure

A *probability,* or a *probability measure,* is a measure μ such that $\mu(X) = 1$. We usually denote the probability measure space by $(\Omega, \mathcal{F}, \mu)$, where $\mathcal{F}$ is a $\sigma-$field (same thing as a $\sigma-$algebra).

Chapter 4

Construction of Measures

4.1 Outer Measures

Definition 4.1.1: Outer Measure

Let X be a set. An *outer measure* is a function μ^* defined on the collection of subsets of X satisfying:

1. $\mu^*(\emptyset) = 0$.
2. If $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$.
3. Whenever $A_i \subset X$, we have that $\mu^*(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$

A common way to generate outer measures is as follows:

Proposition 4.1.2

Suppose $\mathcal{C}$ is a collection of sets of X such that $\emptyset \in \mathcal{C}$ and $D_1, D_2, \cdots \in \mathcal{C}$ such that $\bigcup_{i=1}^{\infty} D_i = X$. Furthermore, suppose that $\ell : \mathcal{C} \to [0, \infty]$ and $\ell(\emptyset) = 0$. Define:

$$\mu^*(E) = \inf\{\sum_{i=1}^{\infty} \ell(A_i) | A_i \in \mathcal{C} \quad \text{and} \quad E \subset \bigcup_{i=1}^{\infty} A_i\}$$

then μ^* is an outer measure

Example.

> Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the collection of intervals of the form $(a, b]$. Let $\ell(I) = b - a$ if $i = (a, b]$. Define μ^* by the above proposition. If we restrict μ^* to a $\sigma-$algebra $\mathcal{L}$ which is smaller than the collection of all subsets of $\mathbb{R}$, then μ^* will be a measure on $\mathcal{L}$. Such measure is known as the *Lebesgue measure*, and such $\sigma-$algebra is the *Lebesgue $\sigma-$algebra.*

Example.

Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the collection of intervals of the form $(a, b]$. Let $\alpha : \mathbb{R} \to \mathbb{R}$ be such that $\lim_{y \to x^+} \alpha(y) = \alpha(x)$ (right continuous) and for $x < y$, we have that $\alpha(x) \leq \alpha(y)$ (increasing). Let $\ell(I) = \alpha(b) - \alpha(a)$ if $I = (a, b]$. Define $\mu*$ by the above, and so restricting μ^* to a smaller $\sigma-$algebra yields the *Lebesgue-Stieltjes measure* corresponding to α. A special case is when $\alpha(x) = x$ for all x (Lebesgue measure).

Definition 4.1.3: μ^* measurable

Let μ^* be an outer measure. A set $A \subset X$ is μ^*-*measurable* if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

for all $E \subset X$.

Theorem 4.1.4: Caratheodory Theorem

If v is an outer measure on X, then the collection $\mathcal{A}$ of μ^*-measurable sets is a $\sigma-$algebra. If μ is the restriction of μ^* to $\mathcal{A}$, then μ is a measure. Moreover, $\mathcal{A}$ contains all the null sets.

4.2 Lebesgue-Stieltjes measures

Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the collection of intervals of the form $(a, b]$. Let $\alpha : \mathbb{R} \to \mathbb{R}$ be such that $\lim_{y \to x^+} \alpha(y) = \alpha(x)$ (right continuous) and for $x < y$, we have that $\alpha(x) \leq \alpha(y)$ (increasing). Let $\ell(I) = \alpha(b) - \alpha(a)$ if $I = (a, b]$. Define

$$m^*(E) = \inf\{\sum_{i=1}^{\infty} \ell(A_i) | A_i \in \mathcal{C} \quad \text{and} \quad E \subset \bigcup_{i=1}^{\infty} A_i\}$$

By an Proposition 4.1.2, m^* is an outer measure. By Theorem 4.1.4, if m^* is a measure on the collection of m^*-measurable sets. Therefore, if $K = (a, b]$ and $L = (b, C]$, then $K \cup L = (a, c]$ and

$$\begin{aligned} \ell(K) + \ell(L) &= [\alpha(b) - \alpha(a)] + [\alpha(c) - \alpha(b)] && (4.1) \\ &= \alpha(c) - \alpha(a) = \ell(K \cup L). && (4.2) \end{aligned}$$

Lemma 4.2.1

Let $J_k = (a_k, b_k), k \in [n]$ be a finite collection of finite open intervals covering a finite closed interval $[C, D]$. Then

$$\sum_{k=1}^{n}[\alpha(b_k) - \alpha(a_k) \geq \alpha(D) - \alpha(C)]$$

Proof. Since $\{J_k\}$ is a cover of $[C, D]$, without loss of generality, there exists intervals $J_{k_1}, \ldots, J_{k_m}$ that cover $[C, D]$ such that $C \in J_{k_1}$ and $D \in J_{k_m}$. By construction:

$$a_{k_1} < C < b_{k_1}, \qquad a_{k_m} < D < b_{k_m}, \qquad a_{k_j} < b_{kj-1} < b_{k_j}(2 \leq j \leq m).$$

□

Thus,

$$\alpha(D) - \alpha(C) \leq \alpha(b_{k_m} - \alpha(a_{k_1}))$$

Proposition 4.2.2

Ife and f are finite and $I = (e, f]$, then $m^*(I) = \ell(I)$.

Proof. First, to show $m^*(I) \leq \ell(I)$, consider letting $A_1 = I$, $A_2 = A_2 = \cdots =$. Then $I \subset \bigcup_{i=1}^{\infty} A_i$ and so by definition

$$m^*(I) \leq \sum_{i=1}^{\infty} \ell(A_i) = \ell(A_1) = \ell(I)$$

The other direction involves creating a cover with C $\epsilon-$close to e (right continuous) and $f = D$ and then using the above Lemma to conclude that

$$\ell(I) \leq \alpha(D) - \alpha(C) + \frac{\epsilon}{2} \leq \sum_{k=1}^{n}(\alpha(d_{k'}) - \alpha(c_k)) + \frac{\epsilon}{2} \leq \sum_{i=1}^{\infty} \ell(A_i) + \epsilon.$$

Taking the infimum over all countable collections $\{A_i\}$, we obtain

$$\ell(I) \leq m^*(I) + \epsilon.$$

□

Proposition 4.2.3

Every set in the Borel $\sigma-$algebra on $\mathbb{R}$ is m^*-measurable.

Remark.

> Therefore, we can drop the asterisk from m^* because we are usually in the Borel $\sigma-$algebra and call m the *Lebesgue-Stieltjes measure*, and when $\alpha(x) = x$, m is simply the *Lebesgue measure*, where the collection of m measurable sets is a *Lebesgue* $\sigma-$*algebra.*

4.3 Examples and related results

Example.

> Let m be Lebesgue measure. If $x \in \mathbb{R}$, then $\{x\}$ is a closed set and hence is Borel measurable. Moreover
>
> $$m(\{x\}) = \lim_{n\to\infty} m((x - (\frac{1}{n}), x]) = \lim_{n\to\infty} [x - (x - \frac{1}{n}))] = 0.$$

Therefore,

$$m([a, b]) = m((a, b]) + m(\{a\}) = b - a + 0 = b - a$$

$$m((a, b)) = m((a, b]) - m(\{b\}) = b - a - 0 = b - a$$

Since $\sigma-$algebras are closed under countable unions, then countable sets are Borel measurable, and the Lebesgue measure of a countable set is 0.

Example.

> The Lebesgue measure of C, the cantor set, is 0, since the measure of F_n, that which remains after the nth cut, is $\frac{2}{3}^n$
>
> Instead, if $f_0 = \frac{1}{2}$ is defined on $(\frac{1}{3}, \frac{2}{3},)$ $f_0 = \frac{1}{4}$ on $(\frac{1}{9}, \frac{2}{9})$, $f_0 = \frac{3}{4}$ on $(\frac{7}{9}, \frac{8}{9})$, and we define
>
> $$f(x) = \inf\{f_0(y)|y \geq x, y \notin C\}$$
>
> Note that f is continuous, and is called the *Cantor function*, and f increases only on the Cantor set

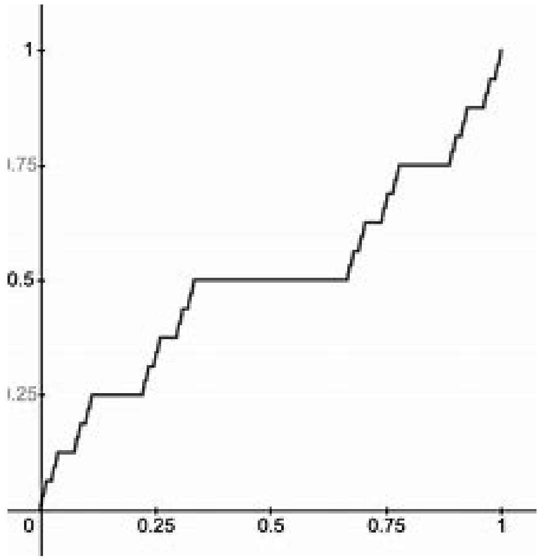

Figure 4.1: Cantor function

Proposition 4.3.1

Suppose $A \subset [0, 1]$ is a Lebesgue measurable set. Let m be Lebesgue measure:

1. For all $\epsilon > 0$, there exists an open set G such that $m(G - A) < \epsilon$ and $A \subset G$.
2. For all $\epsilon > 0$, there exists a closed set F such that $m(A - F) < \epsilon$ and $F \subset A$.
3. There exists a set H which contains A that is the countable intersection of a decreasing sequence of open sets and $m(H - A) = 0$.
4. There exists a set F which is contained in A that is the countable union of an increasing sequence of closed sets which is contained in A and $m(A - F) = 0$.

Corollary 4.3.2

Let μ be a Lebesgue-Stietljes measure on the real line. The conclusion of the above proposition holds with m replaced by μ.

4.4 Nonmeasurable Sets

Theorem 4.4.1: Nonmeasurable sets

Let

$$m^* = \inf\{\sum_{i=1}^{\infty} \ell(A_i) | A_i \in \mathcal{C} \quad \text{and} \quad E \subset \bigcup_{i=1}^{\infty} A_i\}$$

, where $\mathcal{C}$ is the collection of intervals that are open on the left and closed on the right and $\ell((a, b]) = b - a$. M^* is not a measure on the collection of all subsets of $\mathbb{R}$.

Proof. Suppose m^* is a measure, and define $x \sim y$ if $x - y$ is rational. For each equivalence class, pick an element out of that class and call the collection of these points A. Given a set B, define $B + x = \{y + x | y \in B\}$. Note that $\ell((a+q, b+q]) = b - a = \ell((a, b])$ for all a, bq, and thus $m^*(A+q) = m^*(A)$ for each A and q. Moreover, $A + q$ are disjoint for different rationals q. Now, consider that

$$[0,1] \subset \bigcup_{q \in [-1,1] \cap \mathbb{Q}} (A+q),$$

and thus $1 \leq \sum_{q \in [-1,1] \cap \mathbb{Q}} m^*(A+q)$, and thus $m * (A) > 0$. Consider now that

$$\bigcup_{q \in [-1,1] \cap \mathbb{Q}} (A+q) \subset [-1, 2],$$

and thus $3 \geq \sum_{q \in [-1,1] \cap \mathbb{Q}} m^*(A+q)$, implying that $m * (A) = 0$, a contradiction. □

4.5 The Caratheodory Extension Theorem

Let $\mathcal{A}_0$ be an algebra. Saying that ℓ is a measure means that: $\ell(\emptyset) = 0$ and if $A_i \in \mathcal{A}_0$ pairwise disjoint and their union is in $\mathcal{A}_0$, then (2) of the definition of a measure holds. A measure on an algebra is often referred to as a *premeasure*.

Theorem 4.5.1: The Caratheodory Extension Theorem

Suppose $\mathcal{A}_0$ is an algebra and $\ell : \mathcal{A}_0 \to [0, \infty]$ is a measure on $\mathcal{A}_0$. Define

$$\mu^*(E) = \inf\{\sum_{i=1}^{\infty} \ell(A_i) | \text{each} A_i \in \mathcal{A}_0, E \subset \bigcup_{i=1}^{\infty} A_i\}$$

for $E \subset X$. Then:

1. $\mu*$ is an outer measure;
2. $\mu * (A) = \ell(A)$ if $A \in (\mathcal{A}\prime)$;
3. Every set in $\mathcal{A}_0$ and every μ^*-null set is μ^*-measurable;
4. If ℓ is $\sigma-$finite, then there is a unique extension to $\sigma(\mathcal{A}_0)$.

Chapter 5

Measurable functions

5.1 Measurability

Suppose we have a measurable space $(X, \mathcal{A})$.

Definition 5.1.1: Measurable function

A function $f : X \to \mathbb{R}$ is *measurable* or $\mathcal{A}$ measurable if $\{x | f(x) > a\} \in \mathcal{A}$ for all $a \in \mathbb{R}$. A complex-valued function is measurable if both the real and imaginary are measurable.

The preimage of $(a, \infty]$ is measurable under f.

Remark.

> In other words, if $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable sets, $f : \Omega_1 \to \Omega_2$ is measurable if $f^{-1}(A_2) \in \mathcal{A}_1$ for all $A_2 \in \mathcal{A}_2$.

Example.

> Suppose f is a real valued function and identically constant, then the set $\{x | f(x) > \alpha\}$ is either empty or all of X, so f is measurable.

Example.

> Define
>
> $$f(x) := \begin{cases} 1, & x \in A \\ 0 & x \notin A \end{cases},$$
>
> then $\{x | f(x) > a\}$ is either $\emptyset, A, X$, and thus f is measurable if and only if $A \in \mathcal{A}$.

Example.

> Suppose $X \in \mathbb{R}$ with $\mathcal{B}$ measure, then $f(x) = x$ yields $\{x | f(x) > a\} = (a, \infty)$ is measurable.

Proposition 5.1.2

Suppose f is real valued. The following are equivalents:

1. $\{x|f(x) > a\}$ for all $a \in \mathbb{R}$.
2. $\{x|f(x) < a\}$ for all $a \in \mathbb{R}$.
3. $\{x|f(x) \leq a\}$ for all $a \in \mathbb{R}$.
4. $\{x|f(x) \geq a\}$ for all $a \in \mathbb{R}$.

Proof.

$$\{x|f(x) \leq a\} = \{x|f(x) > a\}^c$$

and thus if the latter is $\mathcal{A}$ measurable, then the former will be. Moreover, note that

$$\{x|f(x) \geq a\} = \bigcap_{i=1}^{\infty} \{x|f(x) > a - \frac{1}{n}\}$$

$\square$

Proposition 5.1.3

If X is a metric space, f is continuous, and $\mathcal{A}$ contains all the open sets, then f is measurable

Proof. $\{x|f(x) > a\} = (a, \infty)$ is open, and thus in $\mathcal{A}$. $\square$

Proposition 5.1.4

Let $c \in \mathbb{R}$. If f, g are measurable, then:

1. $f + g$;
2. $f - g$;
3. cf;
4. fg;
5. $f \wedge g$;
6. $f \vee g$,

are all measurable

Proof. :

1. Consider that if $f(x)+g(x) < a$, then there exists an $r \in \mathbb{Q}$ such that $f(x) < r < a+r$,

and thus:

$$\{x|f(x)+g(x)<r\} \qquad \bigcup_{r\in\mathbb{Q}}(\{x|f(x)<r\}\cap\{x|g(x)<a+r\}$$

and thus $f+g$ is measurable.

2. Bootleg using Prop 5.1.3.

3. If $c>0$, then $\{x|cf(x)<a\}=\{x|f(x)<\frac{a}{c}\}$, and thus cf is measurable. If $c=0$, use Example 1. If $c<0$, bootleg as before.

4. Notice first that
$$\{x|fg>a\}=\{x|\frac{1}{2}[(f+g)^2-f^2-g^2\}$$
and thus it suffices to show that f^2 is measurable if f is measurable. Consider that $\{x|f^2(x)>a\}=\{x|f(x)>\sqrt{a}\}\cup\{x|f(x)<-\sqrt{a}\}$.

5. Obvious

□

Proposition 5.1.5

If f_i is a measurable function for each i, then:

1. $\sup(f_i)$
2. $\inf(f_i)$
3. $\lim_{i\to\infty}\sup(f_i)$
4. $\lim_{i\to\infty}\inf(f_i)$

Proof. $\{x|\sup(f_i(x))>a\}=\bigcup_{i=1}^{\infty}\{x|f_i(x)>a\}$, and is therefore measurable. □

Definition 5.1.6: Almost Everywhere

We say $f=g$ *almost everywhere* written $f=g$ a.e. if $\{x|f(x)\neq g(x)\}$ has measure zero.
Similarly, we write that $f_i\to f$ a.e if the set of x where $f_i(x)$ does not converge to $f(x)$ has measure zero.

Remark.

> If X is a metric space and $\mathcal{B}$ is the Borel $\sigma-$algebra and $f:X\to\mathbb{R}$ is measurable with respect to $\mathcal{B}$, we say f is Borel measurable. Similarly, we would say something similar with respect to lebesgue measure.

Proposition 5.1.7

If $f : \mathbb{R} \to \mathbb{R}$ is monotone, then f is Borel measurable.

Proof. Suppose f is increasing. Let $a \in \mathbb{R}$ and define $x_0 = \sup\{y|f(y) \leq a\}$. If $f(x_0) \leq a$, then $\{x|f(x) > a\} = (a, \infty)$. If $f(x_0) > a$, then $\{x|f(x) > a\} = [x, \infty)$. Either case is Borel measurable. □

The following proposition states that if f is measurable, then the inverse of a borel set is measurable:

Proposition 5.1.8

Let $(X, \mathcal{A})$ be a measurable space and let $f : X \to \mathbb{R}$ be an $\mathcal{A}$ measurable function. If $A \in \mathcal{B}$ on $\mathbb{R}$, then $f^{-1}(A) \in \mathcal{A}$.

Proof. Let $\mathcal{B}$ be a Borel $\sigma-$algebra and $\mathcal{C} = \{A \subset \mathbb{R}|f^{-1}(A) \in \mathcal{A}\}$. Note that

$$f^{-1} \bigcup_{i=1}^{\infty} A = \bigcup_{i=1}^{\infty} f^{-1}(A) \in \mathcal{A},$$

and thus $\mathcal{C}$ is closed under unions (and by a similar proof, complements). Thus, $\mathcal{C}$ is a $\sigma-$algebra. Since f is measurable, $\mathcal{C}$ contains (a, ∞) for all a. Therefore, $\mathcal{C}$ contains $\mathcal{B}$. □

Example.

> Constructing a set that is Lebesgue measurable but not Borel measurable:
>
> Let f be the Cantor-Lebesgue function, and define
>
> $$F(x) = \inf\{y|f(y) \geq x\}$$
>
> Note that F is increasing (but not continuous), and is defined to go from $[0, 1]$ to C, the Cantor set. By Proposition 5.1.7, F^{-1} maps Borel sets back to Borel sets. Let m be the Lebesgue measure and A be the nonmeasurable set constructed at the end of the previous chapter. Let $B = F(A)$. Consider that because $F(A) \subset C$, and $m(C) = 0$, then B is a null set, and thus m measurable. However, B is not Borel measurable, for if it where, then by the previous reasoning, $F^{-1}(F(A)) = A$ would be Borel measurable.

5.2 Approximating Functions

Definition 5.2.1: Characteristic Functions

Let $(X, \mathcal{A})$ be a measurable space. If $E \in \mathcal{A}$, then the *characteristic function* of E is defined by

$$\chi_E(x) = \begin{cases} 1, x \in E \\ 0, x \notin E \end{cases}$$

Definition 5.2.2: Simple Function

A *simple function*, s, is a function of the form

$$s(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x)$$

for real numbers a_i and measurable sets E_i.

