Personal Notes on Stochastic Processes

Adel Saleh

January 27, 2021

Abstract

The following document is a comprehensive treatment of the fundamentals of stochastic processes and stochastic analysis. It is compiled from several sources, spanning from personal favorite textbooks [1]–[7] to lecture notes of MATH338 given by Dr. Abbas al Hakim (American University of Beirut), to miscellaneous remarks and observations made by many contributors from the MathStackExchange and MathOveflow communities.

This self-study document is mainly written to shed light on the many possible viewpoints of stochastic processes and to build a solid theoretical framework that emphasizes on the construction of such processes and their sample spaces, thus allowing one to easily link stochastic analysis with other areas of mathematics. Though mainly concentrated on classical results in the field, some to be written sections will be dedicated to discussing more specialized topics such as the stochastic wave equation and stochastic Hasegawa-Mima equation.

Moreover, one large section on measure theory and probability theory is included, heavily inspired by the MATH303 course given by Dr. Bassam Shayya (American University of Beirut), providing a rigorous footing on which all subsequent sections will rely.

For me personally, this document is journal in which I keep track of all that I learn on stochastic analysis. It's a work in progress by nature and proofs for some results are to be written. As this document evolves, I hope it becomes a reference for myself and others too.

Contents

Part I

Fundamental Results in Stochastic Analysis

1 Measure Theory

Inspired by the Math 303 course given by Dr. Bassam Shayya and [8].

1.1 General Lebesgue integral and the L^1 space

Definition (Measurable Space)**.** Let X be any set. Suppose that there exists a collection of subsets $\mathcal{M}$ of X with the following properties:

(i) $\emptyset, X \in \mathcal{M}$.

(ii) If $E \in \mathcal{M}$ then $E^c \in \mathcal{M}$.

(iii) $\mathcal{M}$ is closed under countable union.

Then we call $\mathcal{M}$ a σ-algebra and the pair $(X, \mathcal{M})$ a measurable space. //

Definition. A function f from the measure space $(X, \mathcal{F})$ to the measure space $(Y, \mathcal{G})$ is said to be measurable if for all $E \in \mathcal{G}$ we have $f^{-1}(E) \in \mathcal{F}$. We denote $\mathcal{L}^0(X, Y)$ the set of all measurable functions from X to Y. //

Proposition 1.1. *Let $(X, \mathcal{M}_1)$ and $(Y, \mathcal{M}_2)$ and $(Z, \mathcal{M}_3)$ be measurable spaces and let $f : X \to Y$ and $g : Y \to Z$ be measurable. Then $g \circ f$ is measurable.*

Proof. ■

In many cases when the target space has an additional structure, such as topological or algebraic (or both), one asks whether measurability is compatible with these structures. For example if the target space has a topology, and a sequence of measruable functions $\{f_n\}$ converges pointwise to f, is f measurable? If the target space is a vector space equipped with is a linear combination measurable functions also measurable? The answer is positive if $Y = \overline{\mathbb{R}}$, but even more generally when Y is a separable Banach space or if X is a perfect measure space and Y is any Banach space. In this section we only study the case when $Y = \overline{\mathbb{R}}$ and leave the more general case to Section 8.

First, let us equip the the target space Y with a suitable σ-algebra that encodes the topological structure that it already has.

Definition (Borel σ-algebra)**.** Let $(X, \mathcal{T})$ be a topological space. The Borel σ-algebra $\mathcal{B} = \mathcal{B}(X)$ is the smallest σ-algebra containing the topology $\mathcal{T}$. //

For the rest of the section, we will assume that the target space is always the extended real number line $\overline{\mathbb{R}}$ equipped with the Borel σ-algebra $\mathcal{B}$, and a function $f : X \to \overline{\mathbb{R}}$ is said to be Borel measurable if it's measurable w.r.t the Borel σ-algebra on $\overline{\mathbb{R}}$. When (X, Σ) is a measure space, we denote

$$\mathcal{M}(X) := \{f : X \to \overline{\mathbb{R}} \mid f \text{ is measurable w.r.t } \mathcal{B}\}.$$

Proposition 1.2. *Let (X, Σ) be a measurable space. Then $\mathcal{M}(X)$ is a real vector space.*

Lemma 1.3. *The function $f \in \mathcal{M}(X)$ if and only if for every $x \in \mathbb{R}$ we have $f^{-1}((-\infty, x)) \in \mathcal{M}$.*

Proof of Proposition 1.2. Let $f, g \in \mathcal{M}(X)$. Notice that we can write

$$\{f + g < b\} = \bigcup_{r \in \mathbb{Q}} \{f < r\} \cap \{g < b - r\}.$$

and therefore $f + g$ is measurable by the above lemma. Also, it is clear that for fixed $\alpha, x \in \mathbb{R}$ we have $\{\alpha f < x\} = \{f < x/\alpha\}$ is measurable since f is measurable. Therefore by induction we get that any finite combination of measurable functions is measurable. ■

Proposition 1.4. *Let (X, Σ) be a measurable space and let $\{f_n\} \subset \mathcal{M}(X)$.*

(i) *The functions $f = \inf_n f_n$ and $g = \sup_n f_n$ are in $\mathcal{M}(X)$.*

(ii) *If $f = \limsup f_n$ or $f = \liminf f_n$ then $f \in \mathcal{M}(X)$.*

Proof. ■

Lemma 1.5. *To show that a function f is Borel measurable, it suffices to show that the set $f^{-1}((-\infty, a])$ is measurable for all $a \in \overline{\mathbb{R}}$.*

Proof. ■

Definition. A function $f : (X, \mathcal{M}) \to (\mathbb{R}, \mathcal{B})$ is called simple if there are sets $E_1, \ldots, E_n \in \mathcal{M}$ and constants $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$f(x) = \sum_{k=1}^{n} c_k \cdot \mathbf{1}_{E_k}(x).$$

We denote the space of all simple functions on X as $S(X)$. //

The essential property of measurable functions is that they are pointwise limits of simple functions. This alone helps in understanding and characterizing many other important properties of measurable functions, and most subsequent results in this section are due to this approximation property.

Proposition 1.6. *Let $f : X \to \mathbb{R}$ be a non-negative measurable function. There is a sequence of sets $\{E_n\} \subset \mathcal{M}$ such that*

$$f(x) = \sum_{n=1}^{\infty} \frac{\mathbf{1}_{E_n}(x)}{k}. \tag{1}$$

This implies that the space of simple functions $S(X)$ is dense in $L^0_+(X)$ with respect to the topology of pointwise convergence.

Intuition: A necessary condition for (1) to hold is that for all $x \in X$ and all $n \in \mathbb{N}$ we have

$$f_n(x) := \sum_{k=1}^{n} \frac{\mathbf{1}_{E_k}(x)}{k} \leq f(x).$$

This is the case since the sequence $\{f_n\}$ is increasing with f as it's pointwise limit. Also,

$$f_{n+1}(x) = \begin{cases} f_n(x) & \text{if } x \notin E_{n+1}, \\ f_n(x) + \frac{1}{n+1} & \text{if } x \in E_{n+1}. \end{cases}$$

First let $f_0(x) = 0$ for all $x \in X$. Let

$$E_1 = \{x \in X : f(x) \geq 1\}.$$

If $x \in E_1$ then $f(x) \geq 1$ and hence we define $f_1(x) = f_0(x) + 1 = 1$. If $x \notin E_1$, we set $f_1(x) = f_0(x)$. Hence

$$f_1(x) = f_0(x) + \mathbf{1}_{E_1}(x) = \mathbf{1}_{E_n}(x).$$

Now let

$$E_2 = \left\{x \in X : f(x) \geq f_1(x) + \frac{1}{2}\right\}$$

If $x \in E_2$ then we set $f_2(x) = f_1(x) + 1/2$, otherwise we set $f_2(x) = f_1(x)$. Therefore we can write

$$f_2(x) = f_1(x) + \frac{1}{2}\mathbf{1}_{E_2}(x) = \mathbf{1}_{E_1} + \frac{1}{2}\mathbf{1}_{E_2}.$$

At this point we have that

$$f_2(x) = \begin{cases} 1 + \frac{1}{2} & \text{if } x \in E_1 \cap E_2, \\ 1 & \text{if } x \in E_1 \setminus E_2, \\ \frac{1}{2} & \text{if } x \in E_2 \setminus E_1, \\ 0 & \text{if } x \notin E_1 \cup E_2. \end{cases}$$

What f_2 is doing is checking that if $f_1(x) + 1/2$ exceeds $f(x)$ then keep $f_1(x)$ as is, otherwise add $1/2$ to $f_1(x)$.

Proof of Proposition 1.6. With E_1 defined as above, define recursively

$$E_n = \left\{x \in X : f(x) \geq f_{n-1}(x) + \frac{1}{n}\right\} \quad \text{and} \quad f_n(x) = f_{n-1}(x) + \frac{1}{n}\mathbf{1}_{E_n}(x).$$

For each x, it is clear that the non-negative sequence $\{f_n(x)\}$ is non-decreasing and bounded from above by $f(x)$. To show that $f_n(x) \to f(x)$, it suffices to show that a subsequence of converges to $f(x)$. Let n_0 be the smallest integer such that $1/n_0 \leq f(x)$. Then let $m_0 \geq 1$ be the largest integer such that

$$\frac{1}{n_0} + \frac{1}{n_0 + 1} + \cdots + \frac{1}{n_0 + m_0} \leq f(x).$$

Then let $n_1 > n_0 + m_0$ be the smallest integer such that

$$\sum_{k=1}^{n_0+m_0} \frac{1}{k} + \frac{1}{n_1} \leq f(x),$$

and then m_1 be the largest integer such that

$$\sum_{k=1}^{m_0} \frac{1}{n_0 + k} + \sum_{k=1}^{m_1} \frac{1}{n_1 + k} \leq f(x) \quad \text{so that} \quad \sum_{k=1}^{m_0} \frac{1}{n_0 + k} + \sum_{k=1}^{m_1+1} \frac{1}{n_1 + k} \geq f(x).$$

Then let $n_2 \geq m_1 + n_1 + 1$ be the smallest integer such that

$$\sum_{k=1}^{m_0} \frac{1}{n_0 + k} + \sum_{k=1}^{m_1} \frac{1}{n_1 + k} + \frac{1}{n_2} \leq f(x).$$

We have that

$$f(x) - f_{n_2}(x) = f(x) - \sum_{k=1}^{m_0} \frac{1}{n_0 + k} - \sum_{k=1}^{m_1} \frac{1}{n_1 + k} - \frac{1}{n_2} \leq \frac{1}{n_1 + m_1 + 1} \leq \frac{1}{n_2}.$$

Proceeding in this fashion, we obtain a sequence of integers $n_0 \leq n_1 \leq n_2 \leq \cdots \leq n_k$ and $m_0 \leq m_1 \leq \cdots \leq m_k$ with $n_{j+1} \geq n_j + m_j + 1$ such that

$$\sum_{j=0}^{k-1} \sum_{i=0}^{m_j} \frac{1}{m_j + i} + \frac{1}{n_k} \leq f(x) \quad \text{and} \quad f(x) - f_{n_k}(x) \leq \frac{1}{n_k},$$

and therefore the sequence $f_{n_k}(x)$ converges to $f(x)$ as desired. ■

Proposition 1.7 (Another approximating sequence)**.** *Let $f \in L^0_+(X)$ and for each $n \in \mathbb{N}$ define*

$$f_n(x) = \begin{cases} 2^{-n}(j-1) & \text{if } 2^{-n}(j-1) \leq f(x) < 2^{-n}j, \\ n & \text{if } f(x) \geq n \end{cases}.$$

Then we have that $f_n \nearrow f$. Furthermore, for any set $E \in \mathcal{M}$ on which f is bounded, the convergences is actually uniform.

