Stochastic Mathematics in Finance

Nanyi, UIBE

October 29, 2022

Contents

Foreword

Chapter 1

Set Theory

1.1 Zermelo Fraenkel set theory

1.2 Logit gate

Definition 1.1 (Family).
Suppose Γ is an index set, then a family $\{e_k\}_{k \in \Gamma}$ in a set V is a function from Γ to V.

$$e : \Gamma \to V, \ k \mapsto e_k.$$

Definition 1.2 (Union, Intersection, Complement).
Suppose $A_k \subset \mathbf{X}$. Then the union of A_k is defined by

$$\bigcup_{k \in \Gamma} A_k := \{x : \exists k_0 \in \Gamma, x \in A_{k_0}\}$$

$$\bigcap_{k \in \Gamma} A_k := \{x : \forall k \in \Gamma, x \in A_k\}$$

$$A^c := \{x :\notin A\}$$

$$A \backslash B := \{x : x \in A, \ x \notin B\}$$

Definition 1.3 (Map).
Suppose X and Y are sets. A map f from X to Y is a rule, which can be written by

$$f : X \to Y, \ x \mapsto f(x),$$

where $x \in X$ and $f(x) \in Y$.

Definition 1.4 (injection, bijection, surjection).
Suppose f is a map from X to Y. Then f is an injection if for all

$$f(x_1) = f(x_2) \implies x_1 = x_2.$$

And f is call a bijection if

$$\text{range} f = Y.$$

Definition 1.5.

Definition 1.6 (null space and range).

Theorem 1.1 (De moivre).

$$(\bigcup_{k \in \Gamma} A_k)^c = \bigcap_{k \in \Gamma} A_k^c$$

Definition 1.7 (Collection of set).
The set $\mathcal{F}$ whose elements are subsets of X is called a collection of set on X.

Definition 1.8. Suppose $\mathcal{F}$ is a collection of set on X. Then

- $\mathcal{F}$

1.3   Binary Structure

Definition 1.9 (Binary operation).
A binary operation on a nonempty set V is a map from $V \times V$ to V. In this case, the turple $(V, *)$ is called a binary structure.

Definition 1.10 (Semi-group).
A binary sturcture $(V, *)$ is called a semi-group if the operation $*$ is commutative. More precisely, $\forall a, b, c \in V$, we have

$$a * (b * c) = (a * b) * c.$$

In this case, the order of operation does not matter.

Definition 1.11 (Identity).
Suppose $(V, *)$ is a semi-group, then $e \in V$ is called an identity of $(V, *)$ if

$$e * a = a * e = a$$

holds for all $a \in V$.

Definition 1.12 (Monoid).
A semigroup that has an indentity is called a monoid.

Theorem 1.2. The identity on a monoid is unique.

Proof. Suppose $e, \tilde{e}$ are identities of a monoid $(S, *)$, then

$$e = \tilde{e} * e = e * \tilde{e} = \tilde{e}.$$

$\square$

Definition 1.13 (Inverse). Suppose $(V, *)$ is a monoid. For all $a \in V$, if there exists (another) element $\tilde{a} \in V$ such that

$$a * \tilde{a} = \tilde{a} * a = e,$$

then we say $\tilde{a}$ is the inverse of a.

Definition 1.14 (Group). A monoid $(V, *)$ is said to be a group if for all $a \in V$, there exists an inverse of a.

Theorem 1.3. Operation inverse is unique.

Definition 1.15 (Abelian Group). A group $(V, *)$ is Abelian if $\forall a, b \in V$,

$$a * b = b * a.$$

Example 1.1 (+ in python). Let V denote the class, list, in python. Then $(V, +)$ is a monoid.

Proof. Suppose $l_1, l_2 \in V$, then
$$l_1 = [a_1, \cdots, a_n],$$
$$l_2 = [b_1, \cdots, b_m].$$

And
$$l_1 + l_2 = [a_1, \cdots, a_n, b_1, \cdots, b_m].$$

Note that the identity is []. $\square$

Remark 1.