Proposition 5.2.3

Suppose f is a non-negative and measurable function. Then there exists a sequence of non-negative measurable simple functions, s_n, increasing to f.

Proof. Let $A_{in} = \{x | \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n}\}$ and

$$B_n = \{x | f(x) \geq n\}$$

. Then define

$$s_n = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{A_{in}} + n\chi_{B_n}$$

□

5.3 Lusin's Theorem

Recall that the *support* of a function f is the closure of the set $\{x | f(x) \neq 0\}$

Theorem 5.3.1: Lusin's Theorem

Suppose $f : [0, 1] \to \mathbb{R}$ is Lebesgue measurable, m is Lebesgue measure, and $\epsilon > 0$. There exists a closed set $F \subset [0, 1]$ such that $m([0, 1] - F) < \epsilon$ and the restriction of f to F is a continuous function on F.

Loosely, this means that every measurable function is almost continuous

Chapter 6

The Lebesgue Integral

Definition 6.0.1: The Lebesgue Integral

Let $(X, \mathcal{A}, \mu)$ be a measure space. If

$$s = \sum_{i=1}^{n} a_i \chi_{E_i}$$

is a non-negative measurable simple function, then define the *Lebesgue integral* of s to be:

$$\int s d\mu = \sum_{i=1}^{n} a_i \mu(E_i).$$

If $a_i = 0$ and $\mu(E_i) = \infty$, we the convention that $a_i \mu(E_i) = 0$. If $f \geq 0$ is a measurable function, then define

$$\int f d\mu = \sup \left\{ \int s d\mu : 0 \leq s \leq f, s simple \right\}$$

Let f me measurable and $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$. Provided neither of those are finite, define

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

Finally, if $f = u + iv$ is a complex value function and $\int (|u| + |v|) d\mu$ is finite, then define

$$\int f d\mu = \int u d\mu + i \int v d\mu$$

Definition 6.0.2: Integrable

If f is measurable and $\int |f| d\mu < \infty$, then f is *integrable*.

Proposition 6.0.3

1. If f is a real value measurable function with $0 \leq a \leq f(x) \leq b$ for all x and $\mu(X) < \infty$, then $a\mu(X) \leq \int f d\mu \leq b\mu(X)$.

2. If f and g are measurable, real valued, and integrable and $0 \leq f(x) \leq g(x)$ for all x, then $\int f d\mu \leq g d\mu$.

3. If f is a real valued non-negative, and integrable and c is a non-negative real number, then $\int c f d\mu = c \int f d\mu$.

4. If $\mu(A) = 0$ and f is non-negative and measurable, then $\int f \chi_A d\mu = 0$.

Chapter 7

Probability

Definition 7.0.1: Probability Space

A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is an arbitrary set, $\mathcal{F}$ is a $\sigma-$field of subsets of Ω, and $\mathbb{P}$ is a probability on $(\Omega, \mathbb{P})$.

Definition 7.0.2: Probability

A *probability*, or *probability measure*, is a positive measure whose total mass is 1, so that $\mathbb{P}(\Omega) = 1$.

Remark.

> Elements of $\mathbb{F}$ are called events, elements of Ω are denoted by ω.

Definition 7.0.3: Almost Surely

Instead of saying a property occurs almost everywhere, we talk about properties occurring *almost surely*, or *a.s.*

Definition 7.0.4: Random Variables

Real-valued measurable functions from Ω to $\mathbb{R}$ are *random variable* and usually denoted by X or Y and abbreviated by *r.v.*

Definition 7.0.5: Expectation

The Lebesgue integral of a random variable X with respect to a probability measure $\mathbb{P}$ is called the *expectation* or the *expected value* of X, and we write $\mathbb{E}[X]$ for $\int X d\mathbb{P}$. The notation $\mathbb{E}[X; A]$ is used for $\int_A X d\mathbb{P}$.

Definition 7.0.6: Indicator

The random variable 1_A is the function defined by

$$1_A = \begin{cases}1, & \omega \in \Omega 0, & \omega \notin \Omega\end{cases}$$

is defined to be the *indicator* of A.

Remark.

Events such as $\{\omega|X(\omega) > a\}$ are almost always abbreviated by $(X > a)$. Thus,

$$X > a, Y > b$$

refers to $\{w|X(\omega) > a \quad \text{and} \quad Y(\omega) > b\}$

Definition 7.0.7: $\sigma-$fields genererated by X.

Given a random variable X, the $\sigma-$field generated by X, denoted by $\sigma(X)$, is the collection of events $(X \in A)$, A being a Borel subset of $\mathbb{R}$. If we have several random variables: X_1,X$_2, \ldots, X_n, we write \sigma(X_1, \ldots, X_n)$ for the σ field generated by the collection of events $\{X_i \in A|A \in \mathcal{B}\}$

Remark.

Remember from before that the notation $X \in A$ is used to signify $\{\omega|X(\omega) \in A\}$.

Definition 7.0.8: Distribution

Given a random variable X, we can define a probability on $(\mathbb{R}, \mathcal{B})$, by

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A), \qquad A \in \mathcal{B}.$$

This probability $\mathbb{P}_X$ is known as the *law* or *distribution* of X.

Definition 7.0.9: Distribution function

Define a function $F_X : \mathbb{R} \to [0, 1]$ by

$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X \le x).$$

Such a function F_x is called the *distribution function* of X. Note that F_X is an increasing function whos corresponding Lebesgue-stietjes measure is $\mathbb{P}_x$.

Proposition 7.0.10

The distribution function F_X of a random variable satisfies:

1. F_X is increasing.
2. F_X is right continuous.
3. $$\lim_{x\to\infty} F_X(x) = 1 \qquad \lim_{x\to-\infty} F_X(x) = 0$$

Proposition 7.0.11

Suppose F is a distribution function. There exists a random variable X such that $F = F_X$.

Proof. Let $\Omega = [0,1]$ $\mathcal{F}$ be the Borel $\sigma-$field, and $\mathbb{P}$ the Lebesgue measure. Define $X(\omega) = \sup(y|f(y) < \omega)$ If $X(\omega) \leq x$, then $F(y) > \omega$ for all $y > x$. By right continuity, $F(x) \geq \omega$. On the other hand, if $\omega \leq F(x)$, then $x \notin \{y : F(y) < \omega\}$, and thus $X(\omega) \leq x$. Thus, $\{\omega|X(\omega) \leq x\} = \{\omega|0 \leq e\omega \leq F(x)\}$, and thus $F_x(x) = F(x)$. □

The following are examples of various distributions:

Example.

> Bernoulli: A random variable X is a Bernoulli random variable with parameter p is $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - 0$ for some $p \in [0,1]$.

Example.

> Binomial: A random variable X is a binomial random variable with parameters n and p if $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, where $n \in \mathbb{N}$ and $0 \leq k \leq n$ and $p \in [0,1]$.

Example.

> A random variable X is a geometry random variable with parameter p if $\mathbb{P}(X = k) = (1-p)p^k$, where $p \in \mathbb{Z}_+$

Example.

> Poisson: If $\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k \in \mathbb{Z}_+$ and $\lambda > 0$, then X is a Poisson random variable with parameter λ.

Definition 7.0.12: Density

If F is absolutely continuous, we call $f = F'$ the *density* of F.

Example.

> Exponential: Let $\lambda > 0$. For $x > 0$, let $f(x) = \lambda^{-\lambda x}$. If X has a distribution function who's density is equal to f, then X is an exponential random variable with parameter λ.

Example.

> Standard Normal: Define $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. If the distribution function of X has f as its density, then X is a standard normal variable, and so:
>
> $$\mathbb{P}(X \in A) = \mathbb{P}_X(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{\frac{-x^2}{2}}$$

Proposition 7.0.13

Suppose g is Borel measurable and g is either bounded or non-negative. Then

$$\mathbb{E}[g(X)] = \int g(x)\mathbb{P}_X(dx)$$

Proof. If g is the indicator function of an event A, this becomes the definition of $\mathbb{P}_X$. By linearity, the results holds for simple functions g. By approximating a non-negative measurable function from below by simple functions, the result holds for non-negative functions g and thus by linearlity, for bounded and measurable g. $\square$

Remark.

> If F_X has density f_x, then $\mathbb{P}_X(dx) = f_x(x)dx$. If X is integrable ($\mathbb{E}[X] < \infty$,) we have that
>
> $$\mathbb{E}[X] = \int x f_X(x)dx \qquad \mathbb{E}[X^2] = \int x^2 f_X(x)dx$$

Definition 7.0.14: Mean

We define the *mean* of a random variable to be its expectation

Definition 7.0.15: Variance

We define the *variance* of a random variable to be

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$$

Remark.

> $\mathrm{Var}(X + c) = \mathrm{Var}(X)$

Remark.

> $$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}[X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2.]$$
> Therefore, $\mathrm{Var}(X) \leq \mathbb{E}X^2$ and $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$ for any constant c.

Proposition 7.0.16

If $X \geq 0$ a.s. and $p > 0$, then

$$\mathbb{E}[X^p] = \int_0^\infty p\lambda^{p-1}\mathbb{P}(X > \lambda)d\lambda$$

Proof. Using Fubini's Theorem:

$$\int_0^\infty p\lambda^{p-1}\mathbb{P}(X > \lambda)d\lambda = \mathbb{E}\int_0^\infty p\lambda^{p-1}1_{(\lambda,\infty)}(X)d\lambda = \mathbb{E}\int_0^\infty p\lambda^{p-1}d\lambda = \mathbb{E}X^p$$

□

Proposition 7.0.17

Suppose g is convex and X and $g(X)$ are both integrable. Then $g(\mathbb{E}X) \leq \mathbb{E}(g(X))$.

7.1 Independence

Definition 7.1.1: Independent

Two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. The events $A_1, \ldots, A_n$ are linearly independent if

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_j})$$

whenever $1 \leq i_1 < \cdots < i_j \leq n$.

Example.

> Given $n = 3$, we have that $\mathbb{P}(A_1 \cap A_2 \cap A_3)$ must factor properly, but $\mathbb{P}(A_1 \cap A_2), \mathbb{P}(A_1 \cap A_3)$, and $\mathbb{P}(A_2 \cap A_3)$ must do the same.

Proposition 7.1.2

If A and B are independent, then A^C and B are independent

Proof.

$$\mathbb{P}(A^C \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^C)$$

□

Remark.

We say $\mathcal{F}$ and $\mathcal{G}$ are independent $\sigma-$fields if A and B are independent whenever $A \in \mathbb{F}$ and $B \in \mathbb{G}$.

We say X and Y are two independent random variables if $\sigma(X)$ and $\sigma(Y)$ are independent.

Remark.

Given an infinite sequence of events $\{A_n\}$, we say that they are independent if any finite subset of them is independent.

Remark.

If f and g are Borel functions and X and Y are independent, then $f(X)$ and $g(Y)$ are independent. This is because $\sigma(f(x)) \subset \sigma(X)$ and same for $\sigma(g(y)) \subset \sigma(Y)$.

Definition 7.1.3: Infinitely Often

If $\{A_n\}$ is a sequence of events, define A_n i.o (A_n *infinitely often*), by

$$(A_n \text{i.o.}) = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$$

Lemma 7.1.4: Borel-Cantelli Lemma

Let$\{A_n\}$ be a sequence of events.

1. If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{i.o.}) <= 0$.
2. If A_n are independent events and $\sum_n \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{i.o.}) = 1$.

Proof. :

1. $$\mathbb{P}(A_n \text{i.o.}) = \lim_{n\to\infty} \mathbb{P}(\bigcup_{i=n} {}_i) \leq \sum_{i=n}^{\infty} \mathbb{P}(A_i)$$

   which the RHS tends to zero as $n \to \infty$.

2. $$\mathbb{P}(\bigcup_{i=n} {}_i) = 1 - \mathbb{P}(\bigcap_{i=n} {}_i^C) = 1 - \prod_{i=n}^{N} \mathbb{P}(A_i^C) = 1 - \prod(1 - \mathbb{P}(A_i))$$

   and thus because $1 - e^{-x} \leq x$, we have that the RHS is greater than or equal to

   $$1 - e^{-\sum_{i=n}^{N} \mathbb{P}(A_i)},$$

which tends to 1 as $N \to \infty$.

$\square$

Theorem 7.1.5: Multiplication Theorem

If X, y and XY are integrable and X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Remark.

> If $X_1, \ldots, X_n$ are independent, then so are the random variable $X_1 -_1, \ldots, X_n - \mathbb{E}X_n$. Moreover, using the multiplication theorem to show that cross product terms are zero, we have that
>
> $$\mathbb{E}[(X_1 -_1) + \cdots + (X_n - \mathbb{E}X_n)]^2 = \mathbb{E}(X_1 -_1)^2 + \cdots + \mathbb{E}(X_n -_n)^2$$
>
> Therefore, if the X_i are independent, then the variance of the sum is equal to the sum of the variances.

7.2 Weak Law of Large Numbers

Definition 7.2.1: Independent and Identically Distributed

Given that X_n is a sequence of independent random variables, we have that $\mathbb{P}_{X_n} = \mathbb{P}_{X_1}$ (all r.v. have the same distribution). This situation is *i.i.d*

Remark.

> In i.i.d, we have that $\mathbb{P}(X_n \in A) = \mathbb{P}(X_1 \in A)$ and for all n and Borel sets A. Moreover, we have that $\mathbb{E}X_n = \mathbb{E}X_1$ and $\mathrm{Var}X_n = \mathrm{Var}X_1$.

Definition 7.2.2: Partial Sum Process

Define a *partial sum process* to be

$$S_n = \sum_{i=1}^{n} X_i$$

.

Remark.

> This means that $\frac{S_n}{n}$ is the average value of the first n of the X_i's.

Definition 7.2.3: Converges in probability

We say that a sequence of random variables $\{Y_n\}$ *converges in probability* to a random variable Y if it converges in measure with respect to the measure $\mathbb{P}$. This emans that for all $\epsilon > 0$, we have that, as $n \to \infty$,

$$\mathbb{P}(|Y_n - Y| > \epsilon) \to 0$$

Theorem 7.2.4: Weak Law of Large Numbers

Suppose the X_i are i.i.d. and $\mathbb{E}[X_1^2] < \infty$. Then $\frac{S_n}{n} \to \mathbb{E}[X_1]$ in probability.

Proof. Since X_i are i.i.d, then for n large enough, we have that $\mathbb{E}S_n = n\mathbb{E}X_1$. Therefore, $\mathbb{E}[(\frac{S_n}{n} - \mathbb{E}X_1)^2]$ is the variance of $\frac{S_n}{n}$. Let $\epsilon > 0$. We know by *Chebyshev's inequality* that

$$\mathbb{P}(|\frac{S_n}{n} - \mathbb{E}[X_1]| > \epsilon|) = \mathbb{P}((\frac{S_n}{n} - \mathbb{E}[X_1])^2 > \epsilon^2) \leq \frac{\operatorname{Var}(\frac{S_n}{n})}{\epsilon^2} = \frac{\sum_{i=1}^n \operatorname{Var}X_i}{n^2\epsilon^2} = \frac{n\operatorname{Var}X_1}{n^2\epsilon^2}$$

which the RHS converges to 0 as we let $n \to \infty$. $\square$

Example.

> If S_n is the sum of i.i.d. Bernoulli random variables $X_1, \ldots, X_n$, then S_n is the number of the X_i that are equal to 1. The probability that the first k are $X_i's$ and the rest are 0 is $p^k(1-p)^{n-k}$, and we get the same probability for any configuration of k ones. Therefore, $\mathbb{P}(S_n = k) = \binom{n}{k}p^k(1-p)^{n-k}$.

7.3 Strong Law of Large Numbers

Theorem 7.3.1: Strong Law of Large Numbers

Suppose $\{X_i\}$ is an i.i.d. sequence with $\mathbb{E}[|X_1|] < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n}{n} \to \mathbb{E}X_1, \qquad \text{a.s.}$$

The following Lemmas are useful in proving the SLoFG

Lemma 7.3.2

If $X \geq 0$ a.s. and $\mathbb{E}[X] < \infty$, then

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) < \infty$$

Proof. Since $\mathbb{P}(X \geq x)$ increases as x decreases, then

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) \leq \sum_{n=1}^{\infty} \int_{n-1}^{n} \mathbb{P}(X \geq x)dx = \int_{0}^{\infty} \mathbb{P}(X \geq x)dx = \mathbb{E}[X]$$

□

7.4 Conditional Expectation

Informally, one can think that if X is a random variable, then $\mathbb{E}[X]$ is the best guess for X given no information about the result of the trial. A conditional expectation can be thought of as the best guess given a little bit of information.

It is common in probability theory for there to be more than one $\sigma-$field present. For example, if $X_1, X_2 \ldots$ is a sequence of random variables, then one can define $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, and thus $\mathcal{F}$ is generated by the collection of sets $(X_i \in A)$ and $A \in \mathcal{B}$. Thus, we can think of $\mathbb{F}_n$ as the information that is contained in $X_1, X_2, \ldots, X_n$. Thus, it makes sense to write $\mathbb{E}[Y|X_1, \ldots, X_n]$ and $\mathbb{E}[Y|\mathcal{F}_n]$ and that $\mathcal{F}_0$ stands for no information.

Definition 7.4.1: Conditional Expectation

If $\mathcal{F} \subset \mathcal{G}$ are two $\sigma-$fields and X is an integrable $\mathcal{G}$ measurable random variable, then the *conditional expectation* of X given $\mathcal{F}$, written $\mathbb{E}[X|\mathcal{F}]$, is any $\mathcal{F}$ measurable random variable Y such that $\mathbb{E}[Y|A] = \mathbb{E}[X|A]$ for all $A \in \mathcal{F}$

Definition 7.4.2: Conditional Probability

The *conditional probability* of $A \in \mathcal{G}$ given $\mathbb{F}$ is defined by $\mathbb{P}(A|\mathcal{F}) = \mathbb{E}[1_A|\mathcal{F}]$. When $\mathcal{F} = \sigma(Y)$, one usually writes $\mathbb{E}[X|Y]$ for $\mathbb{E}[C|\mathcal{F}]$.

There are a couple properties this contains:

1. The best guess should just be the expected value if there is no new information, and thus $\mathbb{E}[X|\mathcal{F}_0] = \mathbb{E}[X]$.

Remark.

> If Y_1, Y_2 are two $\mathcal{F}$ measurable random variables such that $\mathbb{E}[Y_1|A] = \mathbb{E}[Y_2|A]$ for all $A \in \mathcal{F}$, then $Y_1 = Y_2$ a.s.

Remark.

> If X is already $\mathcal{F}$ measurable, then $\mathbb{E}[X|\mathcal{F}] = X$.

Example.

Suppose that X, Y have a joint density $f(x,y), \qquad 0 < x, y < \infty$ with marginal densities

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy, \qquad g(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Thus, the conditional density $f(y|x)$ is defined by

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

and thus

$$\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} yf(y|x)dy$$

. An interesting fact arises:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X]] &= \int_{-\infty}^{\infty} [Y = X = x]f(x)dx \\
&= \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} yf(y|x)dy]f(x)fx \\
&= \mathbb{E}[Y]
\end{aligned}$$

Proposition 7.4.3

If X is independent of $\mathcal{F}$, $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$.

Proof. If $A \in \mathcal{F}$, then 1_A and X are independent, and by the multiplication theorem:

$$\mathbb{E}[X|A] = \mathbb{E}[X1_A] = \mathbb{E}[X]\mathbb{E}[1_A] = \mathbb{E}[\mathbb{E}[X|A]].$$

□

Example.