Proof. ■

Now that we have established basic properties of real valued measurable functions, we can move on to define measures on a measurable space $(X, \mathcal{M})$.

Definition. A measure on a measurable space $(X, \mathcal{M})$ is a function $\mu : \mathcal{M} \to [0, \infty]$ such that for $\mu(\emptyset) = 0$ and for any sequence $\{E_n\} \subset \mathcal{M}$ we have

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n). \tag{2}$$

The triple $(X, \mathcal{M}, \mu)$ is called a measure space. //

Theorem 1.8 (Continuity property)**.** *Let $(X, \mathcal{M})$ be a measurable space and let $\mu : \mathcal{M} \to \mathbb{R}$ be a function such that $\mu(\emptyset) = 0$. Then μ is a measure if and only if the following hold.*

(i) *μ is finitely additive.*

(ii) *For any increasing sequence of measurable sets $\{E_n\}$ we have*

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n).$$

(iii) *In addition if $\mu < \infty$ then for any sequence of decreasing measurable sets $\{E_n\}$ we have*

$$\mu\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mu(E_n).$$

Proof. ■

Theorem 1.9 (Borel-Cantelli)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\{E_n\} \subset \mathcal{M}$. Then*

$$\sum_{n=1}^{\infty} \mu(E_n) < \infty \quad \textit{implies} \quad \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = 0.$$

Proof. We have that

$$\mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right) = \lim_{n\to\infty} \mu\left(\bigcup_{k=n}^{\infty} E_k\right) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mu(E_k) = 0.$$

The first equality is due to the continuity property of μ, the inequality is due to countable sub-additivity and the last equality is justified since the limit of the tail of convergent series is 0. ■

Proposition 1.10. *Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{E_n\} \subset \mathcal{M}$ such that*

$$\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k,$$

If E is the set of the above equality then $\lim_{n\to\infty} \mu(E_n) = \mu(E)$.

Theorem 1.11 (Egorov)**.** *Let $(X, \mathcal{M}, \mu)$ be a finite measure space. Suppose that $\{f_n\} \subset L^0(X)$ converges to $f \in L^0(X)$ almost everywhere. Then for every $\epsilon > 0$, there is a set $E \in \mathcal{M}$ such that $\mu(E) < \epsilon$ and $f_n \to f$ uniformly on $X \setminus E$.*

Proof. Let

$$E(n, k) = \left\{ x \in X : |f_n(x) - f(x)| \geq \frac{1}{k} \right\}.$$

Notice that if $x \in X$ is such that $f_n(x) \to f(x)$, then for any fixed $k \in \mathbb{N}$, x cannot be in infinitely many of the $E(n, k)$'s. Since convergence happens for almost all $x \in X$ this means that

$$\mu\big(\{x \in X : x \text{ is in infinitely many } E(n,k)\text{'s}\}\big) = \mu\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E(n,k)\right) = 0, \quad \text{for all } k \in \mathbb{N}.$$

Since $\mu(X) < \infty$ we have by part (iii) of Theorem 1.8 that

$$\lim_{m\to\infty} \mu\left(\bigcup_{n=m}^{\infty} E(n,k)\right) = 0, \quad \text{for all } k \in \mathbb{N},$$

and therefore for fixed $\epsilon > 0$ and fixed k, there is an integer m_k such that for all $m \geq m_k$ we have

$$\mu\left(\bigcup_{n=m}^{\infty} E(n,k)\right) < \frac{\epsilon}{2^k}.$$

Thus if we define

$$E = \bigcup_{k=1}^{\infty} \bigcup_{n=m_k}^{\infty} E(n,k), \quad \text{then} \quad \mu(E) \leq \sum_{k=1}^{\infty} \mu\left(\bigcup_{n=m_k}^{\infty} E(n,k)\right) < \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \epsilon.$$

Also by the definition of E, we have have that for any $k \in \mathbb{N}$ there is an integer $m_k \in \mathbb{N}$ such that for all $x \in X \setminus E$ and all $m \geq m_k$ we have

$$|f_m(x) - f(x)| < \frac{1}{k} \quad \text{so that} \quad \sup_{x \in X \setminus E} |f_m(x) - f(x)| \leq \frac{1}{k},$$

and hence $f_n \to f$ uniformly on $X \setminus E$ as desired. ■

Now the we have finished setting up basic properties of measurable functions, we are in good shape to define the Lebesgue integral for $\overline{\mathbb{R}}$ valued functions. The approach would be to define the integral for simple functions and proving some of it's properties. Then using the density of simple functions to extend the definition to $L^0_+(X)$ and then use Motonone Convergence to extend these properties to functions in $L^0_+(X)$.

Definition (Lebesgue integral)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space. Define

$$\boxed{\int_X f d\mu := \sum_{k=1}^{n} c_k \mu(E_k) \text{ for } f \in S(X).}$$

Then use the above to define

$$\boxed{\int_X f d\mu := \sup\left\{\int_X s d\mu : s \in S(X) \text{ and } 0 \leq s \leq f\right\} \text{ for } f \in L^0_+(X).}$$

We extend this definition for a specific set of functions in $L^0(X)$ namely

$$L^1(X) := \left\{f \in L^0(X) : \int_X |f| d\mu < \infty\right\},$$ [1]

This guarantees that the following definition makes sense

$$\boxed{\int_X f d\mu := \int_X f^+ d\mu - \int_X f^- d\mu \text{ for } f \in L^1(X)}. \tag{3}$$

If $A \in \mathcal{M}$ we define

$$\boxed{\int_A f d\mu := \int_X \mathbf{1}_A \cdot f d\mu.}$$

//

Remark. For functions $f \in L^0(X) \setminus L^0_+(X)$, if we have

$$f \in E := \left\{f \in L^0(X) : \text{ either } \int_X f^+ d\mu < \infty \text{ or } \int_X f^- d\mu < \infty\right\},$$

then we can use (3) to define the integral of f with the integral possibly being $\pm\infty$.

[1] Equivalently equivalently

$$L^1(X) := \left\{f \in L^0(X) : \int_X f^+ d\mu < \infty \text{ and } \int_X f^- d\mu < \infty\right\}.$$

Lemma 1.12. *For simple functions, the Lebesgue integral has the same operational properties of the Riemann integral and satisfies the same inequalities.*

Proof. Let f and g be simple functions on the measure space $(X, \mathcal{M}, \mu)$ and write

$$f = \sum_{i=1}^{m} c_i \mathbf{1}_{E_i} \quad \text{and} \quad g = \sum_{j=1}^{n} d_j \mathbf{1}_{F_j}.$$

We can always assume that each of the collections $\{E_i\}$ and $\{F_j\}$ partition of X. This will allows us to write

$$\mathbf{1}_{E_i} = \sum_{j=1}^{n} \mathbf{1}_{E_i \cap F_j}, \quad \text{so that} \quad f = \sum_{i=1}^{m} \sum_{j=1}^{n} c_i \mathbf{1}_{E_i \cap F_j},$$

and similarly that

$$\mathbf{1}_{F_j} = \sum_{i=1}^{m} \mathbf{1}_{E_i \cap F_j}, \quad \text{so that} \quad g = \sum_{i=1}^{m} \sum_{j=1}^{n} d_j \mathbf{1}_{E_i \cap F_j}.$$

We will use this to prove the lemma.

(i) **(Monotonicity).** Suppose that $f \leq g$. Picking any element $x \in E_i \cap F_j$ tells us that $c_i = f(x) \leq g(x) = d_j$ for any $1 \leq i, j \leq m$ such that $E_i \cap F_j$ is non empty. Therefore,

$$\int_X f d\mu = \sum_{i=1}^{m} \sum_{j=1}^{n} c_i \mu(E_i \cap E_j) \leq \sum_{i=1}^{m} \sum_{j=1}^{n} d_j \mu(E_i \cap E_j) = \int_X g d\mu.$$

(ii) **(Linearity).** We have that

$$\int_X (f+g) d\mu = \sum_{i,j=1}^{m,n} (c_i + d_j) \mathbf{1}_{E_i \cap F_j} = \sum_{i,j=1}^{m,n} c_i \mathbf{1}_{E_i \cap F_j} + \sum_{i,j=1}^{m,n} c_i \mathbf{1}_{E_i \cap F_j} = \int_X f d\mu + \int_X g d\mu.$$

(iii) **(Absolute Value).**

$$\left| \int_X (f+g) d\mu \right| = \left| \sum_{i,j=1}^{m} (c_i + d_j) \mu(E_i \cap F_j) \right| \leq \sum_{i,j=1}^{m,n} |c_i + d_j| \mu(E_i \cap F_j) = \int_X |f+g| d\mu.$$

which completes the proof. ∎

Lemma 1.13. *Let $(X, \mathcal{M}, \mu)$ be a measure space. Suppose that there are measurable functions f and g and sequences $\{f_n\}$ and $\{g_n\}$ that converge pointwise to f and g respectively. If $f < g$ and $G_n = \{f_n \leq g_n\}$, then $\lim_{n\to\infty} \mu(G_n) = \mu(X)$. In addition, if μ is finite then $\lim_{n\to\infty} \mu(E \setminus G_n) = 0$.*

Proof. For each $x \in X$, since $f(x) \leq g(x)$ and $f_n(x) \to f(x)$ and $g_n(x) \to g(x)$ then there is integer $N \in \mathbb{N}$ such that for all $n \geq N$ we have $f_n(x) \leq g_n(x)$. But this means that $x \in G_n$ for all $n \geq N$ and therefore

$$x \in E := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} G_n,$$

and thus $E = X$. By part (iii) of Theorem 1.8 we have that $\mu(E_n) \to \mu(X)$ as desired. ∎

Proposition 1.14. *Let f and g be two positive measurable functions and let $E, F \subset \mathcal{M}$.*

(i) *If $f \leq g$ then*

$$\int_X f d\mu \leq \int_X g d\mu.$$

(ii) *If $E \subset F$ then*

$$\int_E f d\mu \leq \int_F f d\mu.$$

Proof. Let s be a simple any simple function such that $0 \leq s \leq f$. Then $s \leq g$ and hence we have (i). Part (ii) follows by applying part (i) to $\mathbf{1}_E f$ and $\mathbf{1}_F f$. ∎

Theorem 1.15 (Monotone Convergence Theorem)**.** *Let $\{f_n\}$ be an increasing sequence of positive measurable functions that converge pointwise to f. Then*

$$\int_X f d\mu = \lim_{n\to\infty} \int_X f_n d\mu.$$

Proof. Let s be any simple function such that $0 \leq s \leq f$ and let $0 < \alpha < 1$ be arbitrary. Let $G_n = \{f_n \geq \alpha s\}$ then it is clear that every $x \in X$ is eventually in G_n so that $G_n \nearrow X$. Therefore,

$$\lim_{n\to\infty} \int_{G_n} s d\mu = \lim_{n\to\infty} \int_X \mathbf{1}_{G_n} \cdot s d\mu = \lim_{n\to\infty} \sum_{k=1}^{m} c_k \mu(E_k \cap G_n) = \sum_{k=1}^{m} c_k \mu(E_k) = \int_X s d\mu.$$

Also we have that

$$\int_{G_n} \alpha s d\mu \leq \int_{G_n} f_n d\mu \leq \int_X f_n d\mu.$$

And therefore by taking limits

$$\alpha \int_X s d\mu \leq \lim_{n\to\infty} \int_X s d\mu.$$

Since this is true for all $\alpha \in (0,1)$, then by taking limit as $\alpha \to 1$ this is inequality becomes true for $\alpha = 1$. Since s was arbitrary, we get

$$\int_X f d\mu \leq \lim_{n\to\infty} \int_X f_n d\mu,$$

The other reverse inequality follows immediately from monotonicity. ∎

Corollary 1.15.1. *The Lebesgue integral for positive measurable functions has the same operational properties as the Riemann integral and satisfies the same inequalities.*

Proof. Let f, g be positive measurable functions on the measure space $(X, \mathcal{M}, \mu)$. Let $\{f_n\}$ and $\{g_n\}$ be sequences of simple functions increasing to f and g respectively.