- (character,+) is a monoid.

- (list,+) is a monoid.

- (float,+) is a Abelian group.

- (integer,+) is a Abelian group.

Chapter 2

Banach Spaces

We begin this chapter with a quick review of the essentials of metric spaces. Then we extend our results on measurable functions and integration to complex-valued functions. After that, we rapidly review the framework of vector spaces, which allows us to consider natural collections of measurable functions that are closed under addition and scalar multiplication.

Normed vector spaces and Banach spaces, which are introduced in the third section of this chapter, play a hugely important role in modern analysis. Most interest focuses on linear maps on these vector spaces. Key results about linear maps that we develop in this chapter include the Hahn-Banach Theorem, the Open Mapping Theorem, the Closed Graph Theorem, and the Principle of Uniform Boundedness.

2.1   Metric Space

Definition 2.1 (Metric Space)**.**
A metric on a nonempty set V is a function $d : V \times V \to [0, \infty)$ such that

- $d(f, f) = 0$ for all $f \in V$;

- $d(f, g) = 0$ implies $f = g$;

- $d(f, g) = d(g, f$ for all $f, g \in V$;

- $d(f, h) \leq d(f, g) + d(g, h)$ for all $f, g, h \in V$.

A metric space is a pair (V, d), where V is a nonempty set and d is a metric on V.

Example 2.1 (Distance on $\mathbb{R}^n$)**.**
Suppose T is a positive self-adjoint operator. That is for all $f \in V$,

$$\langle Tv, v \rangle = \langle v, T^*v \rangle = \langle v, Tv \rangle > 0.$$

Then we can define a distance on $\mathbb{R}^n$ by:

$$d(x, y)^2 = \langle x - y, T(x - y) \rangle = (x - y)'\mathcal{M}(T)(x - y).$$

Example 2.2 (Minkowski Distance)**.**

Suppose $X_1, \cdots, X_n$ is a random sample, where $X_1 = \begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1n} \end{bmatrix}$ with mean μ and covariance

matrix Σ. If X_i are uncorrelated and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$, then the distance between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$d^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})' \begin{bmatrix} \sigma_{11}^{-1} & & & \\ & \sigma_{22}^{-1} & & \\ & & \ddots & \\ & & & \sigma_{pp}^{-1} \end{bmatrix} (\mathbf{x} - \mathbf{y}).$$

However, if X_i are correlated, we need to translate them into uncorrelated components and then use the above procedure to define the distance. To be specific, the real spectrum theorem yields that

$$\Sigma = P\Lambda P', where \ \Lambda = diag(\lambda_1, \cdots, \lambda_p).$$

Thus the distance of $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$
\begin{aligned}
d^2(\mathbf{x}, \mathbf{y}) &= d^2(P'(\mathbf{x} - \mu), P'(\mathbf{x} - \mathbf{y})) \\
&= (\mathbf{x} - \mathbf{y})' P \Lambda^{-1} P'(\mathbf{x} - \mathbf{y}) \\
&= (\mathbf{x} - \mathbf{y})' \Sigma^{-1}(\mathbf{x} - \mathbf{y})
\end{aligned}
$$

Definition 2.2 (Open ball).
Suppose (V, d) is a metric space, $f \in V$, and $r > 0$.

- The open ball centered at f with radius r is denoted by $B(f, r)$ and id defined by

$$
B(f, r) = \{g \in V : d(f, g) < r\}.
$$

- The closed ball centered at f with radius r is denoted by $\overline{B}(f, r)$ and is defined by

$$
\overline{B}(f, r) = \{g \in V : d(f, g) \le r\}.
$$

Definition 2.3 (Open subset).
A subset G of a metric space V is called open if for all $f \in G$, there exists $r > 0$ such that $B(f, r) \subset G$.

Definition 2.4 (Open subset).
A subset F of a metric space V is called close if $V \backslash F$ is open.