Suppose $\{A_i\}$ is a collection of finite disjoint sets who's union is Ω, $\mathbb{P}(A_i) > 0$ for all i, and $\mathbb{F}$ is the $\sigma-$field generated by the $A_i's$. Then:

$$\mathbb{P}(A|\mathcal{F}) = \sum_i \frac{\mathbb{P}(A \cap A_i)}{\mathbb{P}(A_i)1_{A_i}}$$

This follows because

$$\mathbb{E}[\sum_i \frac{\mathbb{P}(A \cap A_i)}{\mathbb{P}(A_i)} 1_[A_i]|A_j] = \frac{\mathbb{P}(A \cap A_j)}{\mathbb{P}(A_j)} \mathbb{E}[1_{A_j}|A_j] = \mathbb{P}(A \cap A_j).$$

Example.

For a concrete example, suppose we toss a fair coin independently 5 times and let X_i be 1 or 0 depending whether the i^{th} toss was a heads or tails. Let A be the event that there were 5 heads and let $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$. Therefore, $\mathbb{P}(A) = \frac{1}{32}$ while

$$\mathbb{P}(A|\mathcal{F}_1) = \begin{cases} \frac{1}{16}, & X_1 = 1 \\ 0, & X_1 = 0 \end{cases}$$

Proposition 7.4.4

If $\mathcal{F} \subset \mathcal{G}$ and X is integrable and $\mathcal{G}$ measurable, then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[X]$$

Proof.

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\Omega] = \mathbb{E}[X|\Omega] = \mathbb{E}[X]$$

□

Remark.

From this we can derive a property that is sometimes used in the definition of conditional probability: For every $\mathcal{F}_n-$measurable event A, we have that

$$\mathbb{E}[\mathbb{E}[Y|\mathcal{F}_n]1_A] = \mathbb{E}[Y1_A]$$

Proposition 7.4.5

:

1. If $X \geq Y$ are both integrable, then

$$\mathbb{E}[X|\mathcal{F}] \geq \mathbb{E}[Y|\mathcal{F}], \qquad a.s.$$

2. If X and Y are integrable and $a \in \mathbb{R}$, then

$$\mathbb{E}[aX + y|\mathcal{F}] = a\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[Y|\mathcal{F}]$$

Proposition 7.4.6

(Jensen's Inequality for conditional expectations). If g is convex and X and $g(X)$ are integrable, then

$$\mathbb{E}[g(X)|\mathcal{F}] \geq g(\mathbb{E}[X|\mathcal{F}]), \qquad a.s.$$

Proposition 7.4.7

If X and XY are integrable and Y is $\mathcal{F}$ measurable, then

$$\mathbb{E}[XY|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}].$$

Proof. If $A \in \mathcal{F}$, then for any $B \in \mathcal{F}$,

$$\mathbb{E}[1_A\mathbb{E}[X|\mathcal{F}]; B] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]; A \cap B] = \mathbb{E}[X; A \cap B] = \mathbb{E}[1_A X; B].$$

Using the same technique of linearity and approximating the simple random variable to Y shows that the theorem is done. □

Proposition 7.4.8

If $\mathcal{E} \subset \mathcal{F} \subset \mathcal{G}$. then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{E}] = \mathbb{E}[X|\mathcal{E}] = \mathbb{E}[\mathbb{E}[X|\mathcal{E}]|\mathcal{F}]$$

Proof. The second equality holds because $\mathbb{E}[X|\mathcal{E}]$ is $\mathcal{E}$ and thus $\mathcal{F}$ measurable. For the first equality, let $A \in \mathcal{E}$. Thus, since $A \in \mathcal{F}$,

$$\mathbb{E}[\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{E}]; A] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]; A] = \mathbb{E}[X; A] = \mathbb{E}[\mathbb{E}[X|\mathcal{E}]; A].$$

□

Proposition 7.4.9

If X is integrable, then $\mathbb{E}[X|\mathcal{F}]$ exists.

Proof. Using Radon-Nikodym Theorem □

Proposition 7.4.10

(Tower Property): If $m < n$, then

$$\mathbb{E}[\mathbb{E}[Y|\mathcal{F}_n]|\mathcal{F}_m] = \mathbb{E}[Y|\mathcal{F}_m].$$

Example.

> Suppose $X_1, \ldots, X_n$ are independent random variables with $\mathbb{E}[X_j] = \mu$, for each j. Let $S_n = \sum_{j=0}^{n} X_j$ and let $\mathbb{F}_n$ be the information contained in $X_1, \ldots, X_n$ ($\sigma(X_1, \ldots, X_n)$.), then if $m < n$, we have that
>
> $$\begin{aligned} \mathbb{E}[S_n|\mathcal{F}_m] &= \mathbb{E}[S_m|\mathcal{F}_m] + \mathbb{E}[S_n - S_m|\mathcal{F}_m] \\ &= S_m + \mathbb{E}[S_n - S_m] \\ &= S_m + \mu(n - m) \end{aligned}$$

Example.

> Same example as before, but now suppose $\mu = 0$ and $\mathbb{E}[X_j^2] = \sigma^2$:
>
> $$\begin{aligned} E[S_n^2|\mathcal{F}_m] &= \mathbb{E}[(S_m + (S_n - S_m))^2|\mathcal{F}_m] \\ &= \mathbb{E}[S_m^2|\mathcal{F}_m] + 2\mathbb{E}[S_m(S_n - S_m)|\mathcal{F}_m] + \mathbb{E}[(S_n - S_m)^2|\mathcal{F}_m] \\ &= S_m^2 + \sigma^2(n - m) \end{aligned}$$

7.5 Martingales

Definition 7.5.1: Filtration

If $X_1, X_2, \ldots, X_n$ is a sequence of random variables, then the associated (discrete time) *filtration* is the collection $\{\mathcal{F}_n\}$ where $\mathcal{F}_n$ denotes the information in $X_1, \ldots, X_n$, i.e, the $\sigma-$algebra generated by said random variables.

Definition 7.5.2: Adapted

Let $\mathcal{F}$ be a $\sigma-$field and let $\{\mathcal{F}_n\}$ be an increasing sequence of $\sigma-$fields each of which is contained in $\mathcal{F}$. That is, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$ and $\mathcal{F}_n \subset \mathcal{F}$ for each n. A sequence of random variables M_n is *adapted* to $\{\mathcal{F}_n\}$ if for each n, M_n is $\mathcal{F}_n$ measurable.

Definition 7.5.3: Martingale

M_n is a *martingale* with respect to an increasing family of $\sigma-$fields $\{\mathcal{F}_n\}$ if:

1. M_n is adapted to $\mathcal{F}_n$;
2. Each M_n is integrable for each n;
3. $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n, \qquad a.s, \qquad n = 1, 2, \ldots$

Remark.

> It is usually the case that $\mathcal{F}_n = \sigma(M - 1, \ldots, M_n)$.

Remark.

> Note that it is easy to derive that for any martingale, we have that if $m < n$, we have that
>
> $$\mathbb{E}[M_n|\mathcal{F}_m] = M_m$$
>
> and
>
> $$\mathbb{E}[M_n - M_m|\mathcal{F}_m] = 0.$$
>
> Thus, we can say that if M_n are the winnings of a game, no matter what happens up to time m, the expected winnings in the next $n - m$ games is 0.

Proposition 7.5.4

If M_n is a martingale, then

$$\mathbb{E}[M_n] = \mathbb{E}[\mathbb{E}[M_n|\mathcal{F}_0]] = \mathbb{E}[M_0].$$

Proof. Proposition 7.4.4 □

Definition 7.5.5: Submartingale/Supermartingale

If X_n is a sequence of adapted integrable random variables with

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n, \qquad a.s, \qquad n = 1, 2, \ldots$$

, we call X_n a *submartingale.*

If instead we have

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n, \qquad a.s, \qquad n = 1, 2, \ldots$$

, we call X_n a *supermartingale.*

Example.

> If X_i is a sequence of mean zero integrable i.i.d. random variables and S_n is the partial sum process, then $M_n = S_n$ is a martingale, since (using independence and the fact that S_n is measurable with respect to $\mathcal{F}_n$)
>
> $$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n + \mathbb{E}[M_{n+1} - M_n|\mathcal{F}_n] = M_n + \mathbb{E}[M_{n+1} - M_n] = M_n + 0$$

Example.

> If the $X_i's$ have variance one and $M_n = S_n^2 - n$, then, using independence,
>
> $$\mathbb{E}[S_{n+1}^2|\mathcal{F}_n] = \mathbb{E}[(S_{n+1} - S_n)^2|\mathcal{F}_n] + 2s_n\mathbb{E}[S_{n+1}|\mathcal{F}_n] - S_n^2 = 1 + S_n^2$$

Example.

(Discrete Stochastic Integral) Suppose that $M_0, M_1, \ldots$ is a martingale with respect to $\mathcal{F}_n$. For $n \geq 1$, let $\Delta M_n = M_n - M_{n-1.}$ Let B_j denote the "bet" of the jth game. We allot negative value of B_j, which indicate betting that the price will go down or the game will be lost. Let W_n denote the winnings of this strategy: $W_0 = 0$ and for $n \geq 1$, we have that

$$W_n = \sum_{j=1}^{n} B_j[M_j - M_{j-1} = \sum_{j=1}^{n} B_j \Delta M_j$$

Let B_n be $\mathcal{F}_{n-1}$ measurable (we cannot see the result of the nth bet before betting.) The claim is that our winnings, W_n are a martingale:

$$\begin{aligned}
\mathbb{E}[W_{n+1}|\mathcal{F}_n] &= \mathbb{E}[W_n + B_{n+1}(M_{n+1} - M_n)|\mathcal{F}_n] \\
&= \mathbb{E}[W_n|\mathcal{F}_n] + \mathbb{E}[B_{n+1}(M_{n+1} - M_n)|\mathcal{F}_n] \\
&= W_n + B_{n+1}\mathbb{E}[M_{n+1} - M_n|\mathcal{F}_n] && (W_n \text{ is } \mathcal{F}_n \text{ measurable}) \\
&= W_n
\end{aligned}$$

Therefore, one cannot change a discrete time martingale to a game in one's favor with a betting strategy in a finite amount of time.

Example.

(Martingale Betting Strategy) Let $X_1, X_2, \dots$ be independent random variables with

$$\mathbb{P}[X_j = 1] = \mathbb{P}[X_j = -1] = \frac{1}{2}$$

Thus, let $M_0 = 0$, $M_n = X_1 + X_2 + \cdots + X_n$. Our betting strategy will be doubling our bet when we lose, and quitting when we win, thus assuring that once we win (which we will do almost surely), we will win \$1. The winnings in the game can be written as

$$W_n = \sum_{j=1}^{n} B_j \Delta M_j = \sum_{j=1}^{n} B_j X_j$$

where $B_1 = 1$ and for $j > 1$:

$$B_j = 2^{j-1} \qquad \text{if} X_1 = X_2 = \dots X_{j-1} = -1$$

and $B_j = 0$ else.

By the previous example, W_n is a martingale. In particular, for each n, $\mathbb{E}[W_n] = 0$, which we can check that noting that $W_n = 1$ unless $X_1 = \cdots = X_n = -1$, in which case $W_n = -1 - 2 - 2^2 - \cdots - 2^{n-1} = -[2^n - 1]$, which happens with probability $(\frac{1}{2})^n$. Thus,

$$\mathbb{E}[1 \cdot [1 - \frac{1}{2^n}]] - [2^n - 1] \cdot \frac{1}{2^n}$$

However, we know that $\lim_{n\to\infty} W_n = 1$, and thus $1 = \mathbb{E}[W_\infty] > \mathbb{E}[W_0] = 0$, and thus we have a submartingale (or a game in our favor).

Definition 7.5.6: Stopping Time

Suppose we have an increasing sequence of $\sigma-$fields $\{\mathcal{F}_n\}$ contained in a $\sigma-$field $\mathcal{F}$. Let $\mathcal{F}_\infty = \sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$. A random variable N (which is $\mathcal{F}$ measurable) from

$$\Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$$

is called a *stopping time* if for each finite n, we have that $(N \le n) \in \mathcal{F}_n$.

Remark.

Thus, this is useful in the betting strategy in which one bets 1 up to some time and then bets 0 afterwards. Let T be the stopping time for the strategy, then the winnings at time t is

$$M_{n\wedge T} = M_0 + \sum_{j=1}^{n} B_j \Delta M_j$$

where $B_j = 1$ if $j \le T$ and $B_j = 0$ if $j > t$.

Remark.

Intuition:

If $\mathcal{F}_n$ is what you know at time n, then at each time n you know whether to stop or not.

Proposition 7.5.7

:

1. Fixed times n are stopping times.
2. If N_1 and N_2 are stopping times, then so are $N_1 \wedge N_2$ and $N_1 \vee N_2$.
3. If N_n is an increasing sequence of stopping times, then so is $N = \sup(N_n)$, and similar for a decreasing sequence.
4. If N is a stopping time, then so is $N + n$.

y of

Proof. Proof for 2):

is a stopping time. In other words, N is the first time that one of the X_n is in the set A. To show that N is a stopping time, we write

$$(N_1 \wedge N_2 \leq n) = (N_1 \leq n) \cup (N_2 \leq n) \qquad (N_1 \vee N_2 \leq n) = (N_1 \leq n) \cap (N_2 \leq n)$$

Proof for 3)

$$(N \leq n) = \bigcup_{k=1}^{n} (X_k \in A).$$

$$\sup(N_i \leq n) = \cap_i (N_1 \leq n) \in \mathcal{F}$$

□

For a more concrete example, consider some

$$L = \max\{k \leq 9 | X_k \in A\} \wedge 9,$$

which is the last time X_k is in A up to time 9, and $\mathcal{F}_n = \sigma(X_1 m \ldots, X_n)$. It can be shown that L is not a stopping time, as one cannot know whether $L \leq 2$ without looking into the future at $X_3, \ldots, X_9$.

Theorem 7.5.8: Optional Stopping/Sampling Theorem I

Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$, then $Y_n = M_{n \wedge T}$ is a martingale. In particular, for each n, we have that

$$\mathbb{E}[M + n \wedge T = \mathbb{E}[M_0]]$$

If T is bounded, that is, there exists some $k < \infty$ such that $\mathbb{P}(T \leq k) = 1$, then

$$\mathbb{E}[M_t] = \mathbb{E}[M_0].$$

Remark.

> The boundedness condition is necessary as shown by the Martingale Betting Strategy example

Theorem 7.5.9: Optional Stopping/Sampling Theorem II

Suppose T is a stopping time and M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Suppose that $\mathbb{P}(T < \infty) = 1$ and M_n is integrable and for each n, we have that

$$\lim_{n \to \infty} \mathbb{E}[|M_n| 1\{T > n\}] = 0.$$

Then,

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

.

Theorem 7.5.10: Optional Stopping/Sampling Theorem III

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

$$\mathbb{E}[|M_{n \wedge T}| \leq C],$$

then

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

Theorem 7.5.11: Doob's Optional Sampling/Stopping Theorem

Let $\{\mathcal{F}_n\}$ be an increasing family of $\sigma-$fields, each containing in a $\sigma-$field $\mathcal{F}$. Let M_n be a martingale with respect to $\{\mathcal{F}_n\}$ and let N be a stopping time bounded by a positive integer K. Then $\mathbb{E}M_n = \mathbb{E}M_K$.

Proof.

$$\mathbb{E}[M_n] = \sum_{k=0}^{K} \mathbb{E}[M_n; N = K] = \sum_{k=0}^{K} \mathbb{E}[M_k; N = K]$$

Moreover, we know that

$$\mathbb{E}[M_k; N = k] = \mathbb{E}[M_{k+1}; N = k] = \cdots = \mathbb{E}[M_K]; N = k$$

and thus

$$\mathbb{E}M_n = \sum_{k=0}^{K} \mathbb{E}[M_k; N = k] = \mathbb{E}[M_K] = \mathbb{E}[M_0]$$

□

Corollary 7.5.12

If N is bounded by K and M_n is a submartingale, then $\mathbb{E}M_n \leq \mathbb{E}M_k$.

Example.

(Gambler's Ruin for Random Walk) Let $X_1, X_2, \ldots,$ be independent coin-tosses and let $S_n = 1 + X_1 + \ldots X_n$. S_n is a *simple random walk* starting at 1. We have shown that S_n is a martingale. Let $K > 1$ be a positive integer and let T denote the first time n such that $S_n = 0$ or $S_n = K$, then $M_n = S_{n \wedge T}$ is a martingale. We can apply the optional sampling theorem and conclude that

$$1 = M_0 = \mathbb{E}[M_T] = 0 \cdot \mathbb{P}(M_T = 0) + K \cdot \mathbb{P}(M_T = K).$$

and thus by solving we get that

$$\mathbb{P}(M_T = K) = \frac{1}{K}$$

Example.

Let $S_n = X_1 + \cdots + X_n$ be a simple random walk starting at 0. We know that

$$M_n = S_n^2 - n$$

is a martingale, so let $J, K \in \mathbb{N}$ and let

$$T = \min\{n | S_n = -J \quad \text{or} \quad S_n = K\}.$$

Using the same process as the previous example, we get that

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_T] = [1 - \mathbb{P}S_T = K] \cdot (-J) + \mathbb{P}(S_T = K) \cdot K,$$

which, when solving, yields:

$$\mathbb{P}(S_T = K) = \frac{J}{J_K}$$

We can use OST III to conclude that

$$0 = \mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[S_T^2] - \mathbb{E}[T].$$

But we know that

$$\begin{aligned}
\mathbb{E}[S_T^2] &= J^2 \mathbb{P}(S_T = -J) + K^2 \mathbb{P}(S_T = K) \\
&= J^2 \frac{K}{J+K} + K^2 \frac{J}{J+K} \\
&= JK
\end{aligned}$$

And thus $\mathbb{E}[T] = \mathbb{E}[S_T^2] = JK$. Thus, the expected amount of time for the random walker starting at the origin to get a distance K from the origin is K^2.