(i) **(Linearity).** Without loss of generality assume that f and g are non-negative.

$$\int_X (f+g) d\mu = \lim_{n\to\infty} \int_X (f_n + g_n) d\mu = \lim_{n\to\infty} \left[\int_X f_n d\mu + \int_X g_n d\mu \right] = \int_X f d\mu + \int_X g d\mu.$$

(ii) **(Absolute Value).**

$$\begin{aligned}\left|\int_X (f+g)d\mu\right| &= \left|\lim_{n\to\infty}\left(\int_X f_n d\mu + \int_X g_n d\mu\right)\right| = \lim_{n\to\infty}\left|\int_X f_n d\mu + \int_X g_n d\mu\right| \\ &\le \lim_{n\to\infty}\left|\int_X f_n d\mu\right| + \lim_{n\to\infty}\left|\int_X g_n d\mu\right| \\ &\le \lim_{n\to\infty}\int_X |f_n| d\mu + \lim_{n\to\infty}\int_X |g_n| d\mu \\ &= \int_X |f| d\mu + \int_X |g| d\mu.\end{aligned}$$

■

Corollary 1.15.2. *Suppose that $\{f_n\}$ is a non-negative sequence that increases to f almost everywhere on X. Then*

$$\int_X f d\mu = \lim_{n\to\infty}\int_X f_n d\mu.$$

Here we have used the monotone convergence theorem to prove linearity of the Lebesgue integral, an approach similar to the one in [8]. This approach seems natural as the integral is defined as limit of integrals of simple functions, hence we extend the properties of the Lebesgue integral of simple functions to the Lebesgue integral of general measurable functions. However, some authors would argue that proving MCT before the algebraic properties of the Lebesgue integral is premature. Both points are valid, but the former is more suitable in the context of probability theory and more specifically in the construction of the Itô integral in Section 4.

Theorem 1.16 (Fatou's Lemma)**.** *Let $\{f_n\}$ be a sequence of positive measurable functions. Then*

$$\int_X \liminf_{n\to\infty} f_n d\mu \le \liminf_{n\to\infty}\int_X f_n d\mu,$$

where both sides can be equal to $+\infty$.

Proof. Let $g_m = \inf_{k\ge m}\{f_k\}$ so that $g_m \le f_n$ when $m \le n$. This tells us that for all $m \le n$ we have

$$\int_X g_m d\mu \le \int_X f_n d\mu \quad \text{so that} \quad \int_X g_m d\mu \le \liminf_{n\to\infty}\int_X g_n d\mu.$$

Now $\{g_m\}$ is an increasing sequence of measurable functions that converge pointwise to $\liminf f_n$ and therefore by MCT we have

$$\lim_{n\to\infty}\int_X g_n d\mu = \int_X \lim_{n\to\infty} g_n d\mu = \int_X \liminf_{n\to\infty} f_n d\mu,$$

and the inequality is proved. ■

Corollary 1.16.1. *If f is non-negative measurable then*

$$f = 0 \text{ almost everywhere on } X \iff \int_X f d\mu = 0.$$

Proof. Assume that the integral of f is 0 and let

$$E_n = \left\{ x \in X : f(x) > \frac{1}{n} \right\}.$$

Then by definition $f \geq (1/n) \cdot \mathbf{1}_{E_n}$ and therefore

$$0 = \int_X f d\mu \geq \int_{E_n} f d\mu \geq \frac{1}{n}\mu(E_n),$$

and hence $\mu(E_n) = 0$. It follows that

$$\mu\left(\{x \in X : f(x) > 0\}\right) = \mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n) = 0,$$

and hence $f = 0$ almost everywhere. Conversely, suppose that $f = 0$ almost everywhere. This means that $\mu(\{f > 0\}) = 0$. Now let $f_n = n \cdot \mathbf{1}_{\{f>0\}}$. It is clear that $f \leq \liminf f_n$ and

$$\int_X f_n d\mu = n\mu(E) = 0.$$

Then by Fatou's lemma we obtain

$$\int_X f d\mu \leq \int_X \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu = 0,$$

which concludes the proof. ■

Theorem 1.17 (Dominated Convergence)**.** *Suppose we are given sequences* $\{f_n\} \subset L^0(X)$ *and* $\{g_n\} \subset L^0_+(X)$ *such that*

(i) $\lim_{n\to\infty} f_n = f$.

(ii) $\lim_{n\to\infty} g_n = g \in L^1(X)$.

(iii) $\lim_{n\to\infty} \int g_n d\mu = \int g d\mu$.

(iv) $|f_n| \leq g_n$ *for all* $n \in \mathbb{N}$.

Under these assumptions we have

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu.$$

Proof. To start the proof, define

$$\varphi_n = g_n + g - |f_n - f|$$

Then φ_n is positive measurable since

$$|f_n| \leq g_n \implies |f| \leq g \implies \varphi_n \geq g_n + g - (|f_n| + |f|) \geq g_n + g - (g_n + g) = 0.$$

Also notice that $\varphi_n \to 2g$ as $n \to \infty$ almost everywhere on E. Now by Fatou's Lemma

$$\int 2g d\mu = \int 2 \lim_{n\to\infty} g_n d\mu = \int \lim_{n\to\infty} \varphi_n d\mu = \int \liminf_{n\to\infty} \varphi_n d\mu \leq \liminf_{n\to\infty} \int \varphi_n d\mu,$$

and

$$\varphi_n \leq g_n + g \implies \int \varphi_n d\mu \leq \int (g_n + g) d\mu \implies \limsup_{n\to\infty} \int \varphi_n d\mu \leq \int 2g d\mu,$$

and thus

$$\lim_{n\to\infty} \int \varphi_n d\mu = \int 2g d\mu.$$

Therefore we get

$$\int (g_n + g) d\mu = \int \varphi_n d\mu + \int |f_n - f| d\mu$$

Letting $n \to \infty$ we get the desired result. ■

1.2 Outer measures, product measures and the Lebesgue-Stieltjes measure on $\mathbb{R}^n$

Definition. An outer measure on a set X is a function $\mu^* : 2^X \to [0, \infty]$ such that $\mu^*(\emptyset) = 0$ and for any sequence $\{E_n\} \subset 2^X$ that cover a set $E \in 2^X$ we have

$$\mu^*(E) \leq \sum_{n=1}^{\infty} \mu^*(E_n). \tag{4}$$

This property is called countable sub-additivity. //

Outer measures are used in constructing some measures, such as the Lebesgue-Stieltjes measure on $\mathbb{R}^n$ and general product measures. This is highlighted by the following theorem.

Theorem 1.18 (Caratheodory)**.** *Suppose X has an outer-measure μ^*. Let $\mathcal{M}$ be the collection of all subset A of X such that for all $E \in 2^X$ we have*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A) \tag{5}$$

Then $\mathcal{M}$ is a σ-algebra and $\mu^*_{|\mathcal{M}}$ is a measure.

Proof. ■

Condition (5) has an intuitive explanation in terms of events and probability, as will be explained in section 1.6.

Definition (Lebesgue-Stieltjes outer measure on $\mathbb{R}$)**.** Let $F : \mathbb{R} \to \mathbb{R}^+$ be an increasing function. For an interval $I = (a, b)$ in $\mathbb{R}$ define

$$\lambda^*(I) = F(b^-) - F(a^+).$$

Now let E be any subset of $\mathbb{R}$. We define

$$\lambda^*(E) = \inf \left\{ \sum_n \lambda^*(I_n) \,\middle|\, \{I_n\} \text{ countable covering of } E \text{ with bounded open intervals} \right\},$$

where infimum can be $+\infty$. It is clear λ^* satisfies countable subadditivity. //

Notice that we do not assume that $\lambda^*(\emptyset) = 0$ yet since we can deduce it from countable subadditivity, as showcased in the following.

Lemma 1.19. *Let D_f be the set of discontinuities of a real function f. Then D_f is countable.*

Proof. Let D_f be the set of discontinuities of f. Then either we have $f(x^-) \leq f(x) < f(x^+)$ or $f(x^-) < f(x) \leq f(x^+)$. In the first case, there is a rational number $q(x) \in \mathbb{Q}$ such that $f(x) < q < f(x^+)$ and in the second case $f(x^-) < q < f(x)$. It is easy to see that $q : D \to \mathbb{Q}$ is injective. ■

Proposition 1.20. *Suppose $x \notin D_f$, then $\lambda^*(\{x\}) = 0$. Since $\emptyset \subset \{x\}$ this implies that $\lambda^*(\emptyset) = 0$ and hence λ^* is an outer measure.*

Proof. Since D_f is countable, then $\mathbb{R} \setminus D_f$ is dense. For $x \in X \setminus D_f$, let $\{a_n\}$ and $\{b_n\}$ be sequences in $X \setminus D_f$ such that $a_n \nearrow x$ and $b_n \searrow x$. By definition we will then have

$$\lambda^*(\{x\}) \leq \lambda^*((a_n, b_n)) = F(b_n) - F(a_n),$$

and since F is continuous at x then taking limits in the above equation completes the proof. ■

Definition. Let $(\mathbb{R}, \Sigma^{(1)}, \lambda)$ be the measure space obtained by restricting λ^* to measurable sets. When $F(x) = x$ we call λ the Lebesgue measure. //

Proposition 1.21. *$\Sigma^{(1)}$ contains the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$.*

Definition. For every $x \in \mathbb{R}$, pick an element $v \in x + \mathbb{Q}$ such that $v \in [0, 1]$. The collection of all such v's is called a Vitali set. //

Proposition 1.22 (Non measurability of Vitali set)**.** *We have that $V \notin \Sigma^{(1)}$.*

We now proceed with constructing the Lebesgue measure on $\mathbb{R}^n$ for $n \geq 2$. There are two approaches: one would be to use the Lebesgue measure λ defined on $(\mathbb{R}, \Sigma^{(1)})$ and use it to construct a product measure structure on $\mathbb{R}^n$ inductively. This approach draws parallels with the one used to construct the *coin tossing space*, a fundamental and intuitive example of a *probability space* which is product of "smaller" coin tossing spaces. Another approach would be to construct a Lebesgue outer measure on $\mathbb{R}^n$ similar to the Lebesgue outer measure on $\mathbb{R}^n$.

Theorem 1.23 (Hahn-Kolmogorov)**.** *Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ be σ-finite measure spaces and let $\mathcal{F} \otimes \mathcal{G} := \sigma(\mathcal{F} \times \mathcal{G})$. If we define $\eta : \mathcal{F} \times \mathcal{G} \to \mathbb{R}$ as*

$$\eta(F \times G) = \mu(F)\nu(G).$$

then η extends to a unique measure on $\mathcal{F} \otimes \mathcal{G}$.