Definition 2.5 (closure).
Suppose V is a metric space and $E \subset V$. Then the closure of E, denoted $\backslash E$, is defined by

$$
\overline{E} = \{g \in V : B(g, \varepsilon) \cap E \ne \emptyset, \forall \varepsilon > 0\}
$$

Definition 2.6 (Limits in V). Suppose V is a metric space, $f_1, f_2, \cdots$ is a sequence in V, and $f \in V$. Then

$$
\lim_{k \to \infty} f_k = f \text{ means } \lim_{k \to \infty} d(f_k, f) = 0
$$

Theorem 2.1 (closure).
Suppose V is a metric space and $E \subset V$. Then

- $\overline{E} = \{\}$

Chapter 3

L^p Spaces

Fix a measure space $(X, \mathcal{S}, \mu)$ and a positive number p. We begin this chapter by looking at the vector space of measurable functions $f : X \to \mathbf{F}$ such that

$$\int |f|^p d\mu < \infty.$$

Important results called Holder's inequality and Minkowski's inequality help us investigate this vector space. A useful class of Banach spaces appears when we identify functions that differ only on a set of measure 0 and require $p \geq 1$.

Suppose $(\mathbf{X}, \mathcal{S}, \mu)$ is measure space, $0 < p < \infty$.

Definition 3.1.

Theorem 3.1. Suppose $\mathbb{S} = \{\varphi : \varphi \text{ is simple and } \varphi \in L^1(\mathbf{X}, \mathcal{S}, \mu)\}$. Then $\mathbb{S}$ is dense in $L^1(\mathbf{X}, \mathcal{S}, \mu)\}$.

Chapter 4

Hilbert Spaces

Normed vector spaces and Banach spaces, which were introduced in Chapter 1 , capture the notion of distance. In this chapter we introduce inner product spaces, which capture the notion of angle. The concept of orthogonality, which corresponds to right angles in the familiar context of $\mathbf{R}^2$ or $\mathbf{R}^3$, plays a particularly important role in inner product spaces.

Just as a Banach space is defined to be a normed vector space in which every Cauchy sequence converges, a Hilbert space is defined to be an inner product space that is a Banach space. Hilbert spaces are named in honor of David Hilbert (1862-1943), who helped develop parts of the theory in the early twentieth century.

In this chapter, we will see a clean description of the bounded linear functionals on a Hilbert space. We will also see that every Hilbert space has an orthonormal basis, which make Hilbert spaces look much like standard Euclidean spaces but with infinite sums replacing finite sums.

4.1   Inner product space

Definition 4.1 (Inner Product on a vector space).

An inner product is a function from $V \times V$ to $\mathbb{F}$ and has the following properties.

- Positivity: $\langle f, f \rangle \in [0, \infty)$ for all $f \in V$.

- Definiteness: $\langle f, f \rangle = 0$ implies $f = 0$.

- Linearity at first slot:

$$\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$$
$$\langle af, h \rangle = a \langle f, h \rangle$$

- Conjugate symmetry:
$$\langle f, g \rangle = \overline{\langle g, h \rangle}$$

$(V, \langle, \rangle)$ is called an inner product space.

Theorem 4.1 (Properties of inner product).

1. $\langle 0, g \rangle = \langle g, 0 \rangle = 0$.

2. $\langle f, g + h \rangle = \langle f, g \rangle + \langle f, h \rangle$.

3. $\langle f, ag \rangle = \overline{a} \langle f, g \rangle$.

Define $\|f\| = \sqrt{\langle f, f \rangle}$. and we say two elements of an inner product space are called orthogonal if their inner product equals 0.

Theorem 4.2 (Pythagorean Theorem/Cauchy-Schwarz/Triangle Inequality).

1. $\|f + g\|^2 = \|f\|^2 + \|g\|^2$ iff $\langle f, g \rangle = 0$.

2. $|\langle f, g \rangle| \le \|f\| \|g\|$.

3. $\|\|f + g\|\| \le \|f\| + \|g\|$.

4. $\|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2$.