For the next theorem, let $M_N^* = \max_{i \leq n M_i}$:

Theorem 7.5.13: Doob's Inequality

If M_n is a martingale or a positive submartingale, then

$$\mathbb{P}(M_n^* \geq a) \leq \frac{\mathbb{E}|M_n|}{a}.$$

Definition 7.5.14: Upcrossing

The number of *upcrossings* of an interval $[a, b]$ is the number of times a sequence of random variables crosses from below a to above b. Let

$$S_1 = \min\{k : X_k \leq a\} \qquad T_1 = \min\{x > S_1 : X_k \geq b\}$$

and

$$S_{i+1} = \min\{k > T_i : X_k \leq a\} \qquad T_{i+1} = \min\{k > S_{i+1} : X_k \geq b\}$$

Then the number of upcrossings U_n before time n is

$$U_n = \max\{j : T_j \leq n\}$$

Lemma 7.5.15: Upcrossing Lemma

If X_k is a submartingale, then we have that

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[X_n - a]^+}{b - a}.$$

Proof. Number of upcrossings of $[a, b]$ by X_k is the same as the number of upcrossings of $[0, b-a]$ by $Y_k = (X_k - a)^+$, where Y_k is still a submartingale. Fix n and define $Y_{n+1} = Y_n$, define S_i and T_i as above, and let

$$S_i' = S_i \wedge (n+1); \qquad T_i' = T_i \wedge (n+1)$$

Therefore, $T_{n+1}; = n + 1$ and we write

$$\mathbb{E}[Y_{n+1}] = \mathbb{E}Y_{S_1'} + \sum_{i=0}^{n+1} \mathbb{E}[Y_{T_i'} - T_{S_i'}] = \sum_{i=0}^{n+1} \mathbb{E}[Y_{S_{i+1}'} - T_{T_i'}].$$

Since Y_k is a submartingale, all the summands on the right third term are non-negative. For the $j-$th upcrossing, we have that $T_{t_j'} - Y_{S_j'} \geq b - a$, and thus,

$$\sum_{i=0}^{\infty} (Y_{T_i'} - Y_{S_i'}) \geq (b - a)$$

□

Theorem 7.5.16: Martingale Convergence Theorem

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

$$\lim_{n \to \infty} M_n = M_\infty$$

Proof. As Greg Lawler puts it, this proof uses a 'buy low, sell high' strategy:
Suppose $M_0, M_1, \dots$ is a martingale such that

$$\mathbb{E}[|M_n|] \leq C < \infty,$$

for all n, then suppose $a < b$ are real numbers. It will suffice to show that the martingale cannot fluctuate infinitely often below a and above b. Define a sequence of stopping times by

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

and for $j > 1$:

$$S_j = \min\{n > T_{j-1} | M_n \leq a\}$$

$$T_j = \min\{n > S_j | M_n \geq b\}$$

Consider the discrete stochastic integral

$$W_n = \sum_{k=0}^{n} B_k [M_k - M_{k-1}],$$

with $B_n = 0$ if $n = 1 < S_1$ and

$$B_n = 1 \qquad \text{if } S_j \leq n - 1 < T_j\$$$

$$B_n = 0 \qquad \text{if } T_j \leq n - 1 < S_{j+1}$$

Thus, we buy every time the price drops below a and hold on until it rises above b, which is when we sell. Define

$$U_n = j \qquad \text{if} \qquad T_j < n \leq T_{j+1}$$

as the number of *upcrossings* by time n (it denotes the number of times n we see a fluctuation). It is easy to see each upcrossing results in a profit of at least $b - a$. Thus,

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

where the last term represents the loss of holding the last price we've seen. By the Optional Stopping Theorem, we know that

$$\mathbb{E}[W_n] = \mathbb{E}[W_0] = 0,$$

and thus by Jensen's inequality and Lemma 21.27 on Bass, we have that

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[a - M_n]}{b - a} \leq \frac{|a| + \mathbb{E}[|M_n|]}{b - a} \leq \frac{|a| + C}{b - a}$$

Note that because this holds for all n, then we have that

$$\mathbb{E}[U_\infty] \leq \frac{|a| + C}{b - a} < \infty,$$

and thus by Fatou's Lemma, we have that $U_\infty < \infty$ a.s. $\square$

Definition 7.5.17: Markov Property

A discrete time process $Y_0, Y_1, Y_2, \ldots$ if called a *Markov* if for each n, the conditional distribution of

$$Y_{n+1}, Y_{n+2}, \ldots$$

given $Y_0, Y_1, \ldots, Y_n$ is the same as the conditional distribution given Y_n.

In other words, we don't give a shit about the past of the process, only about the current value of it.

Example.

Polya's Urn:

We have an urn with one green and one red ball. At each positive integer time, we randomly choose a ball, look at color, then place it back in along with another ball of the same color. If R_n, G_n denote the number of red and green balls in the urn after the draw at time n, then we have that

$$R_0 = G_0 = 1, \qquad R_n + G_n = n + 2$$

and we define our random variable to be

$$M_N = \frac{R_n}{R_n + G_n} = \frac{R_n}{n+2}$$

to be the fraction of red balls at this time. Let $\mathcal{F} = \sigma(M_1, \ldots, M_n)$. Thus, the probability of a ball being chosen at time n depends only on the fraction of red balls in the urn before choosing, not on the order of red and green balls were put in. This is a Markov Process Note that

$$\mathbb{P}[R_{n+1} = R_n + 1|\mathcal{F}_n] = 1 - \mathbb{P}[R_{n+1} = R_n|\mathcal{F}_n] = \mathbb{P}\{R_{n+1} = R_n + 1|M_n\} = \frac{R_n}{n+2} = M_n$$

We need to check that M_n is a martingale with respect to $\mathcal{F}_n$:

$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}[M_{n+1}|M_n] \\
&= M_n \frac{R_n + 1}{n+3} + [1 - M_n] \frac{R_n}{n+3} \\
&= \frac{R_n(R_n + 1)}{(n+2)(n+3)} + \frac{(n + 2 - R_n)R_n}{(n+2)(n+3)} \\
&= \frac{R_n(n+3)}{(n+2)(n+3)} \\
&= M_n
\end{aligned}$$

This martingale satisfies the martingale convergence theorem because by the optional stopping theorem, we have that

$$\mathbb{E}[|M_n|] = \mathbb{E}[M_n] = \mathbb{E}[M_0] = \frac{1}{2},$$

Thus, with probability one, we have that there exists some M_∞ such that

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

7.5.1 Integrals with Respect to Random Walk

Suppose $X_1, \ldots$ are i.i.d with mean zero and variance σ^2. For example:

Example.

Coin Tossing:

$$\mathbb{P}[X_j = 1] = \mathbb{P}[X_j = -1] = \frac{1}{2}$$

and thus $\sigma^2 = 1$

Example.

Normal increments where $X_j \sim N(0, \sigma^2)$

Definition 7.5.18: Predictability

Let $S_n = \sum_{i=0}^{n} X_n$ and let $\{\mathcal{F}_n\} = \sigma(X_1, \ldots, X_n)$. We say a sequence of random variable $J_1, J_2, \ldots$ is *predictable* (with respect to $\{\mathcal{F}_n\}$ if for each n, J_n is $\mathcal{F}_{n-1}-$measurable)

Definition 7.5.19: Integral with respect to random walk

Suppose $J_1, J_2, \ldots,$ is a predictable sequence with $\mathbb{E}[J_n^2] < \infty$ for each n, then the *integral of J_n with respect to S_n* is defined by

$$Z_n = \sum_{j=1}^{n} J_j X_j = \sum_{j=1}^{n} J_j \Delta S_j$$

Remark.

This integral satisfies the following three rules:

1. Martingale: The integral Z_n is a martingale with respect to $\{\mathcal{F}_n\}$
2. Linearity: If J_n, K_n are predictable sequences and $a, b \in \mathbb{R}$, then $aJ_n + bK_n$ is a predictable sequence and

$$\sum_{j=1}^{n} (aJ_j + bK_j)\Delta S_k = a \sum_{j=1}^{n} J_h \Delta S_j + b \sum_{j=1}^{n} K_j \Delta S_j$$

3. Variance:

$$\text{Var}\left[\sum_{j=1}^{n} J_j \Delta S_j\right] = \mathbb{E}\left[(\sum_{j=1}^{n} J_j \Delta S_j)^2\right] = \sigma^2 \sum_{j=1}^{n} \mathbb{E}[J_j^2]$$

7.6 Brownian Motion

7.6.1 Central Limit Theorem

Theorem 7.6.1: Central Limit Theorem

Suppose $X_1, X_2, \dots, X_n$ are i.i.d with mean μ and variance $\sigma^2 < \infty$ and $S_n = \sum_{i=0}^{n} X_i$, then if

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{\mu}},$$

and

$$\Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-} x^2/2dx$$

is the standard normal distribution function, then as $n \to \infty$, the distribution of Z_n approaches a standard normal distribution. In other words, if $a < b$, then

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

7.6.2 Multivariate Normal Distribution

Definition 7.6.2: Joint/Multivariate Normal/Gaussian Distribution

A finite sequence of random variables has a *joint normal distribution* if they are linear combinations of independent standard normal random variables. In other words, if there exist independent random variables $(Z_1, \dots, Z_m)$, each $N(0, 1)$ and constants M_j, a_{jk} such that for $j = 1, \dots, n$,

$$X_j = m_j + a_{j1}Z_1 + a_{j2}Z_2 + \cdots + a_{jm}Z_m$$

Remark.

$\mathbb{E}[X_j] = m_j$, and in the case of mean-zero joint normals, we can write the equation as $\mathbf{X} = A\mathbf{Z}$, where

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad \mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix}$$

where $A \in M_{n\times m}$ with entries a_{jk}. Note that

$$\mathbb{E}[X_j^2 = a_{j1}^2 + \cdots + a_{jm}^2]$$

7.6.3 Limits of random walks

Suppose $X_1, X_2, \ldots$ are independent random variables with $\mathbb{P}[X_j = 1] = \mathbb{P}[X_j = -1] = \frac{1}{2}$ and let $S_n = \sum_{j=0}^{n} X_j$ be the random walk with time increments $\Delta t = 1$ and space increments $\Delta x = 1$.

Suppose we now choose $\Delta t = \frac{1}{N}$, with N large, then at time $1 = N\Delta t$, the value of the process is

$$W_1^{(N)} = \Delta x(X_1 + \cdots + X_n).$$

However, it will be convenient to choose Δx such that $\mathrm{Var}[W_1^{(N)}] = 1$, and thus by independence,

$$\begin{aligned}
\mathrm{Var}[W_1^{(N)}] &= \mathrm{Var}[\Delta x(X_1 + \cdots + X_n)] \\
&= (\Delta x)^2[\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)] \\
&= (\Delta x)^2 N
\end{aligned}$$

Thus, we need

$$\Delta x = \sqrt{\frac{1}{N}} = \sqrt{\Delta t}$$

When N is large enough, then we know that the distribution of $\frac{S_n}{\sqrt{N}}$ is approximately $N(0,1)$.

7.6.4 Brownian Motion

Brownian motion, or the *Wiener process*, is a model of random continuous motion. Let $B_t = B(t)$ be the random variable of the value of the process at time t.

Definition 7.6.3: Stochastic Process

A collection of random variables indexed by time is called a *stochastic process*.

Remark.

We can view this process in two ways:

1. For each t, there exists a r.b. B_t and there are correlations between the values at different times.

2. The function $t \to B_t$ is a random variable who's value is a function.

Remark.

> The three major assumptions about B_t :
>
> 1. (Stationary Increments): If $s < t$, then the distribution of $B_t - B_s$ is the same as $B_{t-s} - B_0$.
> 2. (Independent Increments) If $s < t$, then the r.b. $B_t - B_s$ is independent of the values B_r for $r \leq s$.
> 3. (Continuous Paths) the function $t \to B_t$ is a continuous function.
>
> This is enough to show that the increments are normally distributed.

Definition 7.6.4: Brownian Motion

A stochastic process B_t is called a (*one dimension*) *Brownian Motion* with *drift* m and *variance* σ^2 starting at the origin if it satisfies:

1. $B_0 = 0$;
2. For $s < t$, the distribution of $B_t - B_s$ is normal with mean $m(t-s)$ and variance $\sigma^2(t-s)$;
3. If $s < t$, the random variable $B_t - B_s$ is independent if the values B_r for $r \leq s; 0$
4. The function $t \to B_t$ is a continuous function of t, almost surely.

Remark.

> If $m = 0$ and $\sigma^2 = 1$, then B_t is *standard Brownian Motion*

Proposition 7.6.5

If B_t is a standard Brownian motion and

$$Y_t = \sigma B_t + mt,$$

then Y_t is a Brownian motion with drift m and variance σ^2.

7.6.5 Properties of Brownian Motion

Consider B_t to be SBM starting at the origin, with origins, we understand that the process is continuous, but extremely rough. We can sample the values

$$B_0, B_{\Delta t}, B_{2\Delta t}, \ldots$$

and thus the increment $B_{(k+1)\Delta t} - B_{k\Delta t}$ is a normal random variable with mean 0 and variance Δt. We set

$$B_{(k+1)\Delta t} = B_{k\Delta t} + \sqrt{\Delta t} N_k,$$

where $N_0, N_1, \ldots,$ denote independent $N(0,1)$ r.v. We can write this as

$$\Delta B_{k\Delta t} = B_{(k+1)\Delta t} - B_{k\Delta t} = \sqrt{\Delta t} N_k$$

Consider that $|\Delta B_t| = B_{(k+1)\Delta t} - B_{k\Delta t} \approx \sqrt{\Delta t}$, thus,

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

and thus as $\Delta t \to 0$, the limit does not exist.

Theorem 7.6.6

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

Definition 7.6.7: Hölder Continuous

Given $\alpha > 0$, we say that a function $f[0,1] \to \mathbb{R}$ is *Hölder continuous or order* α if there exists a $C < \infty$ such that for all $0 \leq s,\, t \leq 1$, we have that

$$|f(s) - f(t)| \leq C|s-t|^{\alpha}$$

Remark.

> Consider that differentiable unctions are Hölder continuous of order 1 since
>
> $$|f(s) - f(t)| \approx |f'(t)||s-t|$$

Theorem 7.6.8: Hölder Continuity Property

With probability one, for all $\alpha < \frac{1}{2}$, B_t is Hölder continuous of order α but is not Hölder continuous of order $\frac{1}{2}$.

7.6.6 Brownian Motion as a Continuous Martingale

Suppose we have an increasing filtration $\{\mathcal{F}_t\}$ and integrable random variables M_t such M_t is adapted of $\{\mathcal{F}_t\}$. Moreover, we have that if $s < t$, then $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$. Thus, if B_t is standard Brownian motion and $s < t$, then

$$\mathbb{E}[B_t|\mathcal{F}_s] = \mathbb{E}[B_s|\mathcal{F}_s] + \mathbb{E}[B_t - B_s|\mathcal{F}_s] = B_s + \mathbb{E}[B_t - B_s] = B_s$$

Thus, we can adjust the definition of Brownian motion, with the third condition replaced by: If B_t is Brownian motion with respect to $\{\mathcal{F}_t\}$, if B_t is $\mathcal{F}_t$ measurable and B_t satisfy the conditions to be a Brownian motion with the third condition being replaced by:

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

i.e: there is nothing useful for predicting the future increments. Thus, B_t is a martingale with respect to $\{\mathcal{F}_t\}$.

Definition 7.6.9: Continuous Martingale

A martingale M_t is called a *continuous martingale* if with probability one the function $t \to M_t$ is a continuous function.

Example.

> SBM with zero drift is a continuous martingale.

Definition 7.6.10: Continuous Time Filtration

A *continuous time filtration* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of of sub $\sigma-$algebras $\{\mathcal{F}_t\}$ of $\mathcal{F}$ such that if $s < t$, then $\mathcal{F}_s \subset \mathcal{F}$.

Definition 7.6.11: Right Continuity

For each t, we have that

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$$

Definition 7.6.12: Strong Completeness

We assume that $\mathcal{F}_t$ contains all the null sets of $\mathcal{F}$.

7.6.7 Brownian Motion as a Markov Process

A continuous time process X_t is a *Markov* if for every t, the conditional distribution of $\{X_s | s \geq t\}$ given $\{X_r | r \leq t\}$ is the same as the conditional distribution given X_t. I.e, the future of the process is conditionally independent of the past given the present value.

If B_t is a Brownian motion with parameters (m, σ^2) and

$$Y_s = B_{t+s}, \qquad 0 \leq s < \infty$$

then the conditional distribution of Y_s given $\mathcal{F}_t$ is that of a Brownian motion with initial condition $Y_0 = B_t$.

7.6.8 Brownian Motion as a Self-Similar Process

Theorem 7.6.13: Scaling Brownian Motion

Suppose B_t is a standard Brownian motion and $a > 0$. Let

$$Y_t = \frac{B_{at}}{\sqrt{a}}.$$

Then Y_t is a standard Brownian motion.

7.6.9 Computations for Brownian Motion

Assume B_t is a standard Brownian motion starting at the origin with respect to a filtration $\{\mathcal{F}_t\}$.

$$\begin{aligned}
\mathbb{E}[|B_t|] &= \mathbb{E}[\sqrt{t}|B_1|] \\
&= \frac{\sqrt{t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x| e^{\frac{-x^2}{2}} dx \\
&= \sqrt{\frac{2t}{\pi}} \int_0^{\infty} x e^{\frac{-x^2}{2}} dx \\
&= \sqrt{\frac{2t}{\pi}}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\{B_t \geq r\} &= \mathbb{P}\{\sqrt{t} B_1 \geq r\} \\
&= \mathbb{P}\{B_1 \geq \frac{r}{\sqrt{t}}\} \\
&= 1 - \Phi(\frac{r}{\sqrt{t}}) \\
&= \int_{\frac{r}{\sqrt{t}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx
\end{aligned}$$

Theorem 7.6.14: Strong Markov Property

If T is a stopping time with $\mathbb{P}\{T < \infty\} = 1$ and

$$Y_t = B_{T+t} - B_t$$

, then Y_t is a standard Brownian motion. Moreover, Y is independent of $\{B_t | 0 \leq t \leq T\}$.

Proposition 7.6.15

(Reflection Principle) If B_t is a standard Brownian motion with $B_0 = 0$, then for every $a > 0$,

$$\mathbb{P}\{\max_{0\leq s\leq t}\} = 2\mathbb{P}\{B_t > a\} = 2[1 - \Phi(\frac{a}{\sqrt{t})]}$$

Proof. Let

$$T_a = \min\{s|B_s = a\}$$

, then

$$\mathbb{P}\{\max_{0\leq s\leq t} B_s \geq a\} = \mathbb{P}\{T_a \leq t\} = \mathbb{P}\{T_a < t\}$$

Since $B_{t_a} = a$, we know that

$$\mathbb{P}\{B_t > a\} = \mathbb{P}\{T_a < t, B_t > a\} = \mathbb{P}\{T_a < t\}\mathbb{P}\{B_t - B_{t_a} > 0|T_a < t\}$$

, where using the Strong Markov Property, we can say that

$$\mathbb{P}\{B_t - B_{t_a} > 0|T_a < t\} = \frac{1}{2}$$

, and thus the first equality is satisfied. The second follows because $\mathbb{P}\{B_t > a\} = \mathbb{P}\{B_1 > \frac{a}{\sqrt{t}}\} = 1 - \Phi(\frac{a}{\sqrt{t}})$ □

Example.

> Let $a > 0$ and let $T_a = \inf\{t|B_t = a\}$. The random variable T_a is called a *passage time.* To find the density of T_a, we consider its distribution function
>
> $$F(t) = \mathbb{P}\{T_a \leq t\} = \mathbb{P}\{\max_{0\leq s\leq t} B_s \geq a\} = 2\left[1 - \Phi(\frac{a}{\sqrt{t}})\right],$$
>
> where the density is therefore
>
> $$f(t) = F'(t) = -2\Phi'(\frac{a}{\sqrt{t}})(-\frac{a}{2t^{3/2}}) = \frac{a}{t^{3/2}\sqrt{2\pi}}e^{\frac{-a^2}{2t}}$$

Example.

$$q(r,t) = \mathbb{P}\{B_s = 0 \quad \text{for some} \quad r \leq s \leq t\}$$

The scaling rule for Brownian motion shows that $q(r,t) = q(1, \frac{t}{r})$ so it suffices to compute that $q(t) = q(1, 1+t)$. Let A be the event that $B_s = 0$ for some $1 \leq s \leq 1+t$, then the Markov property implies that:

$$\begin{aligned} q(t) &= \int_{-\infty}^{\infty} \mathbb{P}[A|B_1 = r] d\mathbb{P}\{B_1 = r\} \\ &= \int_{-\infty}^{\infty} \mathbb{P}[A|B_1 = r] \left[\frac{1}{\sqrt{2\pi} e^{-r^2/2}} dr\right] \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \mathbb{P}[A|B_1 = r] e^{-r^2/2} dr \end{aligned}$$

The reflection principle implies that:

$$\mathbb{P}[A|B_1 = r] = \mathbb{P}\{\min_{1 \leq \leq 1+t} B_s \leq 0 | B_1 = r\} = \mathbb{P}\{\max_{0 \leq s \leq t} B_s \geq r\} = 2\mathbb{P}\{B_t \geq r\} = 2[1 - \Phi(\frac{r}{\sqrt{t}})]$$

Thus, we get that

$$q(t) = \int_{-\infty}^{\infty} 2[1 - \Phi(\frac{r}{\sqrt{t}})] \frac{1}{\sqrt{2\pi}} e^{\frac{-r^2}{2} dr} = q(t) = 1 - \frac{2}{\pi} \arctan(\frac{1}{\sqrt{t}})$$

7.6.10 Quadratic Variation

Definition 7.6.16: Quadratic Variation

If X_t is a process, the *quadratic variation* is defined by

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

where the sum is over all j with $\frac{j}{n} \leq t$.