Proof. ■

Definition. If $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ are σ-finite measure spaces then we denote

$$(X \times Y, \mathcal{F} \otimes \mathcal{G}, \mu \times \nu)$$

the product measure space constructed in Theorem 1.23. //

The above theorem will be later used to construct the Lebesgue-Stiltjies measure on $\mathbb{R}^n$.

Definition (Lebesgue-Stieltjes outer measure on $\mathbb{R}^n$ for $n \geq 2$)**.** Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing function. A box B in $\mathbb{R}^n$ is a set of the form

$$[a_1, b_1] \times \cdots \times [a_n, b_n].$$

We define a function λ^* on boxes as

$$\lambda^*(B) = \prod_{k=1}^{n} (b_k - a_k).$$

For $E \in 2^X$, define the Lebesgue outer measure as

$$\lambda^*(E) = \inf \left\{ \sum_{k=1}^{n} \lambda^*(B_k) : \{B_k\} \text{ collection of boxes s.t } E \subset \bigcup_{n=1}^{\infty} B_k \right\}.$$

Denote $\mathcal{L}(\mathbb{R}^n)$ the σ-algebra on $\mathbb{R}^n$ by restricting the Lebesgue-Stieltjes outer measure as in Theorem 1.18. //

Definition (Lebesgue-Stieltjes product measure on $\mathbb{R}^n$)**.** //

Proposition 1.24. *We have the inclusions*

$$\mathcal{B}(\mathbb{R}^n) \subsetneq \bigotimes^n \mathcal{L}(\mathbb{R}) \subsetneq \mathcal{L}(\mathbb{R}^n).$$

1.3 Measures on topological spaces, Borel and Radon Measures

Definition (Borel σ-algebra, Borel measure)**.** If $(X, \mathcal{T})$ is a topological space, we define the Borel algebra $\mathcal{B} := \mathcal{B}(X)$ to be the smallest σ-algebra containing $\mathcal{T}$. Any measure defined on $\mathcal{B}$ is a Borel measure. //

Proposition 1.25.

Theorem 1.26 (Lusin)**.** *Let X be a locally compact Hausdorff toplogical space equipped with the Borel σ-algebra and a Radon measure μ. Let $f \in L^0(X)$ such that $\mu(\{f \neq 0\}) < \infty$. For every $\epsilon > 0$, there a function $g \in C_c(X)$ such that*

(i) $\mu(\{f \neq g\}) < \epsilon$.

(ii) *If f is bounded then one can have $\|g\|_{L^\infty(\mu)} \leq \|f\|_{L^\infty(\mu)}$.*

1.4 L^p Spaces

This section is inspired by [3], covering most of the fundamental properties of these spaces and some of their uses.

1.5 Absolute continuity and Radon-Nikodym theorem

Theorem 1.27 (Lebesgue-Radon-Nikodym)**.** *Let μ and ν be finite measures on a measurable space $(X, \mathcal{M})$. There is a function $f \in L^0(\mu) \cap L^0(\nu)$ and a μ-null set $F \in \mathcal{M}$ such that for all $E \in \mathcal{M}$ we have*

$$\nu(E) = \int_E f d\mu + \nu(E \cap F).$$

2 Basic Probability Theory

Inspired by [2], [9].

2.1 Sample spaces, measurable events and the probability measure

In many ways this section is inspired by excellent books [2], [9]. I personally took measure theory before taking any probability theory, and the definition of measurability was a bit arbitrary for me at first, especially that my source of intuition was always geometry and areas. The more mysterious equation to me was the Caratheodory condition (5), and how it was used to get a measure from an outer measure. This shroud around the notion of measurability was removed as soon as I took probability theory, and understood measurable sets from the point of view of sample spaces and events.

Definition. A probability space is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}$ takes values in the interval $[0, 1]$. The set Ω is called a sample space and any measurable set is called an event. //

In more grounded terms, Ω contains all possible outcomes ω of an experiment that can be replicated. An event is therefore a collection of outcomes and events containing only one outcome are called simple events. Now let us say that the experiment was done that the outcome ω has been observed. If $\omega \in E$ then we say the event E happened. But this means that $\Omega \setminus E$ did not happen. Also, if we can tell whether $\omega \in E$ or $\omega \in F$ then the event $E \cup F$ happened, meaning that either or F happened. This suggests that following definiton for the set of measurable events.

1. If E is measurable then E^c is measurable.

2. IF E and F are measurable then $E \cup F$ is measurable.

The collection of all such events is called a σ-field. Is it still not a σ-algebra as we still need to have countable unions. However, suppose we further have

3. If $E_1 \subset E_2 \subset E_3 \subset \cdots$ are measurable then $\cup E_n$ is measurable.

4. If $E_1 \supset E_2 \supset E_3 \supset \cdots$ are measurable then $\cap E_n$ is measurable.

The the set of all measurable events becomes closed under countable unions. This makes it easier in some cases to deduce that a collection of sets is actually a σ-algebra.

Remark. Let $\{E_\alpha\}_{\alpha \in J}$ where J is uncountable be a collection of measurable events. From an intuitive point of view, it might seem reasonable to think that if we can tell whether an outcome $\omega \in E_\beta$ for some $\beta \in J$ then we can tell that $E = \bigcup_{\alpha \in J} E_\alpha$ happened (ie E is measurable). In that case, whether a set E is measurable or not is completely determined by whether it contains an outcomes ω such that $\{\omega\}$ is not measurable. There doesn't seem to be a problem at this stage. However, take the case when $\Omega = \mathbb{R}$ and let Σ is a σ-algebra containing the intervals that is also closed under arbitrary unions. It can be easily seen that $\Sigma = 2^{\mathbb{R}}$. But the Vitali set V becomes measurable, contradicting Proposition 1.22. In other words, allowing closure under uncountable unions prevents us from defining the Lebesgue measure on $\mathbb{R}$. In fact one can show that the only such measure on $\Sigma^{(1)}$ is the zero measure.

2.2 Random variables, density functions and the push-forward measure

Definition (Push-forward of a measure)**.** Let $(\Omega_1, \Sigma_1, \mu)$ be a measure space and (Ω_2, Σ_2) be a measurable space. Let $X : \Omega_1 \to \Omega_2$ be measurable. The push-forward is of μ, denoted by $X_*\mu$ is the a function on Σ_2 such defined by

$$X_*\mu(E) = \mu(\{X \in E\}), \quad \text{for all } E \in \Sigma_2.$$

It is clear that $X_*\mu$ is actually a measure on (Ω_2, Σ_2). //

Theorem 2.1 (Change of variables)**.** *Let Let $(\Omega_1, \Sigma_1, \mu)$ be a measure space and (Ω_2, Σ_2) be a measurable space. If $X : \Omega_1 \to \Omega_2$ is measurable and $X_*(\mu)$ is the push-forward measure of X then for any measurable function $g : \Omega_2 \to \mathbb{R}$ we have*

$$\int_{\Omega_2} g dX_*(\mu) = \int_{\Omega_1} g \circ X d\mu.$$

Definition (Random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and equip $\mathbb{R}$ with the Borel σ-algebra. A random variable is a measurable function on $X : \Omega \to \mathbb{R}$.

(i) **(Expected Value)**. The expected value or mean of X is defined as

$$\boxed{\mathbb{E}[X] := \int_\Omega X d\mathbb{P}.}$$

(ii) **(Variance)** We define variance of X to be

$$\boxed{\mathrm{Var(X)} := \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right].}$$

(iii) **(Distribution Measure)** The distribution of measure of X is the push-forward measure

$$\boxed{\mathbb{P}^X := X_*\mathbb{P}.}$$

(iv) **(C.D.F)** The cumulative distribution function of X is the function $F : \mathbb{R} \to \mathbb{R}$ defined by

$$\boxed{F(x) := \mathbb{P}^X((-\infty, x]) = \mathbb{P}[X \leq x].}$$

(v) **(P.D.F)** If $\mathbb{P}^X << \lambda$ then the probability density function is the almost everywhere defined function

$$\boxed{f_X := \frac{d\mathbb{P}^X}{d\lambda}.}$$

Suppose now that Y is another random variable.

(vi) **(Covariance)**. The covariance of X and Y is

$$\boxed{\mathrm{Cov}(X, Y) := \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].}$$

(vii) **(Correlation)**. The correlation coefficient of X and Y is defined as

$$\boxed{\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.}$$

Note that all of the quantities above that involve expectation can very well be infinite. //

Proposition 2.2. *Let X be a random variable with μ_X and F_X defined as above. We have that*

(i) *F_X is increasing.*

(ii) *F_X satisfies the following limits.*

$$\lim_{x\to-\infty} F_X(x) = 0 \quad \textit{and} \quad \lim_{x\to\infty} F_X(x) = 1.$$

(iii) *The Lebesgue-Stieltjes measure induced by F_X is μ_X and*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x d\mu_X = \int_{-\infty}^{\infty} x f_X(x)dx.$$

(iv) *If F is differentiable then*

$$\frac{dF}{dx} = f_X.$$

Proof. ∎

Theorem 2.3 (Chebychev's inequality)**.** *Let X be a random variable with mean μ and variance σ^2. Then for any real positive constant k we have*

$$\mathbb{P}\left[|X-\mu| \geq k\sigma\right] \leq \frac{1}{k^2}.$$

Proof. We prove Markov's inequality first and use it to obtain our desired result. Markov's inequality states that if X is non-negative then

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

which follows from

$$\mathbb{E}[X] \geq \int_{\{X\geq a\}} X(\omega)d\mathbb{P}(\omega) \geq \int_{\{X\geq a\}} a\, d\mathbb{P}(\omega) = a\int_{\Omega} \mathbf{1}_{\{X\geq a\}}(\omega)d\mathbb{P}(\omega) = a\cdot\mathbb{P}[X\geq a].$$

Now we have

$$\mathbb{P}\left[|X-\mu| \geq k\sigma\right] = \mathbb{P}\left[(X-\mu)^2 \geq k^2\sigma^2\right] \leq \frac{\mathbb{E}\left[(X-\mu)^2\right]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2},$$

as desired. ∎

A random variable X induces a probability measure $\mathbb{P}^X$ on $\mathbb{R}$. This measure is referred to as a probability law on $\mathbb{R}$. In many situations it is natural to identify the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the space $(\mathbb{R}, \mathcal{B}, \mathbb{P}^X)$ using X. This happens when one wants to study the properties of X that are irrelevant of the nature of the sample space Ω. So instead of looking at X itself, we study the induced objects such as $\mathbb{P}^X$, F_X or (when it exists) f_X.

Definition (Commonly occurring random variables)**.** //

Definition (Random vector)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random vector $\mathbf{X}$ is a mapping $\mathbf{X} : \Omega \to \mathbb{R}^d$ that is measurable with respect to the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$. In particular, it is a vector $(X_1, \ldots, X_d)$ with each component being a random variable.