A direact consequence is that $(V, \|\|)$ is a normed vector space. And the distance of two elements of V is defined by $d(f, g) = \|f - g\|$.

Definition 4.2 (Distance from a point to a set).
Suppose $f \in V$ and $U \subset V$. Then the distance between f and U is defined by

$$d(f, U) := \inf\{d(f, g) : g \in U\}.$$

Note also that $d(f, U) = 0$ if and only if $f \in \overline{U}$.

Proof. Recall that $\overline{U} = \{f \in V : B(f, \varepsilon) \cap E \neq \emptyset \text{ for all } \varepsilon > 0.\}$ Then

$$
\begin{aligned}
d(f, U) = 0 \quad \Longrightarrow \quad & \forall \varepsilon > 0, \exists g \in U, s.t. \|f - g\| < \varepsilon \\
\Longrightarrow \quad & \forall \varepsilon > 0, B(f, \varepsilon) \cap E \neq \emptyset. \\
\Longrightarrow \quad & f \in \overline{U}.
\end{aligned}
$$

And the other direction trivivally holds. $\qquad\square$

Example 4.1. Suppose T is a positive self-adjoint operator on $V = \mathbb{R}^n$, which implies the matrix of T is symmetric and positive defined. Then $d(u, v) = \langle Tu, v \rangle = v^T \mathcal{M}(T)u$ is a distance on V.

Definition 4.3. A complete inner product space is called a Hilbert space.

Theorem 4.3 (Distance to a closed convex set is attained in a Hilbert space).
Suppose V is a Hilbert space, U is a non-empty closed convex set in V. Then there exists a unique $g \in U$ such that
$$\|f - g\| = d(f, U)$$

Proof. By defition, there exists g_n such that

$$\lim_{n \to \infty} \|f - g_n\| = d(f, U)$$

Then for all m,n, we have

$$
\begin{aligned}
\|g_m - g_n\|^2 &= \|(f - g_n) - (f - g_m)\|^2 \\
&= 2\|f - g_n\|^2 + 2\|f - g_m\|^2 - \|2f - g_m - g_n\|^2 \\
&= 2\|f - g_n\|^2 + 2\|f - g_m\|^2 - 4\|f - (g_m + g_n)/2\|^2 \\
&\leq 2\|f - g_n\|^2 + 2\|f - g_m\|^2 - 4d(f, U).
\end{aligned}
$$

This implies g_n is a Cauchy sequence in V, and thus exists $g \in V$ such that

$$\lim_{n \to \infty} \|g_n - g\| = 0,$$

since V is complete. Also note that U is closed, we have $g \in U$. This implies $d(f-g) \geq d(f, U)$. On the other hand,

$$d(f, g) \leq d(f, g_n) + d(g_n, g),$$

where the right side converges to $d(f, U)$. As for the uniqueness,..   □

4.2   Orthogonality

Definition 4.4 (Orthogonal Projection).
Suppose U is a non-empty closed convex subset of a Hilbert space V. Then orthogonal projection of V onto U is the function

$$P_U : V \to U, \ f \mapsto g.$$

Where g is the only point that satisfies $d(f, g) = d(f, U)$. Theorem 3.3 admits our definition makes sense.

Remark 2.

- $P_U(f) = f$ iff $f \in U$.

- $P_U \circ P_U = P_U$.

- Suppose W is closed subspace and $W \subset U$, then $P_W = P_W \circ P_U$.

Definition 4.5 (Orthogonal Complement). Suppose U is a subset of an inner product space. Then
$$U^\perp := \{h \in V : h \perp U\}$$

Note that $h \perp U$ implies $f \perp g$ for all $g \in U$.

Theorem 4.4. Suppose U is a closed subspace(or complete subspace) of a Hilbert space V. Then

1. $f - P_U f$ is orthogonal to all $g \in U$.

2. If $h \in U$ and $f - h \in U^\perp$, then $h = P_U f$

3. $P_U : V \to U$ is a bounded linear map.