We can write $\langle X \rangle_t$ as

$$\langle X \rangle_t = \frac{1}{n} \sum_{j=1}^{n} Y_j,$$

where

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

Suppose $W_t =_t +mt$ where B_t is a SBM, then for some fixed t,

$$\sum \left[W(\frac{j}{n}) - W(\frac{j-1}{n})\right]^2 =$$

$$\sigma^2 \sum \left[B(\frac{j}{n}) - B(\frac{j-1}{n})\right]^2 + \frac{2\sigma m}{n} \sum \left[X(\frac{j}{n}) - X(\frac{j-1}{n})\right] + \sum \frac{m^2}{n^2}$$

As $n \to \infty$

$$\sigma^2 \sum \left[B(\frac{j}{n}) - B(\frac{j-1}{n}) \right]^2 \to \sigma^2 \langle B \rangle_t = \sigma^2 t$$

And the other terms go to 0. Thus,

Theorem 7.6.17: Quadratic Variation of Brownian Motion

If W_t is a Brownian motion with drift m and variance σ^2, then $\langle W \rangle_t = \sigma^2 t$

Theorem 7.6.18

Suppose B is a SBM, $t > 0$, and Π_n is a sequence of partitions of the form

$$0 = t_{0,n} < t_{1,n} < \cdots < t_{n,n} = t$$

with $||\Pi_n|| \to 0$, then

$$Q(t; \Pi_n) := \sum_{j=1}^{n} [B(t_j) - B(t_{j-1})]^2 \to t$$

in probability. Moreover, if

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

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

7.6.11 Multidimensional Brownian Motion

Definition 7.6.19: Multidimensional Brownian Motion

The $d-$dimensional process

$$B_t = (B_t^1, \ldots, B_t^d)$$

is called a *d-dimensional Brownian motion starting at the origin with drift* $\boldsymbol{m} = (m_1, \ldots, m_d) \in \mathbb{R}^d$ *and* $d \times d$ *covariance matrix* Γ *with respect to the filtration* $\{\mathcal{F}\}$ if each B_t is $\mathcal{F}_t-$ measurable and the following holds:

- $B_0 = 0$;
- If $s < t$, the distribution of $B_t - B_s$ is joint normal with mean $(t - s)\mathbf{m}$ and covariance matrix $(t - s)\Gamma$.
- If $s < t$, the random vector $B_t - B_s$ is independent of $\mathcal{F}_s$.
- With probability one, the function $t \to B_t$ is continuous.

7.6.12 The Heat Equation

Let $p_t(x)$ be the temperature at x at time t. If the heat particles are moving independently and randomly then we can assume they are in Brownian motion. If we also asume that

$$\int_R p_t(x)dx = 1,$$

then we can see $p_t(x)$ as the probability density for Brownian motion. Since $B_t \sim N(0,t)$, then we know that $p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-x^2}{2t}}$ An equation of t, x. If we are interested in the position at time $s+t$, then we can use the Strong Markov Property to consider the position at time s and then the interval to t, leading to the *Chapman-Kolmogorov equation*

$$p_{s+t}(x) = \int_{-\infty}^{\infty} p_s(y) p_t(x-y) dy$$

To understand the evolution of $p_t(x)$, first we will use a binomial approximation, where we view the Brownian motion satisfying

$$\mathbb{P}\{B_{t+\Delta t} = B_t + \Delta x\} = \mathbb{P}\{B_{t+\Delta t} = B_t - \Delta x\} = \frac{1}{2}$$

where $\Delta x = \sqrt{\Delta t}$.
To be at x at time $t + \Delta t$, one must be at $x \pm \Delta x$ at time t, yielding

$$p_{t+\Delta t}(x) \approx \frac{1}{2} p_t(x - \Delta x) + \frac{1}{2} p_t(x + \Delta x)$$

Implying that (using $\Delta t = (\Delta x)^2$)

$$\frac{p_{t+\Delta t}(x) - p_t(x)}{\Delta t} = \frac{p_t(x - \Delta x) + \frac{1}{2} p_t(x + \Delta x) - 2p_t(x)}{2(\Delta x)^2}$$

Recognizing that as $\Delta t \to 0$, we have that the left hand side is equal to $\partial_t p_t(x)$. For the RHS, we can write $f(x) = p_t(x)$, and expand about x :

$$f(x+\epsilon) = f(x) + f'(x)\epsilon + \frac{1}{2} f''(x)\epsilon^2 + o(\epsilon^2),$$

$$f(x-\epsilon) = f(x) - f'(x)\epsilon + \frac{1}{2} f''(x)\epsilon^2 + o(\epsilon^2),$$

where the *little o* notation denotes a term such that

$$\lim_{\epsilon \to 0} \frac{o(\epsilon^2)}{\epsilon^2} = 0.$$

Thus, by adding,

$$f(x+\epsilon) + f(x-\epsilon) - 2f(x) = f''(x)\epsilon^2 + o(\epsilon^2)$$

Thus the RHS results in $\partial_{xx} p_t(x)/2$. Thus, we have derived the heat equation

$$\partial_t p_t(x) = \frac{1}{2} \partial_{xx} p_t(x)$$

7.6.13 Expected Value at a Future Time

Suppose B_t is a Brownian motion with drift m and variance σ^2, and let f be a function on $\mathbb{R}$.

Example.

Consider B_t to be the price of a stock and f to be the worth of a call option at strike price S and time t, then:

$$f(x) = (x - S)_+ = \begin{cases} x - s & \text{ifx} \geq S \\ 0 & \text{ifx¡S} \end{cases}$$

Let $\phi(t, x)$ be the expected value of $f(B_t)$ given that $B_0 = x$. Thus,

$$\phi(t, x) = \mathbb{E}^x[f(B_t)] = \mathbb{E}[f(B_t)|B_0 = x]$$

Chapter 8

Stochastic Integration

8.1 Stochastic Calculus

Going back to good old calculus, consider some $f(t)$ denoting the position of a particle at time t, and we are given that

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

or, more commonly,

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

This is a simple diff eq, where the rate depends on both the time and position. Given some initial $f(0) = x_0$, the function is defined by

$$f(t) = x_0 + \int_0^t C(s, f(s))ds$$

Sometimes this integral is unsolvable, but we can approximate it using *Euler's Method,* with which one uses small increments Δt and writes

$$f((k+1)\Delta t) = f(k\Delta t) + \Delta t C(k\Delta t, f(k\Delta t)).$$

In Stochastic calculus, we add randomness to the mix, and thus,

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

Where B_t is a SBM. This is an example of a *stochastic differential equation,* which we can read as stating that at time t, X_t is evolving like a Brownian motion with drift $m(t, X_t)$ and variance $\sigma(t, X_t)^2$. To solve this, we must use *stochastic Euler method*, which is described by the formula

$$X((k+1)\Delta t) = X(k\Delta t) = X(k\Delta t) + \Delta t m(k\Delta t, X(k\Delta t)) + \sqrt{\Delta t}\sigma(k\Delta t, X(k\Delta t))N_k,$$

where N_k is a $N(0, 1)$ random variable. Really, this should be written as

$$X((k+1)\Delta t) = X(k\Delta t) = X(k\Delta t) + \Delta t m(k\Delta t, X(k\Delta t)) + \Delta B_t \sigma(k\Delta t, X(k\Delta t)),$$

but don't question Daddy Lawler. We say that X_t is a solution to the SFE if

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

The ds integral is chill as fuck, even though the integrand is random. The second integral is tough.

8.1.1 Stochastic Integral

Let B_t be a standard Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$, then define the process

$$X_t = \int_0^t A_s dB_s.$$

Z_t can be though of as a Brownian motion which at time s has variance A_s^2. We can think of A_s as the bet at time s, restricting our betting strategies to those that cannot look into the future.

Riemann Review

Let us define

$$\int_0^1 f(t)dt$$

Consider a partition $P = \{0 = t_0, \ldots, t_n = 1\}$, where if $i < j$, then $t_i < t_j$. Define

$$\int_0^1 f_n(t)dt = \sum_{j=1}^n f(s_j)(t_j - t_{j-1}),$$

where s_j is some point chosen in $[t_{j-1}, t_j]$. We call $f(s_j)$ a *step function.* Then the limit as the mesh size goes to zero yields

$$\lim_{n\to\infty} \int_0^1 f_n(t)dt = \int_0^1 f(t)dt.$$

One can prove that the integral is independent of the choice of s_j.

Fact 8.1.1

$\int_a^b f'(t) = f(b) - f(a).$

Integration of simple processes

The analogue of a step function for the stochastic integral is a simple process:

Definition 8.1.2: Simple Process

A process, A_t, is a *simple process* if there exist times

$$0 = t_0 < t_1 < \cdots < t_n < \infty,$$

and random variables Y_j, $j \in [n]$ that are $\mathcal{F}_{t_j}-$measurable such that

$$A_t = Y_j, t_{j \leq t < t_{j+1}},$$

where $t_{n+1} = \infty$.

Remark.

> Since Y_j is $\mathcal{F}_{t_j}-$measurable, then A_t is $\mathcal{F}_t-$measurable. We also assume Y_j is square integrable for all j.

Definition 8.1.3: The Stochastic Integral

If A_t is a simple process, then we define

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

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

$$Z_t = Z_{t_j} + Y_j[B_t - B_{t_j}] \qquad \text{if} t_j \leq t \leq t_{j+1},$$

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

Proposition 8.1.4

Suppose B_t is a standard Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$, and A_t, C_t are simple processes, then

- If a, b are constants, then $aA_t + bC_t$ is a simple process and
$$\int_0^t (aA_s + bC_s)dB_s = a\int_0^t A_s dB_s + b\int_0^t C_s dB_s$$
and if $0 < r < t$, then
$$\int_0^t A_s dB_s = \int_0^r A_s dB_s + \int_r^t A_s dB_s$$
- The process
$$Z_t = \int_0^t A_s dB_s$$
is a martingale with respect to $\{\mathcal{F}_t\}$
- Z_t is square integrable and
$$\mathrm{Var}[Z_t] = \mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2]ds$$

Proof. For the martingale, we need to show that

$$\mathbb{E}[Z_t|\mathcal{F}_s] = Z_s \qquad \text{if} s < t.$$

Let $t = t_j$ and $s = t_k$ for $j > k$. In this case,

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

and

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

Obviously, $\mathbb{E}[Z_s|\mathcal{F}_s] = Z_s$, and thus

$$\mathbb{E}[Z_t|\mathcal{F}_s] = Z_s + \sum_{i=k}^{j-1} \mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]|\mathcal{F}_s]$$

By the tower property, since $t_k < t_j$, then for $k \le i \le j-1$,

$$\mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]|\mathcal{F}_{t_i}]|\mathcal{F}_s].$$

However, we have that Y_i is $\mathcal{F}_{t_i}$ measurable and $B_{t_{i+1}} - B_{t_i}$ is independent of $\mathcal{F}_{t_i}$, and thus $\mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]|\mathcal{F}_s] = Y_i\mathbb{E}[B_{t_{i+1}} - B_{t_i}|\mathcal{F}_{t_i}] = Y_i\mathbb{E}[B_{t_{i+1}} - B_{t_i}] = 0$.

For the variance rule for $t = t_j$, we have that

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

we invoke a martingale property once again, for $i < k$,:

$$\mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]Y_k[B_{t_{k+1}} - B_{t_k}]] = \mathbb{E}[\mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]Y_k[B_{t_{k+1}} - B_{t_k}]|\mathcal{F}_{t_k}]]$$

using adaptability and independence (for the last term, we get)

$$\mathbb{E}[\mathbb{E}[Y_i[B_{t_{i+1}} - B_{t_i}]Y_k[B_{t_{k+1}} - B_{t_k}]|\mathcal{F}_{t_k}]] = 0$$

A similar argument is used for $i > k$, and we see that

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{j-1} \mathbb{E}[Y_i^2(B_{t_{i+1}} - B_{t_i})^2]$$

again, the first term is $\mathcal{F}_{t_i}-$measurable and $(B_{t_{i+1}} - B_{t_i})^2$ is independent of the same sigmalgebra. Thus, we get

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

and thus,

$$\mathbb{E}[Z_t^2] = \sum_{i=0}^{j-1} \mathbb{E}[Y_i^2](t_{i+1} - t_i) = \int_0^t \mathbb{E}[A_s^2]ds.$$

□

Integration of a Continuous Process

Lemma 8.1.5

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

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

and for all n, t, we have that

$$|A_t^{(n)}| \leq C.$$

Proof. For $t = 1$, let

$$A_t^{(n)} = A(j, n), \qquad \frac{j}{n} \leq t < \frac{j+1}{n}$$

where $A(0,n) = A_0$ and for $j \geq 1$, we have

$$A(j,n) = n \int_{(j-1)/n}^{j/n} A_s ds$$

By construction, $A_t^{(n)}$ are simple processes satisfying $|A_t|^{(n)} \leq C$ and

$$A_t^{(n)} \to A_t,$$

and thus by the bounded convergence theorem, we have

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

, implying that $\lim_{n\to\infty} \mathbb{E}[Y_n] = 0$. □

With this lemma, we can repeat the above proposition but this time with A_t, C_t being bounded, adapted processes and eventually just relax it down for them to be continuous, and even that is much more strict than what is needed.

Remark.

> Suppose T is a stopping time with respect to $\{\mathcal{F}_t\}$. Then if A_t is a continuous, adapted process and
>
> $$Z_t = \int_0^t A_s dB_s,$$
>
> then
>
> $$Z_{t\wedge T} = \int_0^{t\wedge T} A_s dB_s = \int_0^t A_{s,T} dB_s,$$
>
> where $A_{s,T}$ denotes the piecewise continuous process,
>
> $$A_{s,T} = \begin{cases} A_s, & s < T \\ 0, & s \geq T \end{cases}$$
>
> In other words, sopping a stochastic integral is the same as changing bets to zero.

Definition 8.1.6: Stochastic Differential

We write the *stochastic differential*

$$dX_t = A_t dB_t$$

to mean that X_t satisfies

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

Remark.

We can think of X_t as a process that at time t evolves like a Brownian motion with zero drift and variance A_t^2, and is well defined for any adapted, continuous process A_t and X_t is a continuous function of t.

Example.

Let's be dumb dumb and use normal calculus! Consider the integral $Z_t = \int_0^t B_s dB_s$. Although B_t is continuous, adapted, and

$$\int_0^t \mathbb{E}[B_s^2]ds = \int_0^t s ds = \frac{t^2}{2} < \infty$$

and thus Z_t is a square integrable martingale. Using Newtonian calculus, one hopes that

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

However, using the Martingale Optional Stopping Theorem, the left hand side yields that $Z_0 = \mathbb{E}[Z_0] = 0 = \mathbb{E}[Z_t]$, but $\mathbb{E}[\frac{B_t^2}{2}] = \frac{t}{2} \neq 0$.

Theorem 8.1.7: Quadratic Variation

If A_t is an adapted process with continuous or piecewise continuous paths and

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

then

$$\langle Z \rangle_t = \int_0^t A_s^2 ds.$$

Ito's Formula

Let's travel back in time, all the way back to normal calculus. Suppose f is a C^1 function. Then we can expand f with Taylor's help,

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

and write f as a telescoping sum

$$f(1) = f(0) + \sum_{j=1}^{n}[f(\frac{j}{n}) - f(\frac{j-1}{n}).]$$

The Taylor approximation then gives

$$f(\frac{j}{n}) - f(\frac{j-1}{n}) = f'(\frac{j-1}{n}\frac{1}{n}) + o(\frac{1}{n}).$$

and thus

$$f(1) = f(0) + \lim_{n\to\infty} \sum_{j=1}^{n} f'(\frac{j-1}{n})\frac{1}{n} + \lim_{n\to\infty} \sum_{j=1}^{n} o\frac{1}{n}$$

And thus

$$f(1) = f(0) + \int_0^1 f'(t)dt.$$

This parallels Ito's formula, but instead we consider both first and second derivatives:

Theorem 8.1.8: Ito's Formula (I)

Suppose f is a C^2 function and B_t is a standard Brownian motion. Then for every t,

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

or, alternatively,

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

Remark.

> We can think of $f(B_t)$ at time t as a Brownian motion with drift $\frac{1}{2}f''(B_t)$ and variance $f'(B_t)^2$.

Proof. "Proof:"
Assume that $t = 1$ and expand f into a second order Taylor approximation:

$$f(x+y) = f(x) + f'(x)y + \frac{1}{2}f''(x)y^2 + o(y^2).$$

We then write a telescoping sum

$$f(B_1) - f(B_0) = \sum_{j=1}^{n}[f(B_{\frac{j}{n}} - f(B_{\frac{j-1}{n}}))]$$

We can use Taylor Approximation

$$f(B_{\frac{j}{n}}) - f(B_{\frac{j-1}{n}}) = f'(B_{\frac{j-1}{n}})\Delta_{j,n} + \frac{1}{2}f''(B_{\frac{j-1}{n}})\Delta_{j,n}^2 + o(\Delta_{j,n}^2)$$

where we define

$$\Delta_{j,n} = B_{\frac{j}{n}} - B_{\frac{j-1}{n}}.$$

We thus get 3 limits:

$$\lim_{n\to\infty}\sum_{j=1}^{n} f'(B_{\frac{j-1}{n}})[B_{\frac{j}{n}} - B_{\frac{j-1}{n}}] = \int_0^1 f'(B_t)dB_t \tag{8.1}$$

$$\lim_{n\to\infty}\frac{1}{2}\sum_{j=1}^{n} f''(B_{\frac{j-1}{n}})[B_{\frac{j}{n}} - B_{\frac{j-1}{n}}]^2 \tag{8.2}$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} o([B_{\frac{j}{n}} - B_{\frac{j-1}{n}}]^2) \tag{8.3}$$

By quadratic variation, Brownian motion increments satisfy

$$[B_{\frac{j}{n}} - B_{\frac{j-1}{n}}] \approx \frac{1}{n},$$

and thus 8.3 fucks off. 8.1 is just a simple process approximation, and thus 8.1 is easy.

For 8.2, let's say f'' is constant b, then

$$8.2 = \frac{b}{2}\langle B\rangle_1 = \frac{b}{2}.$$

Thus, let $h(t) = f''(B_t)$, be a continuous function. For every $\epsilon > 0$, there exists a step function $h_\epsilon(t)$ such that $|h(t) - h_\epsilon(t)| < \epsilon$ for every t. Fix ϵ, and consider each interval on which h_ϵ is constant to notice that

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

Thus,

$$8.2 = \lim \epsilon \to 0\frac{1}{2}\int_0^1 h_\epsilon(t)dt = \frac{1}{2}\int_0^1 h(t)dt = \frac{1}{2}\int_0^1 f''(B_t)dt$$

$\square$

Example.