(i) **(Mean vector)** The mean of $\mathbf{X}$ is defined as

$$\mu := (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d]).$$

(ii) **(Covariance matrix)** The covariance matrix of $\mathbf{X}$ is defined as

$$\Sigma := \left[\mathrm{Cov}(X_i, X_j)\right]_{i,j=1}^{d}.$$

(iii) **(Joint distribution measure)** The distribution measure of $\mathbf{X}$ is defined as

$$\mathbb{P}^{\mathbf{X}} := \mathbf{X}_* \mathbb{P}.$$

(iv) **(Joint c.d.f)** The joint c.d.f of $\mathbf{X}$ is the function $F_{\mathbf{X}} : \mathbb{R}^d \to \mathbb{R}$ defined as

$$F_{\mathbf{X}}(x_1, \ldots, x_d) = \mathbb{P}^{\mathbf{X}}\left(\prod_{k=1}^{d}(-\infty, x_k]\right) = \mathbb{P}[X_1 \leq x_1, \ldots, X_d \leq x_d].$$

//

Theorem 2.4 (n-dimensional Chebychev's inequality)**.** *Let $X : \Omega \to \mathbb{R}^n$ be a random vector with mean μ and covariance matrix $C = [\mathrm{Cov}(X_i, X_j)]_{i,j=1}^{n}$. If C is positive definite then for any $k \in \mathbb{R}$,*

$$\mathbb{P}\left[\sqrt{(X-\mu)^T C (X-\mu)} \geq k\right] \leq \frac{N}{k^2}.$$

2.3 Conditioning over σ-algebras and independence

Definition (Conditional probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For any two events $A, B \in \mathcal{F}$ we define

$$\mathbb{P}\left[A \mid B\right] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

A and B are called independent if $\mathbb{P}\left[A \mid B\right] = \mathbb{P}(A)$. //

Definition (Independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two events A and B are called independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Let $\{\mathcal{G}_1, \cdots, \mathcal{G}_n\}$ be a collection of sub σ-algebras of $\mathcal{F}$. then we call this collection independent if for all $A_1 \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n$ we have

$$\mathbb{P}\bigg(\bigcap_{k=1}^{n} A_k\bigg) = \prod_{k=1}^{n} \mathbb{P}(A_k).$$

A sequence $\{\mathcal{G}_n\}$ of sub σ-algebras of $\mathcal{F}$ is called independent independent of $\mathcal{G}_{n+1}$ and $\sigma(\mathcal{G}_1 \cup \cdots \cup \mathcal{G}_n)$ are independent for all $n \in \mathbb{N}$. A sequence of random variables $\{X_n\}$ is called independent if $\sigma(X_n)$ is independent of $\sigma(X_1, \ldots, X_n)$ are independent for all $n \in \mathbb{N}$. //

Definition (Conditional expectation)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. The conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is defined as random variable having the following properties.

(i) $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.

(ii) For all $A \in \mathcal{G}$ we have that

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

Property (ii) is called partial averaging. //

Definition. Let X and Y be two random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the joint p.m.f f_{XY} exists. Define

$$f_{X|Y}(x|y) := \frac{f_{XY}(x, y)}{f_Y(y)},$$

as the conditional density of X given Y. //

Proposition 2.5. *If X and Y are two jointly distributed random variables then*

(i) *If X and Y are discrete then*

$$p_X(x) = \sum_{y \in X(\Omega)} p_{X|Y}(x|y) p_Y(y), \quad \forall x \in X(\Omega).$$

(ii) *If X and Y are continuous then*

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy.$$

2.4 Modes of convergence and fundamental theorems

Definition (Convergence in probability)**.** Let $\{X_n\}$ be a sequence of random variables on a sample space. If there is a random variable X such that for every $\epsilon > 0$ one has

$$\lim_{n\to\infty} \mathbb{P}\left[|X_n - X| \geq \epsilon\right] = 0,$$

then one says $\{X_n\}$ converges to X in probability. //

Theorem 2.6. *The function*

$$d(X, Y) = \mathbb{E}[\min(|X - Y|, 1)],$$

is complete metric on $L^0(\Omega)$ and $X_n \to X$ in probability if and only if $d(X_n, X) \to 0$.

Theorem 2.7 (Weak law of large numbers)**.** *Let $\{X_n\}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for some $\mu, \sigma \in \mathbb{R}$ we have $\mathbb{E}[X_n] = \mu$ and $\mathrm{Var}[X_n] = \sigma^2$ for all $n \in \mathbb{N}$. We have that*

$$\lim_{n\to\infty} \overline{X}_n = \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu, \quad \textit{in probability.}$$

Proof. Since the X_n's are independent then

$$\sigma_n^2 := \mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Let $\epsilon > 0$ be given. We have by Chebychev's inequality that

$$\mathbb{P}\left[|\overline{X}_n - \mu| \geq \epsilon\right] \leq \sigma_n^2 \epsilon^{-2} = \frac{\sigma^2}{n\epsilon^2},$$

which gives the desired result. ■

Definition (Convergence in distribution)**.** Let $\{X_n\}$ be a sequence of random variables and for each $n \in \mathbb{N}$ define $F_n := F_{X_n}$. Let X be a random variable with c.d.f $F := F_X$. We say $\{X_n\}$ converges to X in distribution if

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for all $x \in \mathbb{R}$ such that F is continuous at x. //

The proof of the following claim is trivial.

Theorem 2.8 (Continuous mapping theorem)**.** *Let $\{X_n\}$ be a sequence of random vectors in $\mathbb{R}^n$ and let $g : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function. If $\{X_n\}$ converges to X almost surely, in probability or in distribution then $\{g(X_n)\}$ converges to $g(X)$ in the same way $\{X_n\}$ converges to X.*

Theorem 2.9. *Suppose that $\{X_n\}$, $\{A_n\}$ and $\{B_n\}$ are sequences of random vectors in $\mathbb{R}^m$, $\mathbb{R}^m$ and $\mathbb{R}^{mn}$ respectively. Furthermore suppose that*

(i) *$\{X_n\}$ converges in distribution to X.*

(ii) *$\{A_n\}$ converges in probability to a random vector A.*

(iii) $\{B_n\}$ *converges in probability to a non-random vector* B.

Then we have

$$\lim_{n\to\infty} A_n X + B_n = AX + B, \quad \textit{in distribution}\ .$$

Theorem 2.10 (Central limit theorem)**.** *Let* $\{\mathbf{X}_n\}$ *be a sequence of independent random vectors in* $\mathbb{R}^d$ *with common mean vector* μ *and covariance matrix* Σ. *Then*

$$\sqrt{d}\left(\overline{\mathbf{X}}_n - \mu\right) \to N(0, \Sigma) \quad \textit{in distribution},$$

or equivalently

$$\sqrt{d} \cdot \Sigma^{-\frac{1}{2}}\left(\overline{\mathbf{X}}_n - \mu\right) \to N(0, I_d) \quad \textit{in distribution.}$$

Theorem 2.11 (Generalized Central Limit Theorem)**.** *Under the assumptions of Theorem 2.10, if* $f : \mathbb{R}^d \to \mathbb{R}^m$ *is a continuously differentiable function with Jacobian matrix* $J(\mathbf{x})$ *then*

$$\sqrt{n}\left(f(\overline{\mathbf{X}}_n) - f(\mu)\right) \to N\left(0, J(\mu)\, \Sigma\, J(\mu)^T\right) \quad \textit{in distribution.}$$

3 Vector Valued Measures and Measure Valued Random Variables

3.1 Bochner spaces: measurability, integration and duality

Based on [4].

3.2 Vector measures

3.3 Introduction to random measures

4 General Stochastic Processes and Martingales

4.1 Definitions and Properties

Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A stochastic process is a collection $\{X_t\}_{t\in T}$ of random variables $X_t : \Omega \to \mathbb{R}^n$ indexed by T. For fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called a path or a realization of the process. Consider the following notation.

(i) $X(t) = X(t, \cdot) = X_t$ is the random variable $\omega \mapsto X_t(\omega)$.

(ii) $X(\omega) = X(\cdot, \omega) = X_\omega$ is the path $t \mapsto X_t(\omega)$.

It will be often useful to view $\{X_t\}$ as a

1. function of two variables $X : T \times \Omega \to \mathbb{R}^n$ defined by $X(t, \omega) = X_t(\omega)$.
2. a random variable $X : \Omega \to (\mathbb{R}^n)^T$ such that $\omega \mapsto X(\omega)$.

We alternate freely between the above notation and viewpoints depending on convenience. //

Analogously to real valued random variables, a stochastic process X has distribution space as a $(\mathbb{R}^n)^T$ random variable. Indeed, consider first the Borel σ-algebra $\mathcal{B}$ generated by the product topology on $(\mathbb{R}^n)^I$. Sets of the form

$$S(t, U) := \{f \in (\mathbb{R}^n)^I : f(t) \in U\}, \quad U \text{ open in } \mathbb{R}^n,\ t \in T,$$

are a subbasis for the product topology and hence a basis for the product topology contains sets of the form,

$$\bigcap_{j=1}^{k} S(t_j, F_j) = \{f \in (\mathbb{R}^n)^I : f(t_1) \in F_1, \ldots, f(t_k) \in F_k\}, \quad F_j \in \mathcal{B}(\mathbb{R}^n). \tag{6}$$

It is clear that $\omega \to X(\cdot, \omega)$ is measurable from $(\Omega, \mathcal{F})$ to $((\mathbb{R}^n)^I, \mathcal{B})$. Hence we can define it's distribution measure

$$\mu_X(F) := \mathbb{P}[X \in F] = \mathbb{P}\{\omega \in \Omega : X(\cdot, \omega) \in F\} \text{ where } F \in \mathcal{B}((\mathbb{R}^n)^I).$$

For Borel sets B of the form (6) we have

$$\mu_X(B) = \mathbb{P}\big[\, X \in B \,\big] = \mathbb{P}\big(\{\omega \in \Omega : X(\cdot, \omega) \in B\}\big) = \mathbb{P}\left[X_{t_1}(\omega) \in F_1, \ldots, X_{t_k}(\omega) \in F_k\right].$$

Therefore we can identify Ω with a subset of $(\mathbb{R}^n)^I$ and X can be viewed as a probability measure or probability law μ_X on the measure space $((\mathbb{R}^n)^I, \mathcal{B})$. This is useful in studying properties of stochastic processes when the nature of sample space is not relevant.

Definition (Distinguishing between stochastic processes)**.** Let $\{X\}_{t\in T}$ and $\{Y\}_{t\in T}$ be two stochastic processes.

(i) The processes have the same finite dimensional distributions if for all $t_1, \ldots, t_k \in T$ the random vectors $(X_{t_1}, \ldots, X_{t_k})$ and $(Y_{t_1}, \ldots, Y_{t_k})$ are equal in distribution.

(ii) X is called a version of Y if for all $t \in T$ we have $\mathbb{P}[X(t) = Y(t)] = 1$.

(iii) The two processes are called indistinguishable if for almost all $\omega \in \Omega$ we have $X(\cdot,\omega) = Y(\cdot,\omega)$.

It is immediate that (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i). //

The next theorem allows us to construct stochastic processes form a probability law on $(\mathbb{R}^n)^T$ equipped with the Borel σ-algebra $\mathcal{B}$, given that the law satisfies natural consistency properties.

Theorem 4.1 (Kolmogorov's extension theorem)**.** *Suppose for every $k \in \mathbb{N}$ and every $t_1, \dots, t_k \in T$ we are given a measure $\nu_{t_1,\dots,t_k}$ on $\mathbb{R}^n$. Suppose furthermore that*

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

and for all $m \in \mathbb{N}$

$$\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\dots t_k,\dots,t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbb{R}^{mn}).$$

Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process X such that $t_1, \dots, t_k \in T$ we have

$$\mathbb{P}[X_{t_1} \in F_1, \dots, X_{t_k} \in F_k] = \nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k), \quad F_j \in \mathcal{B}(\mathbb{R}^n).$$

In other words, there is a probability measure μ_X on $(\mathbb{R}^n)^I$ such that if B is a Borel set in $(\mathbb{R}^n)^I$ of the form (6) *then*

$$\mu_X(B) = \nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k).$$

Proof. ■

It is often asked whether a stochastic process has continuous paths for almost all $\omega \in \Omega$. This requires a topological structure on T.