4. $\|P_U f\| \leq \|f\|$.

Proof. Suppose $g \in U$ and $a = -t\langle f - P_U f, g \rangle$ where $t > 0$. Then we have

$$
\begin{aligned}
\|f - P_U f\|^2 &\leq \|f - P_U f + ag\|^2 \\
&= \|f - P_U f\|^2 + |a|^2 \|g\|^2 + 2\mathrm{Re}\,\bar{a}\langle f - P_U f, g \rangle.
\end{aligned}
$$

The inequality implies

$$
2|\langle f - P_u, g \rangle|^2 \leq t|\langle f - P_u, g \rangle|^2 \|g\|^2.
$$

Thus $\langle f - P_U f, g \rangle = 0$.

(2) Suppose $g \in U$. Since U is a vector space, $g - h \in U$ and $\langle f - h, g - h \rangle = 0$, thus the triangle inequality implies

$$
\begin{aligned}
\|f - h\|^2 &\leq \|f - h\|^2 + \|h - g\|^2 \\
&= \|f - g\|^2.
\end{aligned}
$$

The inequality suggests that h is the closest point with respect to f in U and thus $h = P_u f$.

(3) Suppose $f_1, f_2 \in V$. Since

$$
\langle f_1 + f_2 - P_U f_1 - P_U f_2, g \rangle = 0
$$

for all $g \in U$. Now (2) implies that $P_U(f_1 + f_2) = P_U f_1 + P_U f_2$.

(4) At last, suppose $f = P_U f + h$.

$$
\langle P_U f, h \rangle = \langle P_U f, f - P_U f \rangle = 0,
$$

since $P_U f \in U$ and $f - P_U f \perp U$. Thus

$$
\|f\|^2 = \|P_U f\|^2 + \|h\|^2.
$$

$\square$

Theorem 4.5 (Properties of Orthogonal Complement)**.**

1. $U^\perp$ is a subspace of V.

2. $U \cap U^\perp \subset \{0\}$.

3. $W \subset U \implies U^\perp \subset W^\perp$.

4. $(\overline{U})^\perp = U^\perp$.

5. $U \subset (U^\perp)^\perp$.

Proof. $\square$

Theorem 4.6 (Orthogonal Decomposition)**.**
Suppose U is a closed subspace of a Hilbert space V(and thus complete). Then

$$V = U \oplus U^\perp.$$

In other words, $f \in V$ can can be uniquely written in the form

$$f = g + h, \ g \in U, \ h \in U^\perp.$$

Theorem 4.7 (Range and Null space of Orthogonal projections)**.**
Suppose U is a closed subspace of a Hilbert space V. Then

- $\text{range}P_U = U$ and $P_U = U^\perp$.

- $\text{range}P_{U^\perp}$

Theorem 4.8 (Riesz Reprentation Theorem I)**.** Suppose $\varphi \in B(V, \mathbb{F})$ where V is a Hilbert space. Then there exists a unique $h \in V$ such that

$$\varphi(f) = \langle f, h \rangle$$

for all $f \in V$. Moreover, $\|\varphi\| = \|h\|$.

Proof. Since $\text{null}\varphi \subset \{h\}^\perp$, which implies $h \subset (\text{null}\varphi)^\perp$. Thus we begin our proof with $(\text{null}\varphi)^\perp$. First suppose $\varphi \neq 0$, we claim that $\text{null}\varphi$ is a closed subspace of V and $\text{null}\varphi \neq V$ since φ is continuous. Thus there exists $g \in (\text{null}\varphi)^\perp$ with $\|g\| = 1$. Let

$$h = \overline{\varphi(g)}g.$$

with $\|h\|^2 = \langle h, h \rangle = |\varphi(g)|^2$. Thus

$$\varphi(h) = |\varphi(g)|^2 = \|h\|^2.$$

Now suppose $f \in V$,

$$
\begin{aligned}
\langle f, h \rangle &= \langle f - \frac{\varphi(f)}{\|h\|^2} h, h \rangle + \langle \frac{\varphi(f)}{\|h\|^2} h, h \rangle \\
&= \langle \frac{\varphi(f)}{\|h\|^2} h, h \rangle \\
&= \varphi(f),
\end{aligned}
$$

where the equality holds because $f - \frac{\varphi(f)}{\|h\|^2} h \in \text{null}\varphi$.