Let $f(x) = x^2$, thus $f'(x) = 2x$, and $f''(x) = 2$. Ito's formula gives,

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

and thus

$$\int_0^t B_s dB_s = \frac{1}{2}[B_t^2 - t].$$

Example.

Let $f(x) = e^{\sigma x}$, where $\sigma > 0$. Then $f'(x) = \sigma e^{\sigma x}$ and $f''(x) = \sigma^2 e^{\sigma x} = \sigma f(x)$. Then let $X_t = f(B_t) = e^{\sigma B_t}$. By Ito's formula,

$$dX_t = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt = \sigma X_t dB_t + \frac{\sigma^2}{2}X_t dt.$$

Remark.

This is an example of *geomtric Brownian motion*

More Ito's Fomula

Theorem 8.1.9: Ito's Formula (II)

Suppose $f(t,x)$ is a function that is C^1 in t and C^2 in x. If B_t is a standard Brownian motion, then

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

or

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

Proof. This derivation is very similar, except that we now have more terms in the Taylor approximation □

Example.

Let $f(t,x) = e^{at+bx}$, then

$$\partial_x f(t,x) = be^{at+bx}; \qquad \partial_t f(t,x) = ae^{at+bx}; \qquad \partial_{xx} f(t,x) = b^2 e^{at+bx}$$

and thus if we let $X_t = f(t, B_t) = e^{at+bB_t}$, then

$$\partial_x f(t,x) = bX_t; \qquad \partial_t f(t,x) = aX_t; \qquad \partial_{xx} f(t,x) = b^2 X_t$$

and thus

$$dX_t = bX_t dB_t + [aX_t + \frac{1}{2}b^2 X_t]dt = (a + \frac{b^2}{2})X_t dt + bX_t dB_t$$

Definition 8.1.10: Geometric Brownian Motion

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

$$dX_t = mX_t dt + \sigma X_t dB_t = X_t[mdt + \sigma dB_t]$$

where B_t is a standard Brownian motion

Thus, if B_t is a SBM, the above example shows that

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

is a solution to the above.

We can say that if one uses the same Brownian motion in both of the above equations, then one gets the same function since the latter is a strong solution to the former. To explain this, let Δt be small, and define $B_{k\Delta t}$ by $B_0 = 0$ and

$$B_{k\Delta t} = B_{(k-1)\Delta t} + \sqrt{\Delta t} N_k,$$

where N_i is a sequence of independent $N(0,1)$ r.v. Choose $X_0 = e^0 = 1$ and

$$X_{k\Delta t} = X_{(k-1)\Delta t}[m\Delta t + \sigma\sqrt{\Delta t} N_j]$$

thus, if Δt is small enough, this should be close to

$$Y_{k\Delta t} = \exp\{(m - \frac{\sigma^2}{2})(k\Delta t) + \sigma B_{k\Delta t}\}$$

Remark.

> If usual calc, we throw away terms smaller order than dt. In Stochastic calc, we can use the rules that
> $$(dB_t)^2 = dt, \qquad (dB_t)(dt) = 0, \qquad (dt^2) = 0$$
> With these rules, we can formally derive Ito (II) by looking at the fact that
> $$df(t, B_t) = \partial_t f(t, B_t)dt + \partial_x f(t, B_t)dB_t + \frac{1}{2}\partial_{xx} f(t, B_t)(dB_t)^2 + o(dt) + o((dt)(dB_t)) + o((dB_t)^2)$$

Suppose that X satisfies

$$dX_t = R_t dt + A_t dB_t \rightleftharpoons X_t = X_0 + \int_0^t R_s ds + \int_0^t A_s dB_s$$

and define the quadratic variationby by

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

and, similar to Brownian motion, the drift term does not contribute to the quadratic variation:

$$\langle X \rangle_t = \langle \int A dB \rangle_t = \int_0^t A_s^2 ds \leftrightarrow d\langle X \rangle_t = A_t^2 dt$$

Theorem 8.1.11: Ito's Formula (III)

Suppose X_t satisfies

$$dX_t = R_t dt + A_t dB_t,$$

where R_t and A_t are adapted processes with piece wise continuous paths, $f(t, x)$ is C^1 in t and C^2 in x. Then,

$$df(t, X_t) = \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial_{xx} f(t, X_t)d\langle X \rangle_t \tag{8.4}$$

$$= \left[\partial_t f(t, X_t) + R_t \partial_x f(t, X_t) + \frac{A_t^2}{2}\partial_{xx} f(t, X_t) \right] dt + A_t \partial_x f(t, X_t) dB_t \tag{8.5}$$

Example.

Suppose X is a geometric Brownian motion satisfying

$$dX_t = mX_t dt + \sigma X_t dB_t$$

Let $f(t,x) = e^{-t}x^3$, then

$$\partial_t f(t,x) = -e^{-t}x^3 = -f(t,x)$$

$$\partial_x f(t,x) = 3x^2 e^{-t} = \frac{3}{x} f(t,x)$$

$$\partial_{xx} f(t,x) = 6xe^{-t} = \frac{6}{x^2} f(t,x)$$

and

$$df(t,X_t) = \partial_t f(t,X_t)dt + \partial_x f(t,X_t)dX_t + \frac{1}{2}\partial_{xx} f(t,X_t)d\langle X\rangle_t \tag{8.6}$$

$$= \left[\partial_t f(t,X_t) + mX_t\partial_x f(t,X_t) + \frac{\sigma^2 X_t^2}{2}\partial_{xxf(t,X_t)}\right] dt + \sigma X_t \partial_x f(t,X_t)dB_t \tag{8.7}$$

$$= \left(-1 + 3m + \frac{6\sigma^2}{2}\right) f(t,X_t)dt + 2\sigma f(t,X_t)dB_t \tag{8.8}$$

$$\tag{8.9}$$

and thus

$$d[e^{-t}X_t^3] = 3e^{-t}X_t^3\left[(\frac{-1}{3} + m + \sigma^2)dt + \sigma dB_t\right]$$

Example.

The exponential SDE is

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

as we claim that the solution is

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

This is trivial, but let

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

and thus satisfies

$$dY_t = -\frac{A_t^2}{2}dt + A_t dB_t, \qquad d\langle Y\rangle_t = A_t^2 dt.$$

Thus, if $f(x) = x_0 e^x$, then $f(x) = f'(x) = f''(x)$ and

$$df(Y_t) = f'(Y_t)dY_t + \frac{1}{2}f''(Y_t)d\langle Y\rangle_t = A_t X_t dB_t$$

Theorem 8.1.12: Ito's Formula (III - Local Form)

Suppose X_t satisfies

$$dX_t = R_t dt + a_t dB_t$$

with $a < X_0 < b$ and $f(t,x)$ is C^1 in t and C^2 in x for $a < x < b$. Let $T = \in \{t : X_t = a \quad or \quad X_t = b\}$. Then if $t < T$,

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

or

$$df(t,X_t) = f(t,X_t)dt + f'(t,X_t)dX_t + \frac{1}{2} f''(t,X_t) d\langle X\rangle_t, \quad t < T$$

Example.

> Suppose B_t is a standard Brownian motion and $Y_t = \frac{t}{B_t^2}$. Let $T = \inf\{t : B_t = 0\}$. Then for $t < T$,
>
> $$dY_t = f(t,B_t)dt + f'(t,B_t)dB_t + \frac{1}{2} f''(t,B_t)dt = [B_t^{-2} + 3tB_t^{-4}]dt - 2tB_t^{-3}dB_t.$$

Diffusions

Definition 8.1.13: Diffusion

We say that X_t is a *diffusion* if it is a solution to an SDE of the form

$$dX_t = m(t,X_t)dt + \sigma(t,X_t)dB_t \tag{8.10}$$

where $m(t,x), \sigma(t,x)$ are functions. These processes are also known as Markov processes

Remark.

> A process if called *time homogeneous* if the functions do not depend on t,
>
> $$dX_t = m(X_t)dt + \sigma(X_t)dB_t$$

Remark.

> Simulations of diffusions can be done using the stochastic Euler rule:
>
> $$X_{t+\Delta t} = X_t + m(x,X_t)\Delta t + \sigma(t,X_t)\sqrt{\Delta t} N$$

Definition 8.1.14: Generator

We define the *generator of a time homogeneous Markov process* X_t, $L = L^0$, by

$$L(f, x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}$$

If m and σ are bounded continuous functions, and f is a C^2 function, then Ito gives:

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

In other words,

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

since the second RHS terms is a martingale with expectation zero, then expectation of RHS is

$$t\mathbb{E}[Y_t],$$

where

$$Y_t = \frac{1}{t}\int_0^t \left[m(s, X_s)f'(X_s) + \frac{\sigma^2(s, X_s)}{2}f''(X_s)\right]ds$$

implying that

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

Because Y_t are uniformly bounded, we can take the limit inside the expectation, and thus

$$Lf(x) = \lim_{t \downarrow 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t} = m(0, X_0)f'(x) + \frac{\sigma^2(0, X_0)}{2}f''(x)$$

Similarly, if we define

$$L^t f(x) = \lim_{s \downarrow 0} \frac{\mathbb{E}[f(X_{t+s}) - f(X_t)|X_t = x]}{s},$$

then by the same logic, we get

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

Covariation and the Product Rule

Suppose X_t and Y_t satisfy

$$dX_t = H_t dt + A_t dB_t, \qquad dY_t = K_t dt + C_t dB_t \tag{8.11}$$

where B_t is the same SBM in both equations. The covariation process is then defined by

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

Formally, we get that

$$\begin{aligned}[dX_t][dY_t] &= [H_t dt + A_t dB_t][K_t dt + C_t dB_t] \\ &= A_t C_t dt + O((dt)^2) + O((dt)(dB_t))\end{aligned}$$

and thus we get that

$$\langle X, Y\rangle_t = \int_0^t A_s C_s ds$$

or

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

In a similar method as formally deriving the calculus product rule, we get that

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

Theorem 8.1.15: Ito Product Rule

If X_t, Y_t satisfy (8.11), then

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

or in integral form,

$$X_t Y_t = X_0 Y_0 + \int_0^t [X_s K_s + Y_s H_s + A_s C_s] ds + \int_0^t [X_s C_s + Y_s A_s] dB_s$$

Example.

Suppose B_t is a standard Brownian motion and X_t is the geometric Brownian motion satisfying

$$dX_t = mX_t dt + \sigma X_t dB_t,$$

then we can say that

$$H_t = mX_t, \quad A_t = \sigma X_t$$

and if we set $Y_t = B_t$, then

$$K_t = 0, \quad C_t = 1$$

Thus,

$$\begin{aligned}d(B_t X_t) &= B_t dX_t + X_t + dB_t + d\langle B_t, X_t\rangle_t \\ &=\end{aligned}$$

Several Brownian Motions

Suppose that $B_t^1, \ldots, B_t^d$ are independent Brownian motions with respect to the filtration $\{\mathcal{F}_t\}$ and that $A_t^1, \ldots, A_t^d$ are adapted processes. We write

$$dX_t = H_t dt + \sum_{j=1}^{d} A_t^j dB_t^j$$

to mean

$$X_t = X_0 + \int_0^t H_s ds + \sum_{j=1}^{d} \int_0^t A_s^j dB_s^j$$

Remark.

> If
>
> $$Y_t = Y_0 + \int_0^t K_s ds + \sum_{j=1}^{d} \int_0^t C_s^j dB_s^j,$$
>
> then remembering the orthogonality rule,
>
> $$d\langle X, Y\rangle_t = \sum_{j=1}^{d} A_t^j C_t^j dt$$

Remark.

> Recall that if $X_t = (X_t^1, \ldots, X_t^n)$, then the gradient is defined by
>
> $$\nabla f(t, \mathbf{X}_t) \cdot d\mathbf{X}_t = \sum_{k=1}^{n} \partial_k f(t, \mathbf{X}_t) dX_t^k$$

Theorem 8.1.16: Super Ito Formula

Suppose $B_t^1, \ldots, B_t^d$ are independent standard Brownian motions, and $X_t^1, \ldots, X_t^n$ are processes satisfying

$$dX_t^k = H_t^k dt + \sum_{i=1}^{d} A_t^{i,k} dB_t^i.$$

Furthermore, suppose that $f(t, \mathbf{x}), t \geq 0$ and $\mathbf{x} \in \mathbb{R}^n$ is a function which is C^1 in t and C^2 in $\mathbf{x}$. Thus, if $\mathbf{X}_t = (X_t^1, \ldots, X_t^n)$, then

$$df(t, \mathbf{X}_t) = f(t, \mathbf{X}_t)dt + \nabla f(t, \mathbf{X}_t) \cdot d\mathbf{X}_t + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \partial_{jk} f(t, \mathbf{X}_t) d\langle X^j, X^k\rangle_t$$

.

A particular case of this theorem is the following:

Theorem 8.1.17

Suppose $B_t = (B_t^1, \ldots, B_t^d)$ is a SBM in $\mathbb{R}^d$. If $f : [0, \infty) \times \mathbb{R}^d$ is C^1 in t and C^2 in $\mathbf{x} \in \mathbb{R}^d$, then

$$df(t, B_t) = \nabla f(t, B_t) \cdot dB_t + \left[f(t, B_t) + \frac{1}{2} \nabla^2 f(t, B_t) \right] dt$$

where

Definition 8.1.18: Laplacian

We define the *Laplacian* of a function $f : \mathbb{R}^d \to \mathbb{R}$ to be

$$\nabla^2 f(\mathbf{x}) = \sum_{j=1}^{d} \partial_{jj} f(\mathbf{x})$$

Chapter 9

More Stochastic Calculus

9.1 Martingales and Local Martingales

In the last chapter, we said that if

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

where B_t is a standard Brownian motion and A_s is a continuous or piecewise adapted process and

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

for each t, then Z_t is a square integrable martingale. What happens if the inequality doesn't hold?

Example.

> Martingale betting strategy. Consider
>
> $$Z_t = \int_0^t A_s dB_s$$
>
> where A_s is a betting strategy that changes at times $t_0 < t_1 < t_2 < \cdots < 1$, where
>
> $$t_n = 1 - 2^{-n}$$
>
> Set $A_t = 1$ for $0 \leq t < \frac{1}{2}$, and thus, $Z_{1/2} = B_{1/2}$. Note that
>
> $$\mathbb{P}\{Z_{1/2} \geq 1\} = \mathbb{P}\{B_{1/2} \geq 1\} = \mathbb{P}\{B_1 \geq \sqrt{2}\} = 1 - \Phi(\sqrt{2}) =: q > 0$$
>
> If $Z_{1/2} \geq 1$, we set $A_t = 0$ for $\frac{1}{2} \leq t \leq 1$. Otherwise, let $x = 1 - Z_{1/2} > 0$, and define a by
>
> $$\mathbb{P}\{a[B_{3/4} - B_{1/2}] \geq x\} = q$$
>
> and set $A_t = a$ for $\frac{1}{2} \leq t < \frac{3}{4}$. We need to know only $Z_{1/2}$ to determine a and thus a is $\mathcal{F}_{1/2}-$measurable. If $a[B_{3/4} - B_{1/2}] \geq x$, then
>
> $$Z_{3/4} = \int_0^{3/4} A_s dB_s = Z_{1/2} + a[B_{3/4} - B_{1/2}] \geq 1.$$
>
> Thus,
>
> $$\mathbb{P}\{Z_{3/4} \geq 1 | Z_{1/2} < 1\} = 1$$
>
> and hence
>
> $$\mathbb{P}\{Z_{3/4} < 1\} = (1 - q)^2$$
>
> If $Z_{3/4} \geq 1$, then we set $A_t = 0$ for $\frac{3}{4} \leq t \leq 1$ and otherwise proceed as above.
> At each time t, adjust the bet so that
>
> $$\mathbb{P}\{Z_{t_{n+1}} \geq 1 | z_{t_n} < 1\} = 1$$
>
> and thus
>
> $$\mathbb{P}\{Z_{t_n} < 1\} \leq (1 - q)^n.$$
>
> Thus, with probability one, $Z_t \geq 1$, and thus $\mathbb{E}[Z_1] \geq 1$. Thus, Z_t is not a martingale.

```
    import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# Parameters
N = 10000000  # number of steps
dt = 1.0 / N
times = np.linspace(0, 1, N+1)
```

```
# Function to generate Brownian motion
def generate_brownian_motion(N, dt):
    dB = np.random.normal(0, np.sqrt(dt), N)
    B = np.cumsum(dB)
    B = np.insert(B, 0, 0)  # Insert B(0) = 0
    return B

# Initializing A_t and Z_t
A = np.ones(N+1)
Z = np.zeros(N+1)

# Define change points t_n
change_points = [1 - 2**(-n) for n in range(1, int(np.log2(N))+2)]
change_points_indices = [int(t * N) for t in change_points]

# Phi value
q = 1 - stats.norm.cdf(np.sqrt(2))

# Function to calculate the value of a
def calculate_a(x, q, delta_t):
    return x / stats.norm.ppf(1 - q) / np.sqrt(delta_t)

# Generate Brownian motion
B = generate_brownian_motion(N, dt)

# Set A_t for 0 <= t < 1/2
A[:int(0.5 * N)] = 1

# Simulate the process
for i in range(1, len(change_points)):
    tn_idx = change_points_indices[i-1]
    t_next_idx = change_points_indices[i]

    # Calculate Z at tn
    Z[tn_idx] = np.sum(A[:tn_idx] * np.diff(B[:tn_idx+1]))

    if Z[tn_idx] >= 1:
        A[tn_idx:t_next_idx] = 0
    else:
        x = 1 - Z[tn_idx]
        delta_t = times[t_next_idx] - times[tn_idx]
        a = calculate_a(x, q, delta_t)
        A[tn_idx:t_next_idx] = a
```

```
# Final Z calculation
for t in range(1, N+1):
    Z[t] = Z[t-1] + A[t-1] * (B[t] - B[t-1])

# Plotting the results
plt.figure(figsize=(10, 6))
plt.plot(times, Z, label='Z_t')
plt.axhline(y=1, color='r', linestyle='--', label='Threshold')
plt.xlabel('Time t')
plt.ylabel('Z_t')
plt.legend()
plt.title('Simulation of Z_t with the described betting strategy')
plt.show()
```

Gives:

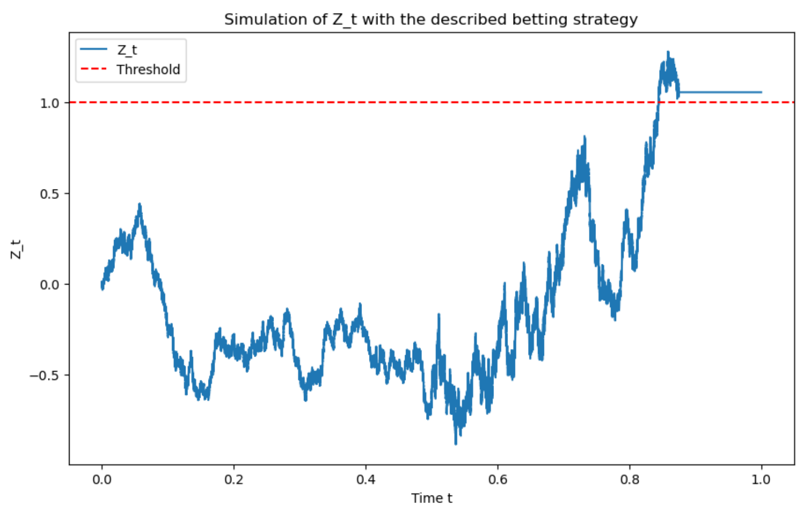

Figure 9.1: Martingale Betting Strategy

Definition 9.1.1: Local Martingale

A continuous process M_t is adapted to the filtration $\{\mathcal{F}_t\}$ is called a *local martingale* on $[0, T)$ if there exists an icnreasing sequence of stopping times

$$\tau_1 \leq \tau_2 \leq \cdots$$

such that with probability one,

$$\lim_{j \to \infty} \tau_j = T,$$

and for each j,

$$M_t^{(j)} = M_{t \wedge \tau_j},$$

is a martingale

Remark.