Definition (Continuity of a stochastic process)**.** Let T be any topological space and let $\{X_t\}_{t\in T}$ be a stochastic process.

(i) X continuous at $t_0 \in T$ if for almost all $\omega \in \Omega$ we have

$$\lim_{t\to t_0} X_t(\omega) - X_{t_0}(\omega) = 0.$$

It is continuous if the above holds for all $t_0 \in T$.

(ii) X is continuous in mean at t_0 if

$$\lim_{t\to t_0} \mathbb{E}[X_t - X_{t_0}] = 0.$$

It is continuous in mean if (iii) holds for all $t_0 \in t$.

(iii) continuous in probability at t_0 if $\lim_{t\to t_0} \mathbb{P}[X_t - X_{t_0}] = 0$.

(iv) continuous if (v) holds for all $t_0 \in T$.

(v) Feller continuous if for every Borel function φ the function $t \mapsto \mathbb{E}\left[\varphi\left(X_t\right)\right]$ is continuous.

Condition (ii) is often stated as X has continuous sample paths //

In the following we let $T = [0, \infty)$ and equip T with the Borel σ-algebra $\mathcal{B}(T)$.

Definition (Measurability and adaptability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Equip $T \times \Omega$ and $\mathbb{R}^n$ with the σ-algebras $\mathcal{B}(T) \otimes \mathcal{F}$ and $\mathcal{B}(\mathbb{R}^n)$ respectively. Suppose also we are given a collection $\{\mathcal{F}_t\}_{t \in T}$ of increasing sub σ-algebras of $\mathcal{F}$. A stochastic process $\{X_t\}_{t \in T}$ is called

(i) measurable if the function $X : T \times \Omega \to \mathbb{R}^n$ defined by $X(t, \omega) = X_t(\omega)$ is measurable.

(ii) adapted if for all $t \in T$ we have $\sigma(X_t) \subset \mathcal{F}_t$.

(iii) progressively measurable for each $t \in T$ we have that $X : T \times \Omega \to \mathbb{R}^n$ is measurable when the domain and target space are equipped with the σ-algebras $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ and $\mathcal{B}(\mathbb{R}^n)$ respectively.

We denote the natural filtration of X to be $\{\mathcal{F}_t^X\}$ where $\mathcal{F}_t^X = \sigma(X_s;\ 0 \leq s \leq t)$. //

Proposition 4.2 (Chung and Doob, 1965)**.** *If a stochastic process X is measurable and adapted to a filtration $\{\mathcal{F}_t\}$ then X is progressively measurable.*

Theorem 4.3 (Kolmogorov's continuity condition)**.** *Let T be a closed cube in $\mathbb{R}^n$. Suppose the stochastic process $\{X_t\}_{t \in T}$ satisfies the following condition: there are positive constants $C, p \in \mathbb{R}$ and $\gamma > N$ such that*

$$\mathbb{E}\left[|X_t - X_s|^p\right] \leq C|t - s|^\gamma, \quad \forall s, t \in T.$$

Then there is a continuous version of $\{X_t\}$. In addition, if we call call $\{\tilde{X}_t\}_{t \in T}$ this modification and θ is chosen so that $1 \leq \theta < (\gamma - N)/p$ then

$$\sup_{s \neq t} \frac{|X_s - X_t|}{|t - s|^\theta} \in L^p(\Omega).$$

Proof. ∎

4.2 Gaussian processes, tempered measures and white noise

This section is based on [10], [11]

Definition. A process $\{X_t\}_{t \in T}$ is called a Gaussian process if for all $t_1, \ldots, t_n \in T$ the random vector $(X(t_1), \ldots, X(t_k))$ is a k-dimensional Gaussian random vector vector. //

Definition. Let $(M, \mathcal{G}, \sigma)$ be a σ-finite measure space. A Gaussian process associated with $(M, \mathcal{G}, \sigma)$ is a stochastic process $\{W_A^{(\sigma)}\}_{A \in \mathcal{G}}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

(i) For all $A \in \mathcal{M}$ we have $\mathbb{E}[W_A] = 0$.

(ii) For every $A_1, \ldots, A_n \in \mathcal{M}$ the random variables $W_{A_1}, \ldots, W_{A_n}$ are jointly normally distributed with covariances $\mathbb{E}[W_{A_i} W_{A_j}] = \sigma(A_i \cap A_j)$.

//

Theorem 4.4. *Let $(M, \mathcal{G}, \sigma)$ be a σ-finite measure space. Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Gaussian process $\{W_A^{(\sigma)}\}_{A \in \mathcal{G}}$ on that space that is indexed by $\mathcal{G}$.*

Theorem 4.5. *Let σ be a tempered measure. Let $(\mathcal{S}', \mathscr{C})$ be the space of tempered distributions equipped with the cylinder algebra $\mathscr{C}$. Then there is a probability measure $\mathbb{P}^{(\sigma)}$ on $\mathcal{S}'$ and a real valued Gaussian process $\{W_\varphi^{(\sigma)}\}_{\varphi \in \mathcal{S}}$ such that for all $\varphi \in \mathcal{S}$ we have*

(i) $W_\varphi^{(\sigma)}(\xi) = \langle \xi, \varphi \rangle$ *for all* $\xi \in \mathcal{S}'$.

(ii) $\mathbb{E}[W_\varphi^{(\sigma)}] = 0$.

(iii) $\mathbb{E}[\exp(iW_\varphi^{(\sigma)})] = \exp\Big(-\frac{1}{2} \int_{\mathbb{R}} |\hat{\varphi}(u)|^2 d\sigma(u) \Big)$.

4.3 Lévy and jump processes

Definition. A stochastic process X is called a Lévy process if

(i) $X(0) = 0$ almost surely.

(ii) X has independent increments.

(iii) X has stationary increments.

(iv) X is stochastically continuous in the sense that $\lim_{s \to t} \mathbb{P}(|X(t) - X(s)| > \epsilon) = 0$.

//

Theorem 4.6. *Any Lévy process X then admits a cadlag modification.*

Proof. ∎

Theorem 4.7 (Lévy-Itô decomposition)**.** *Let X be a Lévy process. Then X admits a decomposition*

$$X(t) = \gamma t + \sigma B(t) + X^P(t) + X^M(t),$$

where

(i) $B(t)$ *is standard Brownian motion.*

(ii) $X^P(t)$ *is a compound Poisson process.*

(iii) $X^M(t)$ *square integrable pure jump process.*

Proof. ∎

4.4 Brownian motion

Let $\{B_t\}_{t \in I}$ be n-dimensional Brownian motion defined as a stochastic process whose finite dimensional distributions are given by

$$\nu_{t_1, \ldots, t_k}(F_1 \times \cdots \times F_k) = \int\limits_{F_1 \times \cdots \times F_k} \prod_{j=1}^{k} (2\pi \Delta t_j)^{-\frac{n}{2}} \exp\Big(- \frac{|x - x_j|^2}{2\Delta t_j} \Big) dx_1 \ldots dx_k.$$

Equivalently, one can define Brownian motion as a stochastic process $\{B_t\}_{t \in I}$ such that for any $0 \le t_1 < \cdots < t_n \le T$ we have that

$$B_{t_1}, \;\; B_{t_2} - B_{t_1}, \; \ldots \; , \;\; B_{t_n} - B_{t_{n-1}},$$

are independent normal random variables with $\mathbb{E}[B_{t_{i+1}} - B_{t_i}] = 0$ and $\mathrm{Var}[B_{t_{i+1}} - B_{t_i}] = t_{i+1} - t_i$.

We have all the following properties derived from either definitions.

There is a continuous version of Brownian motion.

B_t is a Gaussian process i.e $(B_{t_1}, \dots, B_{t_k})$ is multi-normal on $\mathbb{R}^{nk}$.

Each component of $B_t = (B_t^{(1)}, \dots, B_t^{(n)})$ is standard Brownian motion on $\mathbb{R}$.

$E[|B_t - B_s|^4] = n(n+2)|t-s|^2$.

$\langle B, B\rangle_t = \lim_{n\to\infty} \sum_{t_i \in \Pi_n} |B_{t_i} - B_{t_{i-1}}|^2 = t$ almost surely. +

For any $t_0 \geq 0$, $\{B_{t_0+t} - B_{t_0}\}$ is Brownian motion.

If $UU^T = I$, then $\{UB_t\}$ is a Brownian motion.

For $c \in \mathbb{R}$, $\{c^{-1}B_{c^2t}\}$ is also a Brownian motion.

$E[\exp(\lambda(B_s - B_t))] = \exp(\lambda^2(s-t)/2)$.

If $\{B_t\}$ is standard one dimensional Brownian motion then $\int_0^t B_s dB_s = \frac{1}{2}B_t - t$.

For all $m > 0$ and $w \leq m$ we have $\mathbb{P}\{\tau_m \leq t,\, B_t \leq w\} = \mathbb{P}\{B_t \geq 2m - w\}$.[2]

We have $f_{\tau_m}(t) = |m|(2\pi)^{-1/2}t^{-3/2}\exp(-m^2/2t)$.

We have the joint density $f_{M_t,B_t}(m,w) = 2(2m-w)(2\pi)^{-1/2}t^{-3/2}\exp(-(2m-w)^2/2t)$.

4.5 Martingales

Lemma 4.8 (Doob's martingale inequality). *Let $\{M_t\}_{t\in I}$ be a continuous martingale. Then for all $p \geq 1$ and $T \geq 0$ then*

$$\mathbb{P}\left[\sup_I M(t) \geq \lambda\right] \leq \left(\frac{\|M(T)\|_{L^p(\Omega)}}{\lambda}\right)^p = \frac{1}{\lambda^p}\mathbb{E}[M_T^p].$$

Proof. ■

4.6 Markov and Feller processes

[2]Hitting time $\tau_m(\omega) = \inf\{t \geq 0 : B_t(\omega) = m\}$.

5 Itô Integral and Stochastic Partial Differential Equations

5.1 Integration with respect to Brownian motion

Definition. Let $I = [0, T]$ be an interval and let $\{X_t\}_{t \in I}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the following holds.

(i) $X : I \times \Omega \to \mathbb{R}^n$ is measurable where I is equipped with the Borel σ-algebra.

(ii) $\{X_t\}$ is adapted to the natural filtration of $\{B_t\}$.

(iii) For fixed $\omega \in \Omega$, if we denote $X(\omega, \cdot) : I \to \mathbb{R}$ the function $t \mapsto X(\omega, t)$, then the random variable $Y(\omega) = \|X(\omega, \cdot)\|_{L^2(I)}$ is in $L^1(\Omega)$.