> In the case of the stochastic integral, we let $\{\tau_j\}$ be the stopping times defined by
>
> $$\tau_j = \inf\{t : \langle Z \rangle_t = \int_0^t A_s^2 ds = j\}.$$
>
> Thus, for each j, $M_t^{(j)}$ is a square integrate martingale, and so Z_t is a local martingale on $[0, T)$, where
>
> $$T = \inf\{t : \int_0^t A_s^2 = \infty\}$$

Remark.

> Suppose that
>
> $$dX_t = R_t dt + A_t dB_t,$$
>
> if $R_t \neq 0$, then X_t is not a martingale, and thus for X_t to be a martingale, we need $R_t \equiv 0$. It is a local martingale if $R_t \equiv 0$ but X_t is not a martingale. We usually refer to $A_t dB_t$ as the local martingale part.

Theorem 9.1.2: Optional Sampling Theorem

Suppose Z_t is a continuous martingale and T is a stopping time, both with respect to the filtration $\{\mathcal{F}_t\}$.

- If $M_t = Z_{t \wedge T}$, then M_t is a continuous martingale with respect to $\{\mathcal{F}_t\}$ and $\mathbb{E}[Z_{t \wedge T}] = \mathbb{E}[Z_0]$.
- Suppose there exists $C < \infty$ such that for all t, we have that $\mathbb{E}[Z_{t \wedge T}^2] \leq C$. Then if $\mathbb{P}\{T < \infty\} = 1$, $\mathbb{E}[Z_t] = \mathbb{E}[Z_0]$

Example.

> Suppose Z_t is a continuous martingale with $Z_0 = 0$. Suppose that $a, b > 0$ and let $T = \inf\{t | Z_t = -a \;\; or \;\; Z_t = b\}$. Suppose that $\mathbb{P}\{T < \infty\} = 1$, (which happens if Z_t is a standard Brownian motion. Then $Z_{t\wedge T}$ is abounded martingale and thus
>
> and thus
> $$0 = \mathbb{E}[Z_0] = \mathbb{E}[Z_t] = -a\mathbb{P}\{Z_T = a\} + b\mathbb{P}\{Z_t = b\},$$
> $$\mathbb{P}\{Z_t = b\} = \frac{a}{a+b}.$$
>
>)

Remark.

> The Martingale Convergence Theorem and the Maximal inequality also hold.

9.2 The Bessel Process

The Bessel process with parameter a is the solution to the SDE

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

Let $T_\epsilon = \inf\{t : X_t \leq \epsilon\}$, and note that there is no problem to find a solution for $t < T_\epsilon$, and thus it is well defined for $t < T$, where $T = T_0 = \inf\{t | X_t = 0\}$. Thus, there is a strong drift close to 0.

Suppose that $0 < r < R < \infty$, and let

$$\tau = \tau(r, R) = \inf\{t | X_t = r \;\; or \;\; X_t = R\}.$$

For $r \leq x \leq R$, let

$$\phi(x) = \mathbb{P}\{X_\tau = R | X_0 = x\},$$

and we'll use Ito's formula to compute ϕ. Note that $\phi(r) = 0, \phi(R) = 1$. Let J denote the indicator function of the event $\{X_\tau = R\}$ and let $M_t = \mathbb{E}[J|\mathcal{F}_t]$, we can use the tower property to show that M_t is a martingale, and by the Markovian nature of the diffusion of X,

$$\mathbb{E}[J|\mathcal{F}_t] = \phi(X_{t\wedge\tau})$$

Thus, if $\tau \leq t$, then we already know if $\{X_\tau = R\}$, but if $\tau > t$, then the only useful information for predicting if $X_\tau = R$ is X_t (Markov Property!). Thus, $\phi(X_{t\wedge\tau})$ must be a martingale. Ito's formula gives:

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

But as seen before, if it is a martingale, then the dt term must vanish. Thus, we must choose ϕ by solving the ODE

$$x\phi''(x) + 2a\phi'(x) = 0.$$

The solutions are of the form

$$\phi(x) = c_1 + c_2 x^{1-2a}, \qquad a \neq \frac{1}{2}$$

$$\phi(x) = c_1 + c_2 \ln(x), \qquad a = \frac{1}{2}$$

The boundary conditions, $\phi(r) = 0, \phi(R) = 1$, determine that

$$\phi(x) = \frac{x^{1-2a} - r^{1-2a}}{R^{1-2a} - r^{1-2a}}, \quad a \neq \frac{1}{2} \tag{9.1}$$

$$\phi(x) = \frac{\ln(x) - \ln(r)}{\ln(R) - \ln(r)}, \quad a = \frac{1}{2} \tag{9.2}$$

Proposition 9.2.1

If $a \geq \frac{1}{2}$, then $\mathbb{P}\{T = \infty\} = 1$, that is, with probability one, the Bessel process never reaches zero. If $a < \frac{1}{2}$, then $\mathbb{P}\{T < \infty\} = 1$.

Proof. Assume $X_0 = x < R$ and let $\tau(r, R)$ be defined as above. If $T < \infty$, then there must be some $R < \infty$ such that $X_{\tau(r,R)} = r$ for all $r > 0$. Using the above equations, we see that

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

□

9.2.1 Feynman-Kac Formula

Suppose that the price of a stock follows a geometric Brownian motion

$$dX_t = mX_t dt + \sigma X_t dB_t \tag{9.3}$$

Suppose further that at some future time T we have an option to buy a share of the stock at price S, but we'll only exercise the option if $X_t \geq S$ and the value of the option at time T is $F(X_T)$, where

$$F(x) = (x - S)_+ = \max\{x - S, 0\}$$

This makes sense, because if $X_T \leq S$, then $F(X_T) = 0$, and thus there is no payoff, otherwise, you can buy the stock and sell it immediately on the market to make stonks.

Furthermore, suppose that there is an inflation rate of r so that x dollars at time t in future is worth only $e^{-rt}x$ in current dollars. Let $\phi(t, x)$ be the expected value of this optino at time t, measured in dollars at time t, given that the current price of the stock is x,

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

The Feynman-Kac formula gives a PDE for this quantity.

Let's step it up, as it doesn't make a difference, and make the process a diffusion: Assume that X_t satisfies the SDE

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = x_0$$

and there is a payoff $F(X_T)$ at some future time T. Suppose that there is an inflation rate $r(t, x)$ so that if R_t denotes the value at time t of R_0 dollars at time 0, then

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

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

If $\phi(t, x)$ denotes the expected value of the payoff in time t dollars given $X_t = x$, then

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

Using Ito's formula to derive a PDE for ϕ (assuming sufficient smoothness in t and x), let

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

One can use the tower property to show that M_t is a martingale, and since R_t is $\mathcal{F}_t-$measurable,

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

But since X_t ia markov process, it only cares about the information given by X_t, and thus, by definition,

$$M_t = R_t^{-1}\phi(t, X_t) \tag{9.4}$$

and thus M_t is a martingale. Ito's formula gives that

$$d\phi(t, X_t) = \partial_t\phi(t, X_t) + \partial_x\phi(t, X_t)dX_t + \frac{1}{2}\partial_{xx}\phi(t, X_t)d\langle X\rangle_t.$$

In particular,

$$d\phi(t, X_t) = H_t dt + A_t dB_t$$

with

$$H_t = \partial_t\phi(t, X_t) + m(t, X_t)\partial_x\phi(t, X_t) + \frac{1}{2}\sigma(t, X_t)^2\partial_{xx}\phi(t, X_t)$$
$$A_t = \sigma(t, X_t)\partial_x\phi(t, X_t).$$

Since $\langle R\rangle_t = 0$, the stochastic product rule implies that

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

and thus the dt term is R_t^{-1} times

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

However, since M_t is a martingale, this only happens if the dt term is 0, and thus

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

Theorem 9.2.2: Feynman-Kac Formula

Suppose X_t satisfies

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = x_0$$

and $r(t, x) \geq 0$ is a discounting rate, and suppose that the payoff F at time T is given with $\mathbb{E}[|F(X_T)|] < \infty$. If $\phi(t, x), 0 \leq t \leq T$ is defined as

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

and $\phi(t, x)$ is C^1 in t and C^2 in x, then $\phi(t, x)$ satisfies the PDE

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

with $0 \leq t < T$, with terminal condition $\phi(T, x) = F(x)$.

Example.

Suppose X_t satisfies

$$dX_t = mX_tdt + \sigma X_tdB_t,$$

and $m(t, x) = mx, \sigma(t, x) = \sigma x$, and ϕ is defined the same as above. Thus, the Feynman-Kac formula gives

$$\partial_t\phi = r\phi - mx\partial_x\phi - \frac{\sigma^2x^2}{2}\partial_{xx}\phi \tag{9.5}$$

which is a version of the *Black-Scholes PDE.*

Remark.

Another derivation is given using the generator. Suppose that X_t satisfies

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

and that F is a function which doesn't grow too quickly. Let

$$f(t, x) = \mathbb{E}[F(X_T)|X_T = x].$$

Let $r(t) \geq 0$ be a discount rate and

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

Chapter 10

Change of Measure and Girasanov Theorem

10.1 Absolutely Continuous Measures

Definition 10.1.1: Absolutely and Mutually Continuous

Supposse μ, ν are measures on $(\Omega, \mathcal{F})$, then we say that

- μ is *absolutely continuous* with respect to μ, written $\nu \ll \mu$ if for every $E \in \mathcal{F}$, if $\mu(E) = 0$, then $\nu(E) = 0$.
- μ and ν are *mutually absolutely continuous* or *equivalent* measures if $\nu \ll \mu$ and $\mu \ll \nu$.
- μ and ν are *singular* measures, written $\mu \perp \nu$ if there exists an event $E \in \mathcal{F}$ with $\mu(E = 0$ and $\nu(\Omega \setminus E) = 0$

Example.

Let Ω be a countable set and $\mathcal{F} = 2^\Omega$, if $p : \Omega \to [0, \infty)$ is a function, then there exists a corresponding measure μ defined by

$$\mu(E) = \sum_{\omega \in E} p(\omega)$$

Suppose ν is another measure given by the function q, then let

$$A_\mu = \{\omega : p(\omega) > 0\}, \qquad A_\nu = \{\omega : q(\omega) > 0\}$$

Then $\nu \ll \mu$ if and only if $A_\nu \subset A_\mu$ and

$$q(\omega) = \frac{d\nu}{d\mu}(\omega)p(w)$$

where we define $d\nu/d\mu$ on A_μ by

$$\frac{d\nu}{d\mu}(\omega) = \frac{q(w)}{p(w)}$$

Thus, μ and ν are equivalent if $A_\nu = A_\mu$ and $\nu \perp \mu$ if $A_\nu \cap A_\mu = \emptyset$.

Theorem 10.1.2: Radon-Nikodym Theorem

Suppose μ and ν are $\sigma-$finite measures on $(\Omega, \mathcal{F})$ with $\nu \ll \mu$, that is,

$$\Omega = \bigcup_{n=1}^{\infty} A_n,$$

with $\mu(A_n), \nu(A_n) < \infty$ for each n. Then there exists a function f such that for every E,

$$\nu(E) = \int_E f d\mu$$

Remark.

The function f is called the *Radon-Nikodym derivative* of ν with respect to μ and is denoted

$$f = \frac{d\nu}{d\mu}.$$

Example.

If $(\Omega, \mathcal{F}, P)$ is a probability space, and Q is a probability measure with $Q \ll P$, then the Radon-Nikodym derivatve

$$X = \frac{dQ}{dP}$$

is a nonnegative random variable with $\mathbb{E}[X] = 1$ satisfying

$$Q(E) = \mathbb{E}_P[X 1_E]$$

and extending it,

$$\mathbb{E}_Q[Y] = \mathbb{E}_P[Y \frac{dQ}{dP}]$$

Example.

Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probabilty space and $\mathcal{G} \subset \mathcal{F}$. Let X be a nonnegative integrable $\mathcal{F}-$measurable random variable. Then

$$Q(A) = \mathbb{E}[X 1_A], \quad A \in \mathcal{G}$$

defined a measure on $(\Omega, \mathcal{G})$ that satisfied $Q \ll \mathbb{P}$. This, there exists a $\mathcal{G}-$measurable random variable Y such that for all $A \in \mathcal{G}$,

$$Q(A) = \mathbb{E}[1_A Y],$$

we can say that Y is the conditional expectation $E(X|\mathcal{G})$.

10.1.1 Giving Drift to a Brownian Motion

To take a fair game and make it unfair (or vice-versa):

- Add a deterministic amount in one direction.
- Change the probabilities of the outcome.

For the first method, define a Brownian motion with drift m by setting

$$W_t = mt + B_t$$

For the second way, suppose B_t is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\{\mathcal{F}_t\}$. To change the probability, we must consider a measure Q instead of $\mathbb{P}$, and so let

$$M_t = e^{mB_t - \frac{m^2 t}{2}}.$$

It can be derived that M_t is a martingale, and that M_t satisfies

$$dM_t = mM_t dB_t, M_0 = 1$$

If V is $\mathcal{F}_t$ measurable, then define

$$Q_t(V) = \mathbb{E}[1_V M_t]$$

and thus

$$\frac{dQ_t}{d\mathbb{P}} = M_t$$

it is easy to see that if $s < t$ and V is $\mathcal{F}_s-$measurable, then it is also $\mathcal{F}_t-$Measurable, that is, $Q_s(V) = Q_t(V)$:

$$\begin{aligned} Q_s(V) &= \mathbb{E}[1_V M_t] \\ &= \mathbb{E}[\mathbb{E}[1_V M_t|\mathcal{F}_s]] \\ &= \mathbb{E}[1_V M_s] \\ &= Q_s(V) \end{aligned}$$

Moreover, we claim that the process $t \to B_t$ under the measure Q is a Brownian motion with drift m and $\sigma^2 = 1$.

Example.

> Suppose X_t is a geometric Brownian motion satisfying
>
> $$dX_t = X_t[mdt + \sigma dB_t],$$
>
> where B_t is a standard Brownian motion defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P}.)$ If $r \in \mathbb{R}$, then we can find a new probability measure Q such that $dB_t = rdt + dW_t$, where W_t is a Brownian motion with respect to Q. Then
>
> $$dX_t = X_t[(m + \sigma r)dt + \sigma dW_t]$$
>
> Thus, with respect to Q, X_t is a geometric Brownian motion with the same volatility but new drift. Thus, measures for GBM with the same σ are equivalent.

10.2 Girsanov Theorem

Suppose M_t is a nonnegative martingale satisfying the exponential SDE

$$dM_t = A_t M_t dB_t, \quad M_0 = 1, \tag{10.1}$$

where B_t is SBM. The solution to this equation, derived before, is

$$M_t = e^{Y_t}, \qquad Y_t = \int_0^t A_s dB_s - \frac{1}{2}\int_0^t A_s^2 ds \tag{10.2}$$

Assume M_t is a martingale, and thus we can define a probability measure $\mathbb{P}^*$ such that is V is a $\mathcal{F}+_t$ measurable event, then

$$\mathbb{P}^*(V) = \mathbb{E}[1_V M_t.] \tag{10.3}$$

Thus,

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = M_t$$

If $s < t$ and V is $\mathcal{F}_s$ measurable, then V is also $\mathcal{F}_t-$measurable, and thus we need that for such V,

$$\mathbb{E}[1_V M_s] = \mathbb{E}[1_V M_t]$$

Let $\mathbb{E}^*$ denote expectations with respect to $\mathbb{P}^*$. If X is $\mathcal{F}_t-$measurable, then

$$\mathbb{E}^*[X] = \mathbb{E}[XM_t]$$

Theorem 10.2.1: Girsanov Theorem

Suppose M_t is a nonnegative martingale satisfying (10.1) and let $\mathbb{P}^*$ be the probability measure define in (10.3). If

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

then with respect to the measure $\mathbb{P}^*$, W_t is a standard Brownian motion. Thus

$$dB_t = A_t dt + dW_t,$$

where W is a $\mathbb{P}^*-$Brownian motion.

If we weight the probability measure $\mathbb{P}$ by the martingale, then in the new measure $\mathbb{P}^*$, the Brownian motion acquires a drift A_t.

To prove this, consider two lemmas:

Lemma 10.2.2: Lemma A

Let $0 \leq t \leq T$ and let Y be a $\mathcal{F}_t$ measurable random variable, then

$$\mathbb{E}^*[Y] = \mathbb{E}[YZ(t)]$$

The proof is definitions.

Lemma 10.2.3: Lemma B

Let $s \leq t$ such that $0 \leq s \leq t \leq T$ and Y be an $\mathcal{F}(t)-$measurable random variable, then

$$\mathbb{E}^*[Y|\mathcal{F}(s)] = \frac{1}{Z(s)} E[YZ(t)|\mathcal{F}(s)], \qquad a.s$$

Proof. Note that by definition, the RHS is $\mathcal{F}_s$ measurable.

$$\begin{aligned}
\int_A \frac{1}{Z(s)} E[YZ(t)|\mathcal{F}_s] d\mathbb{P}^* &= \int_\Omega 1_A \frac{1}{Z(s)} E[YZ(t)|\mathcal{F}_s] d\mathbb{P}^* \\
&= \mathbb{E}^*[1_A \frac{1}{Z(s)} E[YZ(t)|\mathcal{F}_s]] \\
&= \mathbb{E}[1_A E[YZ(t)|\mathcal{F}_s]] \\
&= \mathbb{E}[E[1_A YZ(t)|\mathcal{F}_s]] \\
&= \mathbb{E}[1_A YZ(t)] \\
&= \mathbb{E}^*[1_A Y] \\
&= \int_A Y d\mathbb{P}^*
\end{aligned}$$

where the second to last equality follows from Lemma A.

$$\int_A \mathbb{E}^*[Y|\mathcal{F}_s] d\mathbb{P}^* = \int_A Y d\mathbb{P}^*$$

for all $A \in \mathcal{F}_s$ □

Proof. By definition, $W(0) = 0$, and is continuous. We know that

$$dW_t = dW_t + A_t dt$$

and thus, by a formal derivation,

$$\langle W \rangle_t = t$$

We have showed via Ito's formula already that M_t is a martingale and $\mathbb{E}[M_t] = 1$. Because

$$M_t = \mathbb{E}[M_t|\mathcal{F}_t] = \mathbb{E}[M|\mathcal{F}_t],$$

then M_t is a Radon-Nikodyn process. We claim that $\{W_t M_t\}$ is martingale:

$$\begin{aligned}
d(W_t M_t) &= W_t dM_t + M_t dW_t + dW_t dM_t, \quad \text{(Ito's Product rule)} \quad = -W_t A_t M_t dW_t + Z_t(dW_t - A_t dt) \\
&= (-W_t - A_t + 1) M_t dW_t
\end{aligned}$$

and thus $W_t M_t$ is an Ito integral, and thus a martingale.
Thus, by Lemma B, if $s < t$:

$$\mathbb{E}^*[W_t|\mathcal{F}_s] = \frac{1}{M_s} \mathbb{E}[W_t M_t|\mathcal{F}_s] \tag{10.4}$$

$$= \frac{1}{Z(s)} W_s M_s \tag{10.5}$$

$$= W_s \tag{10.6}$$

and thus W_t is a martingale, and by Levy's theorem, W_t is a Brownian motion. □

Theorem 10.2.4: Girsanov Theorem, local martingale form

Suppose $M_t = e^{Y_t}$ satisfies (10.1) - (10.2), and let

$$T_n = \inf\{t | M_t + |A_t| = n\}, \qquad T = T_\infty = \lim_{n\to\infty} T_n.$$

Let $\mathbb{P}^*$ be the probability measure as above. If

$$W_t = B_t - \int_0^t A_s ds, \quad 0 \le t < T,$$

then with respect to the measure $\mathbb{P}^*$, W_t, $t < T$, is a standard Brownian motion. In other words,

$$dB_t = A_t dt + dW_t, \qquad t < T,$$

where W is a $\mathbb{P}^*-$Brownian motion. If any of the three following conditions holds, then M_s, $0 \le s \le t$, is actually a martingale:

$$\mathbb{P}^*\{T > t\} = 1$$

$$\mathbb{E}[M_t] = 1$$

$$\mathbb{E}[\exp\{\frac{\langle T\rangle_t}{2}\}] < \infty/$$

Remark.

> The last condition is called the novikov condition

10.3 The Black-Scholes Formula

Definition 10.3.1: Arbitrage

An *arbitrage* is a system that guarantees that a player will not lose money while also giving a positive probability of making money.

Remark.

> If P and Q are equivalent probability measures, then an arbitrage under probability P is the same as an arbitrage under probability Q, because for probability measures:
>
> $$P(V) = 0 \Leftrightarrow Q(V) = 0$$
>
> $$P(V) > 0 \Leftrightarrow Q(V) > 0$$

Consider a simple European call option for a stock whose price moves according to a geometric Brownian motion:

$$dS_t = S_t[mdt + \sigma dB_t]$$

and there exists a risk-free bound R_t such that

$$dR_t = rR_t dt$$

That is, $R_t = e^{rt}R_0$.

Let T be a time in the future and suppose we have the option to buy a share of stock at time T for strike price K. The value of this option at time T is:

$$F(S_T) = (S_T - K)_+ = \begin{cases} (S_T - K), & \text{if} S_T > K, \\ 0, & \text{if} S_T \leq K. \end{cases}$$

The goal is to find the price $f(t, x)$ of the option at time $t < T$ given $S_t = x$ One option, using the Feynman-Kac formula, is to price the option using

$$f(t, x) = \mathbb{E}[e^{-r(T-t)F(S_T)|S_t=x}],$$

which satisfies the PDE:

$$\partial_t f(t, x) = rf(t, x) - mx\partial_x f(t, x) - \frac{\sigma^2 x^2}{2}\partial_{xx} f(t, x)$$

The Black-Scholes approach is to let $f(t, x)$ be the value of a portfolio at time t, given that $S_t = x$, that can be hedged in order to guarantee a portfolio of value $F(S_T)$ at time T. Where a portfolio is an ordered pair (a_t, b_t), where a_t, b_t denote the number of units of stocks and bonds, respectively. Let V_t be the value of the portfolio at time t,

$$V_t = a_t S_t + b_t R_t \tag{10.7}$$

We will manage the portfolio by switching between stocks and bonds, so that no matter how the price of the stock moves, the value at time T will be

$$V_T = (S_T - K)_+$$

Assume that the portfolio is *self financing*, that one does not add outside resources to the portfolio, and thus the change of value of the portfolio is given by the change of the price of the assets,

$$dV_t = a_t dS_t + b_t dR_t \tag{10.8}$$

This is not a direct consequence of 10.7, since a_t, b_t vary with time and are not constant, we need to utilize the product rule:

$$d(a_t S_t) = da_t + a_t dS_t + d\langle a, S\rangle_t$$

$$d(b_t R_t) = b_t dR_t + R_t db_t$$

Thus we can see that 10.8 is a strong assumption about the portfolio. Thus, assuming 10.7

and plugging in:

$$\begin{aligned}dV_t &= a_t S_t[mdt + \sigma dB_t] + b_r r R_t dt \\ &= a_t S_t[mdt + \sigma dB_t] + r[V_t - a_t S_t]dt \\ &= [m a_t S_t + r(V_t - a_t S_t)]dt + \sigma a_t S_t dB_t \end{aligned} \tag{10.9}$$

By definition, $V_t = f(t, S_t)$, and thus assuming sufficient smoothness, Ito's formula shows that

$$\begin{aligned}dB_t &= df(t, S_t) \\ &= \partial_t f(t, S_t)dt + \partial_x f(t, S_t)dS_t + \frac{1}{2}\partial_{xx} f(t, S_t)d\langle S\rangle_t \\ &= \left[\partial_t f(t, S_t) + mS_t\partial_x f(t, S_t) + \frac{\sigma^2 S_t^2}{2}\partial_{xx} f(t, S_t)\right] dt\sigma S_t\partial_x f(t, S_t)dB_t \end{aligned} \tag{10.10}$$

By equating dB_t terms in the two above equation, we can see that the portfolio is given by

$$a_t = \partial_x f(t, S_t) \qquad b_t = \frac{V_t - a_t S_t}{R_t} \tag{10.11}$$

By equating the dt terms we get the Black-Scholes equation

$$d_t f(t, x) = rf(t, x) - rx\partial_x f(t, x) - \frac{\sigma^2 x^2}{2}\partial_{xx} f(t, x)$$

- m, the drift term does not appear, and is the same as Feynman-kac but with m replaced with r, and thus we can write

  $$f(t, x) = \mathbb{E}[e^{-r(T-t)F(S_t)}|S_t = x]$$

  where S satisfies

  $$dS_t = S_t[rdt + \sigma dB_t]$$

We can use this and compute the Black-Scholes formula:

$$f(T - t, x) = x\Phi\left(\frac{\log(x/K) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}\right) - Ke^{-rt}\Phi\left(\frac{\log(x/K) + (r - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}\right) \tag{10.12}$$

We can generalize this to the case when the stock price satisfies

$$dS_t = S_t[m(t, S_t)dt + \sigma(t, S_t)dB_t]$$

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

We again get (10.9) and (10.10) and by equating coefficients we get the Black-Scholes equation

$$\partial_t f(t, x) = r(t, x)f(t, x) - r(t, x)x\partial_x f(t, x) - \frac{\sigma(t, x)^2 x^2}{2}\partial_{xx} f(t, x) \tag{10.13}$$

The function f can be given by

$$f(t,x) = \mathbb{E}[R_t/R_T]F(S_T)|S_t = x$$

given that S_t, R_t, satisfy

$$dS_t = S_t[r(t,S_t)dt + \sigma(t,S_t)dB_t]$$

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

10.4 Martingale approach to Black-Scholes equation

Suppose the risk-free bond has rate $r(t,x)$ and the volatility is given by $\sigma(t,x)$. If R_t denotes the value of the bond at time t, then

$$dR_t = r(t,S_t)R_t dt, \qquad R_t = R_0 \exp\{\int_0^t r(s,S_s)ds\},$$

we must also assume that the stock satisfies

$$dS_t = S_t[r(t,S_t)dt + \sigma(t,S_t)dB_t] \tag{10.14}$$

and the value of the portfolio at time t satisfies

$$V_t = f(t,S_t) = \mathbb{E}_Q[(R_t/R_T)F(S_T)|S_t] = \mathbb{E}_Q[(R_t/R_T)F(S_T)|\mathcal{F}_t].$$

where $\mathbb{E}_Q$ is the expectation taken with respect to the probability measure which S_t satisfies (10.14).

Let $\tilde{S}_t = \frac{S_t}{R_t}$, $\tilde{V}_t = \frac{V_t}{R_t}$ be the stock price and the porfolio bounded discounted by the bond rate. The product rule shows that $\tilde{S}_t$ satisfies

$$d\tilde{S}_t = \sigma\tilde{S}_t dB_t$$

and thus $\tilde{S}_t$ is a martingale, and

$$\tilde{V}_t = \frac{V_t}{R_t} = \frac{\mathbb{E}_Q[(R_t/R_T)F(S_T)|\mathcal{F}_t]}{R_t} = \mathbb{E}_Q[\frac{F(S_T)}{R_T}|\mathcal{F}_t] = \mathbb{E}_Q[\tilde{V}_T|\mathcal{F}_t]$$

Theorem 10.4.1: What we derived above

Suppose S_t satisfies

$$dS_t = S_t[m(t, S_t)dt + \sigma(t, S_t)dB_t]$$

and a risk-free bond R_t is available at rate $r(t, S_t)$,

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

Suppose that the Brownian motion is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that there exists a probability measure Q that is an equivalent measure to $\mathbb{P}$ such that under $\mathbb{Q}$, the discounted stock price $\tilde{S}_t = \frac{S_t}{R_t}$ is a martingale. Suppose there is an option at time T with value $F(S_T)$ such that $\mathbb{E}[\frac{|F(S_T)|}{R_T}] < \infty$, then the arbitrage-free price of the option at time t is

$$V_t = R_t \mathbb{E}_Q(\frac{f(S_T)}{R_T}|\mathcal{F}_t)$$

Example.

Ok. This is fucking it. Let's derive this shit.

Suppose r, σ are constants and $F(S_T) = (S_T - K)_+$. The discounted values are $\tilde{S}_t = e^{-rt}S_t$, and $\tilde{V}_t = e^{-rt}V_t$, and

$$\tilde{V}_T = e^{-rT}F(S_T) = e^{-rT}(S_T - K)_+ = (\tilde{S}_T - \tilde{K})_+$$

where $\tilde{K} = e^{-rt}K$. nder the measure Q, $\tilde{S}_t$ satisfies

$$d\tilde{S}_t = \sigma\tilde{S}_t dB_t$$

implying that

$$\begin{aligned}\tilde{S}_T &= \tilde{S}_t \exp\left\{\int_t^T \sigma dB_s - \frac{1}{2}\int_t^T \sigma^2 ds\right\} \\ &= \tilde{S}_t \exp\left\{\sigma(B_T - B_t) - \frac{\sigma^2(T-t)}{2}\right\}\end{aligned}$$

Thus, the conditional distribution of $\tilde{S}_T$ given $\tilde{S}_t$ is that of

$$Z = \exp\{aN + y\}$$

given that $a = \sigma\sqrt{T-t}$, N is standard normal, and

$$y = \log\tilde{S}_t - \frac{a^2}{2}$$

It takes little work to show that Z has density

$$g(z) = \frac{1}{az}\phi(\frac{-y + \log(z)}{a})$$

where ϕ is the standard normal density, and thus,

$$\tilde{V}_t = \int_K^\infty (z - \tilde{K})g(z)dz$$

More calculus gives

$$\tilde{V}_t = \tilde{S}_t\Phi\left(\frac{\log(\tilde{S}_t/\tilde{K}) + \frac{a^2}{2}}{a}\right) - \tilde{K}\Phi\left(\frac{\log(\tilde{S}_t/\tilde{K} - \frac{a^2}{2})}{a}\right)$$

Implying that

$$V_t = e^{rt}\tilde{V}_t = S_t\Phi\left(\frac{\log(S_t/K) + rs + \frac{a^2}{2}}{a}\right) - e^{-rs}K\Phi\left(\frac{\log(S_t/K + rs - \frac{a^2}{2})}{a}\right)$$

where $s = T - t$, and thus plugging in $a = \sigma\sqrt{s}$ gives the Black Scholes Formula!

One of the hypothesis in the above Theorem is that if S_t satisfies

$$dS_t = S_t[m(t, S_t)dt + \sigma(t, S_t)dW_t],$$

then there exists a probability measure Q under which

$$dS_t = S_t[r(t, S_t)dt + \sigma(t, S_t)dW_t] \tag{10.15}$$

Where W_t is a $Q-$Brownian motion. Thus, if S_t satisfies (10.15), then the discount price satisfies

$$d\tilde{S}_t = \tilde{S}_t\sigma(t, S_t)dW_t \tag{10.16}$$

The Girsanov theorem tells us that the way to obtain Q is to tilt by the local martingale M_t, where

$$dM_t = M_t\frac{r(t, S_t) - m(t, S_t)}{\sigma(t, S_t)}dB_t$$

and thus in the measure Q,

$$dB_t = \frac{r(t, S_t) - m(t, S_t)}{\sigma(t, S_t)}dt + dW_t$$

10.5 Martingale approach to pricing

Figure 10.1: Agus Thrm

Suppose S_t denotes the price of an asset satisfying

$$dS_t = S_t[m_t dt + \sigma_t dB_t]$$

where B_t is a standard Brownian motion. Let $\mathcal{F}$ denote the information in $\{B_s | 0 \leq s \leq t\}$, and as usual assume that m_t, σ_t are processes adapted to the filtration $\{\mathcal{F}_t\}$. Assume that there is a risk-free bond R_t satisfying

$$dR_t = r_t R_t dt, \qquad R_t = R_0 \exp\{\int_0^t r_s ds\}$$

where r_t is adapted. Let T be a fixed future time and assume that V is an $\mathcal{F}_t-$measurable random valuable. We say that V is a claim at time T, such as

$$V = F(S_T), \quad V = \max\{0 \leq t \leq T\}S_t, \quad V = \frac{1}{T}\int_0^T S_t dt$$

Definition 10.5.1: Arbitrage-free price

If V is a claim at time T, then the *(arbitrage-free) price* V_t, $0 \leq t \leq T$, of a claim V_T is the minimum value of a self-financing portfolio that can be hedged to guarantee that its value at time T is V.

Remark.

> The goal is to determine the price V_t and the portfolio, $(a_t, b_t,)$ where a_t denotes the number of units of S_t and b_t the number of units f R.

Recall:

$$V_t = a_t S_t + b_t R_t$$

and (a_t, b_t) is self-financing if

$$dV_t = a_t dS_t + b_t dR_t$$

Let

$$\tilde{S}_t = \frac{R_0}{R_t} S_t, \qquad \tilde{V}_t = \frac{R_0}{R_t} V_t$$

denote the discount stock price and discount value, the latter of which is given by

$$\tilde{V}_t = a_t \tilde{S}_t + b_t R_0$$

And thus, using the product formula, we get that

$$d\tilde{S}_t = \tilde{S}_t[(m - r_t)dt + \sigma_t dB_t]$$

Remark.

> Our goal is to find a self-financing portfolio (a_t, b_t) such that with probability one,
>
> $$\tilde{V}_T = a_T \tilde{S}_t + b_T R_0 = \tilde{V}$$

Let Q be the probability measure that is an equivalent measure with respect to $\mathbb{P}$ such that $\tilde{S}_t$ under Q is a martingale. By Example 5.3.3 on Lawler, Girsanov tells us to choose

$$dQ = M_t d\mathbb{P}$$

where M_t satisfies

$$dM_t = \frac{r_t - m_t}{\sigma_t} M_t dB_t, \quad M_0 = 1 \tag{10.17}$$

- Assumption 1: The local martingale defined in 10.17 is a martingale. Let

$$W_t = B_t - \int_0^t \frac{r_s - m_s}{\sigma_s} ds$$

be the Brownian motion with respect to Q. Plugging into the $d\tilde{S}_t$ differential,

$$d\tilde{S}_t = \sigma \tilde{S}_t dW_t \tag{10.18}$$

- Assumption 2: The $Q-$local martingale $\tilde{S}_t$ satisfying (10.18) is a $Q-$martingale

Definition 10.5.2: Contingent Claim

A claim V at time T is called a *contingent claim* if $V \geq 0$ and

$$\mathbb{E}_Q[\tilde{V}^2] < \infty$$

Remark.

> The *(arbitrage-free) price* V_t $0 \leq t \leq T$ of a contingent claim V_T is the minimum value of a self-financing portfolio that can be hedged to guarantee that its value never drops below zero and at time T equals V.

Given a contingent claim, set

$$\tilde{V}_t = \mathbb{E}_Q[\tilde{V}|\mathcal{F}_t]$$

Thus, $\tilde{V}_t$ is a square integrable martingale. Assume there exists a process A_t such that

$$d\tilde{V}_t = A_t dW_t \tag{10.19}$$

then:

$$\begin{aligned}
dV_t &= R_t d\tilde{V}_t + \tilde{V}_t dR_t \\
&= R_t A_t dW_t + \tilde{V}_t dR_t \\
&= \frac{A_t}{\sigma^t \tilde{S}_t} R_t d\tilde{S}_t + \tilde{V}_t dR_t \\
&= \frac{A_t}{\sigma^t \tilde{S}_t} [dS_t - \tilde{S}_t dR_t] + \tilde{V}_t dR_t \\
&= \frac{A_t}{\sigma \tilde{S}_t} dS_t + \left[\tilde{V}_t - \frac{A_t}{\sigma_t}\right] dR_t
\end{aligned}$$

Thus, if

$$a_t = \frac{A_t}{\sigma_t \tilde{S}_t}, \qquad b_t = \tilde{V}_t - \frac{A_t}{\sigma_t} \tag{10.20}$$

then the portfolio is self financing and a simple calculation shows that

$$a_t S_t + b_t R_t = V_t$$

- Assumption 3: We can write $\tilde{V}_t$ as (10.19) and if a_t, b_t are defined as in (10.20), then

$$V_t = \int_0^t a_s dS_s + \int_0^t b_s dR_s$$

is well defined.

Theorem 10.5.3: Above stuff

If V is a contingent claim and assumptions 1-3 hold, then the arbitrage-free price is $V_t = R_t E_Q[\tilde{V}_T | \mathcal{F}_t]$

Example.

> This is the end. Close your eyes and count to 10.
>
> Assume that the stock price is a diffusion satisfying
>
> $$dS_t = S_t[m(t, S_t)dt + \sigma(t, S_t)dB_t]$$
>
> and the bond rate satisfies
>
> $$dR_t = r(t, S_t)R_t dt.$$
>
> The product rule then implies that the discounted stock price satisfies
>
> $$d\tilde{S}_t = \tilde{S}_t[(m(t, S_t) - r(t, S_t))dt + \sigma(t, S_t)dB_t]$$
>
> Given some claim V of the form $V = F(S_T)$, let ϕ be the function
>
> $$\phi(t, x) = \mathbb{E}_Q[(R_t/R_T)V | S_T = x]$$
>
> and note that
>
> $$V_t = \phi(t, S_t)$$

Using Ito's formula and recalling that

$$dS_t = S_t[r(t, S_t)dt + \sigma(t, S_t)dW_t,$$

we see that

$$d\tilde{V}_t = d[R_t^{-1}\phi(t, S_t)] = J_t dt + A_t dW_t$$

where

$$J_t = R_t^{-1}[\partial_t \phi(t, S_t) + \frac{\sigma(t, S_t)^2 S_t^2}{2} \partial_{xx}\phi(t, S_t) + r(t, S_t)S_t \partial_x \phi(t, S_t) - r(t, S_t)\phi(t, S_t)]$$

$$A_t = R_t^{-1} S_t \sigma(t, S_t)\partial_x \phi(t, S_t) = \tilde{S}_t \sigma(t, S_t)\partial_x \phi(t, S_t)$$

Since $\tilde{V}_t$ is a $Q-$martingale, $J_t = 0$, giving the Black-Scholes PDE. Since $d\tilde{V}_t = A_t dW_t$, then plugging into (10.20), we get

$$a_t = \partial_x(t, S_t)$$

$$b_t = R_t^{-1}[V_t - S_t \partial_x(t, S_t)]$$