Denote the space of such stochastic processes as $\mathcal{V} = \mathcal{V}(I)$. It can be shown that there is a sequence of simple processes $\{\Delta_n\} \subset \mathcal{V}$ such that

$$\lim_{n\to\infty} \int_\Omega \int_0^T \big(X(\omega, t) - \Delta_n(\omega, t)\big)^2 dt\, d\omega = 0, \quad \text{ie} \quad \mathbb{E}\big[\|X - \Delta_n\|_{L^2(I)}\big] \to 0.$$

The Itô isometry for simple processes says that

$$\mathbb{E}\left[\left(\int_0^T \Delta_n(t) - \Delta_m(t) dB_t\right)^2\right] = \mathbb{E}\left[\int_0^T (\Delta_n(t) - \Delta_m(t))^2 dt\right],$$

and therefore $\big\{\int_0^T \Delta_n dB_t\big\}$ is a Cauchy sequence in $L^2(\Omega)$. Hence, we can define the Itô integral of $\{X_t\}$ as

$$\int_0^T X(\cdot, t) dB_t := \lim_{n\to\infty} \int_0^T \Delta_n(\cdot, t) dB_t \in L^2(\Omega).$$

//

Lemma 5.1. *The definition of $\mathcal{I}$ is independent of the choice of simple processes. More precisely, if $\{\Delta_n\}$ is any sequence of simple process satisfying the requirements of the above definition then*

$$\left\|\int_0^T \Delta_n(t) dB_t - \int_0^T X(t) dB_t\right\|_{L^2(\Omega)} \to 0.$$

Lemma 5.2. *The Itô integral $\mathcal{I}(\omega, t)$ of a simple process $\Delta(\omega, t)$ is a continuous Martingale.*

Proof. For all $\omega \in \Omega$ and $t \in I$ write

$$\Delta(t, \omega) = \sum_k e_k(\omega) \cdot \mathbf{1}_{[t_k, t_{k+1})}(t), \quad \mathcal{I}(t, \omega) = \sum_{t_k \le t} e_k(\omega) \big(B(t_{k+1}, \omega) - B(t_k, \omega)\big).$$

Assume that $t \in (t_k, t_{k+1})$ for some k and let $h \in \mathbb{R}$ such that $t + h \in (t_k, t_{k+1})$

$$|\mathcal{I}(t + h, \omega) - \mathcal{I}(t, \omega)| = |e_k(\omega)| \left|B(t + h, \omega) - B(t, \omega)\right|,$$

and therefore $I(\cdot,\omega)$ is continuous at t. On the other hand,

$$|\mathcal{I}(t_k+h,\omega)-\mathcal{I}(t_k,\omega)| = \begin{cases} |e_{k-1}||B(t_k+h,\omega)-B(t_k,\omega)| & \text{if } h<0, \\ |e_k||B(t_k+h,\omega)-B(t_k,\omega)| & \text{if } h>0. \end{cases}$$

Therefore, $\mathcal{I}(\cdot,\omega)$ is continuous at t_k.

Now $\mathcal{I}$ is a Martingale. Indeed, let $t \in I$ and $h>0$ and let ℓ be the index for which $t \in [t_\ell, t_{\ell+1})$ then we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{I}(t+h) \mid \mathcal{F}_t\right] &= \mathbb{E}\left[\int_0^t \Delta(s)dB_s + \int_t^{t+h} \Delta(s)dB_s \,\middle|\, \mathcal{F}_t\right] \\
&= \int_0^t \Delta(s)dB_s + \mathbb{E}\left[\sum_{t\le t_k\le t_{k+1}\le t+h} e_k\left(B(t_{k+1})-B(t_k)\right) \,\middle|\, \mathcal{F}_t\right] \\
&= \int_0^t \Delta(s)dB_s + \mathbb{E}\left[e_\ell\left(B(t_{\ell+1})-B(t)\right)\middle|\mathcal{F}_t\right] \\
&\qquad + \sum_{t_{\ell+1}\le t_k\le t+h} e_k\mathbb{E}\left[B(t_{k+1})-B(t_k)\right] \\
&= \int_0^t \Delta(s)dB_s = \mathcal{I}(t),
\end{aligned}$$

as desired. ■

Theorem 5.3 (Properties)**.** *The Itô integral $\mathcal{I}(\omega,t)$ of a stochastic process $X(\omega,t)$ that is adapted to the natural filtration $\{\mathcal{F}_t\}_{t\in I}$ of Brownian motion satisfies the following.*

(i) *For each $t\in I$, $\mathcal{I}(t,\cdot)$ is $\mathcal{F}_t$-measurable.*

(ii) *$\mathcal{I}$ satisfies the Itô isomerty.*

(iii) *$\mathcal{I}$ is a Martingale.*

(iv) *Almost all paths $\mathcal{I}(\cdot,\omega)$ can be chosen to be continuous.*

(v) *The quadratic variation of the Itô process is given by*

$$[\mathcal{I},\mathcal{I}](t,\omega) = \int_0^t X(t,\omega)dB_t(\omega).$$

Proof. Part (i)-(iii) follow directly from the fact that $\mathcal{I}(t)$ is a pointwise limit of $\mathcal{F}_t$-measurable functions. For part (iv), let $\{\Delta_n\}$ be a sequence of simple processes such that

$$\lim_{n\to\infty} \|X-\Delta_n\|_{L^2(I),L^1(\Omega)} := \lim_{n\to\infty}\int_\Omega\int_I |\Delta_n(\omega,t)-X(\omega,t)|^2 dt\, d\omega = 0.$$

We will show that there is a subsequence $\{\Delta_{n_k}\}$ such that for almost all $\omega\in\Omega$

$$\|\mathcal{I}_{n_{k+1}}(\omega,\cdot)-\mathcal{I}_{n_k}(\omega,\cdot)\|_{L^\infty(I)} = \sup_{t\in[0,T]}\left|\int_0^t \Delta_{n_{k+1}}(s,\omega)dB_s - \int_0^t \Delta_{n_k}(s,\omega)dB_s\right| \to 0.$$

Hence for almost all $\omega \in \Omega$, the sequence of continuous functions $\{\mathcal{I}_{n_k}(\omega, \cdot)\}_{k\in\mathbb{N}}$ is Cauchy in $L^\infty(I)$ and therefore converges to a continuous element $\mathcal{J}(\omega, \cdot) \in L^\infty(I)$, which by Lemma 4.1 will be almost surely equal to $\mathcal{I}(\omega, \cdot)$.

By Doob's martingale inequality applied on $\mathcal{I}_n - \mathcal{I}_m$ with $p = 2$ we have that for any $\epsilon > 0$ that

$$\begin{aligned}\mathbb{P}\left[\|\mathcal{I}_n(\cdot,\omega) - \mathcal{I}_m(\cdot,\omega)\|_{L^\infty(I)} \geq \epsilon\right] &\leq \frac{1}{\epsilon^2}\|\mathcal{I}_n(T) - \mathcal{I}_m(T)\|^2_{L^2(\Omega)} \\ &= \frac{1}{\epsilon^2}\mathbb{E}\left[\int_0^T \left(\Delta_n(t) - \Delta_m(t)\right)^2 dt\right] \quad \text{(by Itô isometry)} \\ &= \frac{1}{\epsilon^2}\|\Delta_n - \Delta_m\|_{L^2(I),L^1(\Omega)} \xrightarrow{m,n\to\infty} 0.\end{aligned}$$

Therefore there is a subsequence $\{\Delta_{n_k}\}$ such that

$$\mathbb{P}\left[\left\|\mathcal{I}_{n_{k+1}} - \mathcal{I}_{n_k}\right\|_{L^\infty(I)} \geq 2^{-k}\right] < 2^{-k},$$

so that by the Borel-Cantelli lemma

$$\mathbb{P}\left\{\omega \in \Omega : \|\mathcal{I}_{n_{k+1}}(\cdot,\omega) - \mathcal{I}_{n_k}(\cdot,\omega)\|_{L^\infty(I)} \geq 2^{-k} \text{ for infinitely many } k\right\} = 0.$$

Therefore for almost all $\omega \in \Omega$ there is an integer N_ω such that for all $k \geq N_\omega$ we have

$$\|\mathcal{I}_{n_{k+1}}(\omega,\cdot) - \mathcal{I}_{n_k}(\omega,\cdot)\|_{L^\infty(I)} < 2^{-k},$$

and thus $\{\mathcal{I}_{n_k}(\cdot,\omega)\}$ is Cauchy for almost all ω as desired. ■

5.2 Weakening defining conditions for Itô integral

Definition. Let $I = [0, T]$ be an interval and $\mathcal{W} = \mathcal{W}(I)$ be the set of all stochastic processes $X(t, \omega)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the following conditions.

(i) $X : I \times \Omega \to \mathbb{R}$ is measurable.

(ii) There is filtration $\{\mathcal{H}_t\}$ such that $B(t, \omega)$ is Martingale with respect to this filtration and $X(t, \cdot)$ is $\mathcal{H}_t$-adapted.

(iii) For almost all $\omega \in \Omega$, $X(\cdot, \omega) \in L^2(I)$.

For such functions one can show that there is a sequence of simple processes $\{\Delta_n\} \subset \mathcal{W}$ that converge to X in probability for each $t \in [0, T]$. By defining the integral of simple processes in the usual way, defines

$$\int_0^T X(t)dB_t := \lim_{n\to\infty} \int_0^T \Delta_n(t)dB_t \quad \text{in probability.}$$

//

Theorem 5.4. *The Itô integral of functions in $\mathcal{W}(I)$ has the same properties of the integral for functions $\mathcal{V}(I)$, except that it's not a Martingale but rather a local Martingale.*

5.3 Itô process and Itô-Doeblin Formula

Definition (Itô Process)**.** Let $\{\mathcal{F}_t\}$ be the natural filtration for Brownian motion. Suppose there are adapted processes $\{\alpha(t,\cdot)\}$ and $\{\sigma(t,\cdot)\}$ and a stochastic process $\{X(t,\cdot)\}$ such that

$$X(t) = X(0) + \int_0^t \alpha(s,\omega)ds + \int_0^t \sigma(s,\omega)dB_s.$$

Then X is called an Itô process. //

Proposition 5.5 (Quadratic variation of Itô process)**.** *If X is an Itô process then*

$$[X,X](t,\omega) = \int_0^t \sigma(s,\omega)dB_s(\omega).$$

Theorem 5.6 (Itô-Doeblin Formula)**.** *Let $\{X(t,\cdot)\}$ be an Itô process and $g \in C^2([0,\infty)\times\mathbb{R})$. Define the stochastic process $Y(t,\omega) := g(t, X(t,\omega))$. Then $\{Y(t,\cdot)\}$ is an Itô process and for all $T \geq 0$*

$$Y(T) - Y(0) = \int_0^T \frac{\partial g}{\partial t}(t,X(t))dt + \int_0^T f_x(t,X(t))dX_t + \frac{1}{2}\int_0^T \frac{\partial^2 f}{\partial x^2}(t,X(t))d[X,X](t).$$

5.4 SDE's and the Markov property

Definition. Let $I \subset \mathbb{R}^+$ be any interval (possibly unbounded) and $t_0 = \inf I$. Let $\mu, \sigma : I \times \mathbb{R} \to \mathbb{R}$ be meaurable functions. A first order linear stochastic differential equation is an equation of the form

$$\begin{cases} dX(u) = \mu(u,X(u))du + \sigma(u,X(u))dW(u), & u \in I, \\ X(t_0) = x, & x \in \mathbb{R}. \end{cases} \tag{7}$$

The function μ is called the *drift*, the function $\sigma^2/2$ is called the *diffusion* and $X(t) = x$ is called an initial condition. //

Definition. A stochatic process X satisfying (7) is called a strong solution if

(i) X is adapted to W.

(ii) $\mu(t, X_t) \in L^1(I)$ almost surely.

(iii) $\sigma(t, X_t) \in L^2(I)$ almost surely.

(iv) X is continuous almost surely.

//

Proposition 5.7 (Uniqueness of strong solutions)**.** *Suppose that the functions μ and σ are Lipschitz. If X and Y are two strong solutions to* (7) *then X and Y are indistinguishable.*

Proof. For all $t \in I$ we have that

$$X(t) - Y(t) = \int_{t_0}^t [\mu(u,X(u)) - \mu(u,Y(u))]du + \int_{t_0}^t [\sigma(u,X(u)) - \sigma(u,Y(u))]dW(u).$$

Squaring both sides we get

$$(X(t)-Y(t))^2 \leq 2\left(\int_0^t \mu(u,X(u))-\mu(u,Y(u))du\right)^2 + 2\left(\int_0^t \sigma(u,X(u))-\sigma(u,Y(u))dW(u)\right)^2.$$

Taking expectations we obtain

$$\begin{aligned}\mathbb{E}\left[(X(t)-Y(t))^2\right] \leq 2\mathbb{E}&\left[\left(\int_0^t \mu(u,X(u))-\mu(u,Y(u))du\right)^2\right]\\ &+ 2\mathbb{E}\left[\left(\int_0^t \sigma(u,X(u))-\sigma(u,Y(u))dW(u)\right)^2\right].\end{aligned}$$

On one hand, we have by the Cauchy-Shwarz inequality that

$$\left(\int_0^t \mu(u,X(u))-\mu(u,Y(u))du\right)^2 \leq T\int_0^t \big(\mu(u,X(u))-\mu(u,Y(u))\big)^2 du,$$

and therefore by taking expectations and using the Lipschitz continuity of μ we get

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t \mu(u,X(u))-\mu(u,Y(u))du\right)^2\right] &\leq t\int_0^t \mathbb{E}\left[\big(\mu(u,X(u))-\mu(u,Y(u))\big)^2\right] du\\ &\leq tK\int_0^t \mathbb{E}\left[\big(X(u)-Y(u)\big)^2\right] du.\end{aligned}$$

On the other hand, by the Itô isometry we have that

$$\left(\int_0^t \sigma(u,X(u))-\sigma(u,Y(u))dW(u)\right)^2 = \int_0^t \big(\sigma(u,X(u))-\sigma(u,Y(u))\big)^2 du,$$

and therefore by the taking expectations and using the Lipschitz continuity of σ we get

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^t \sigma(u,X(u))-\sigma(u,Y(u))dW(u)\right)^2\right] &= \int_0^t \mathbb{E}\left[\big(\sigma(u,X(u))-\sigma(u,Y(u))\big)^2\right] du\\ &\leq M\int_0^t \mathbb{E}\left[(X(u)-Y(u))^2\right] du.\end{aligned}$$

All of the above imply that for all $t\in[0,T]$ we have

$$\mathbb{E}\left[(X(t)-Y(t))^2\right] \leq 2(TK+M)\int_0^t \mathbb{E}\left[(X(u)-Y(u))^2\right] du.$$

Therefore, by Gronwall's inequality we have

$$\mathbb{E}\left[\big(X(t)-Y(t)\big)^2\right] = 0, \quad \text{for all } t\in[0,T].$$

This means $X(t)=Y(t)$ almost surely. Now let

$$F = \big\{\omega\in\Omega : X_r(\omega)=Y_r(\omega),\ \forall r\in\mathbb{Q}\cap[0,T]\big\}.$$

For each $r \in \mathbb{Q} \cap [0,T]$ we have that the event

$$E_r = \{\omega \in \Omega : X_r(\omega) \neq Y_r(\omega)\},$$

has probability 0. This means that

$$\mathbb{P}(F) = \mathbb{P}\left(\Omega \ \setminus \bigcup_{r \in \mathbb{Q} \cap [0,T]} E_r\right) = 1.$$

This means that almost surely $X(r) - Y(r) = 0$ for all $r \in \mathbb{Q} \cap [0,T]$. Since for almost all $\omega \in \Omega$ we have $X(\cdot,\omega)$ and $Y(\cdot,\omega)$ are continuous then by density we have that almost surely $X(t) - Y(t) = 0$ and thererfore X and Y are indistinguishable. ■

Definition (Markov property)**.** Let $I = [t,T]$ and $x \in \mathbb{R}$ be given. Suppose X is a stochastic process that solves (7) with initial condition $X(t) = x$. Let $h : \mathbb{R} \to \mathbb{R}$ be a Borel function. //

5.5 Feynman-Kac and Fokker-Planck Equations

Theorem 5.8 (Fokker-Planck)**.** *Let $\{X_t\}$ be a real valued stochastic process satisfying the following stochastic differential equation*

$$dX(u) = \mu(u, X(u)) + \sigma(u, X(u))dW(u),$$

where W is any Weiner process. If $p(x,t)$ is the p.d.f of $X(t)$ then

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[\mu(x,t)p(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x,t)p(x,t)].$$

Theorem 5.9 (Feynman-Kac)**.** *Let X be a real valued stochastic process satisfying the following stochastic differential equation*

$$dX(u) = \beta(u, X(u)) + \gamma(u, X(u))dW(u).$$

Let h be a real Borel function. Fix $T > 0$ and let $t \in [0,T]$. Let

$$g(x,t) = \mathbb{E}^{t,x}\left[h(X(T))\right] = \mathbb{E}\left[h(X(T)) \mid X(t) = x\right].$$

Then the function g satisfies the following

$$\begin{cases} \dfrac{\partial g}{\partial t} = -\beta\dfrac{\partial g}{\partial x} - \dfrac{1}{2}\gamma^2\dfrac{\partial^2 g}{\partial x^2}, & \textit{for all } (x,t) \in \mathbb{R} \times [0,T], \\ g(x,T) = h(x), & \textit{for all } x \in \mathbb{R}. \end{cases} \tag{8}$$

5.6 Itô-Doeblin formula for jump processes

5.7 Examples from finance and economics

5.8 *Functional Itô calculus and stochastic integral representation of martingales

Part II

Functional Analysis and Partial Differential Equations

6 Classical theory and fundamental equations

6.1 Fundamental existence theorems

6.2 Poisson equation

6.3 Diffusion equation

6.4 Wave equation

7 Hilbert Spaces

7.1 Elementary properties

7.2 Lax-Milgram

7.3 Reproducing kernel Hilbert spaces

8 Sobolev Spaces

8.1 Defintion, characterization, completeness and duality of $W^{m,p}(\Omega)$

8.2 Sobolev embeddings

8.3 The trace operator and fractional Sobolev spaces

8.4 Weak formulation of boundary value problems

8.5 *Weighted Sobolev Spaces and Non-Linear Potential Theory

8.6 *Sobolev spaces on manifolds

Part III

Special Topics

9 Stochastic Heat Equation

10 Stochastic Wave Equation

11 The Classical and Stochastic Hasegawa-Mima equation

12 Stochastic Integration in UMD spaces

13 Miscellaneous Remarks and Observations

Understanding $dB_t \cdot dB_t = dt$ and $dB_t dt = 0$. Not to be understood in the sense of

$$(B_{t_{j+1}} - B_{t_j})^2 \simeq t_{j+1} - t_j.$$

First of all, we have

$$\begin{aligned}
\Big|\sum_{j=1}^{n-1}(B_{t_{j+1}} - B_{t_j})(t_{j+1} - t_j)\Big| &\leq n \cdot \sup_{0\leq j\leq n-1} |B_{t_{j+1}} - B_{t_j}||t_{j+1} - t_j| \\
&\leq T \cdot \sup_{0\leq j\leq n-1} |B_{t_{j+1}} - B_{t_j}|,
\end{aligned}$$

which goes to zero since B_t is continuous. Second,

$$[B, B](t) = \lim_{|\Pi|\to 0} \sum_{j=0}^{n-1}(B_{t_{j+1}} - B_{t_j})^2 = t \quad \text{(a.s)},$$

which follows from the inequality

$$\left\| t - \sum_{j=0}^{n-1}(B_{t_{j+1}} - B_{t_j})^2 \right\|_{L^2(\Omega)} \leq 2|\Pi|t.$$

If $\Pi_n = \{0, t/n, 2t/n, \ldots, t\}$ and we define

$$Z_{j+1} = \frac{B_{t_{j+1}} - B_{t_j}}{\sqrt{t_{j+1} - t_j}} = \sqrt{\frac{n}{t}}(B_{t_{j+1}} - B_{t_j}),$$

then it can be shown using the law of large numbers on the independent random variables $\{Z_{j+1}^2\}$ with common mean $\overline{\mu} = 1$ that

$$\frac{1}{t}\sum_{j=0}^{n-1}(B_{t_{j+1}} - B_{t_j})^2 = \sum_{j=0}^{n-1} \frac{Z_{j+1}^2}{n} \longrightarrow \overline{\mu} = 1, \quad \text{(a.s)}.$$

Remark on definition of Itô integral.

References

[1] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, ser. Hochschultext / Universitext. Springer, 2003, ISBN: 9783540047582. [Online]. Available: https://books.google.com.lb/books?id=kXw9hB4EEpUC.

[2] S. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*, ser. Springer Finance Textbooks v. 11. Springer, 2004, ISBN: 9780387401010. [Online]. Available: https://books.google.com.lb/books?id=O8kD1NwQBsQC.

[3] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, ser. Universitext. Springer New York, 2010, ISBN: 9780387709130. [Online]. Available: https://books.google.com.lb/books?id=GAA2XqOIIGoC.

[4] T. Hytonen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces, Volume I: Martingales and Littlewood-Paley Theory*. Dec. 2016, ISBN: 978-3-319-48519-5 (print), 978-3-319-48520-1 (online). DOI: 10.1007/978-3-319-48520-1.

[5] H. Holden, B. Oksendal, and J. Uboe, *Stochastic Partial Differential Equations*. Birkhauser Boston, 2014, ISBN: 9781468492163. [Online]. Available: https://books.google.com.lb/books?id=IBsgswEACAAJ.

[6] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, ser. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2004, ISBN: 9783540643258. [Online]. Available: https://books.google.com.lb/books?id=1ml95FLM5koC.

[7] I. Karatzas, I. Shreve, S. Shreve, and S. Shreve, *Brownian Motion and Stochastic Calculus*, ser. Graduate Texts in Mathematics (113) (Book 113). Springer New York, 1991, ISBN: 9780387976556. [Online]. Available: https://books.google.com.lb/books?id=ATNy%5C_Zg3PSsC.

[8] R. Bartle and K. M. R. Collection, *The Elements of Integration*, ser. Wiley Classics Library. Wiley, 1966. [Online]. Available: https://books.google.com.lb/books?id=UrLvAAAAMAAJ.

[9] J. Walsh, *Knowing the Odds: An Introduction to Probability*, ser. Graduate studies in mathematics. American Mathematical Society, 2012, ISBN: 9780821890325. [Online]. Available: https://books.google.com.lb/books?id=4uC0uEXpvyoC.

[10] R. Dalang, D. Khoshnevisan, F. Rassoul-Agha, C. Mueller, D. Nualart, and Y. Xiao, *A Minicourse on Stochastic Partial Differential Equations*, ser. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2008, ISBN: 9783540859932. [Online]. Available: https://books.google.com.lb/books?id=EBeVxgEACAAJ.

[11] D. Alpay, P. Jorgensen, and D. Levanony, *On the equivalence of probability spaces*, 2016. arXiv: 1601.00639 [math.PR].