We have now proved the existence of h. To prove the uniqueness, suppose $\tilde{h} \in V$ has the same property. Then

$$
\langle h - \tilde{h}, h - \tilde{h} \rangle = \langle h - \tilde{h}, h \rangle - \langle h - \tilde{h}, h \rangle = \varphi(h - \tilde{h}) - \varphi(h - \tilde{h}) = 0,
$$

which implies that $h = \tilde{h}$, which proves the uniqueness.

The Cauchy-Schwarz inequality implies

$$
|\varphi(f)| = |\langle f, h \rangle| \le \|f\| \|h\|,
$$

and hence $\|\varphi\| \le \|h\|$. On the other hand,

$$
\varphi(h) = \|h\|^2,
$$

we also have $\|\varphi\| \ge \|h\|$. $\qquad \square$

Theorem 4.9 (Riesz Reprentation Theorem II)**.** Suppose $\varphi \in B(V, \mathbb{F})$ where V is a Hilbert space. Let

$$
h = \sum_{k \in \Gamma} \overline{\varphi(e_k)} e_k.
$$

Then

$$
\varphi(f) = \langle f, h \rangle
$$

for all $f \in V$. Furthermore, $\|\varphi\| = (\sum_{k \in \Gamma} |\varphi(e_k)|^2)^{\frac{1}{2}}$.

Proof. $\qquad \square$

4.3 Orthonormal Bases

Definition 4.6. $\{e_k\}_{k\in\Gamma}$ is called an orthonormal family if

$$\langle e_i, e_j \rangle = \delta_{ij}, i, j \in \Gamma.$$

Definition 4.7 (Unordered sum).

Suppose $\{f_k\}_{k\in\Gamma}$ is a family in a normed vector space V. The unordered sum $\sum_{k\in\Gamma}$ is called convergent if $\exists g \in V$, $\forall \varepsilon > 0$, $\exists \Omega \subset \Gamma$, for all $\Omega \subset \Omega' \subset \Gamma$, we have

$$\|g - \sum_{k\in\Omega'} f_k\| < \varepsilon$$

If this happens, we set $\sum_{s\in\Gamma} f_k = g$. Moreover, if $\forall N > 0$, $\exists \Omega \subset \Gamma$, for all $\Omega \subset \Omega' \subset \Gamma$, we have

$$\|\sum_{k\in\Omega'} f_k\| > N$$

We we say $\sum_{s\in\Gamma} f_k$ 'converges' to infinity.

Example 4.2 (Unordered sum in $\mathbb{R}$).

1. $\{a_k\}_{k\in\Gamma}$ where a_k is nonnegative.

$$\sum_{k\in\Gamma} a_k = \sup\{\sum_{k\in\Omega} a_k : \Omega \text{ is a finite subset of } \Gamma\}.$$

If $\sup\{\sum_{k\in\Omega} a_k : \Omega \text{ is a finite subset of } \Gamma\} < \infty$, then $\{a_k\}_{k\in\Gamma}$ is called convergent.

2. General case.

$$\sum_{k\in\Gamma} a_k \text{ converges } \iff \sum_{k\in\Gamma} |a_k| < \infty.$$

Theorem 4.10. Suppose V is a Hilbert space, then the inner product defined on V and the unordered sum operation are commutitative if the unordered sum makes sense.

Theorem 4.11 (Linear combination of an orthonormal family).

Suppose $\{e_k\}_{k\in\Gamma}$ is an orthonormal family in a Hilbert space V, and also suppose $\{a_k\}_{k\in\Gamma}$ is a family in $\mathbb{F}$. Then

- $\sum_{k\in\Gamma} a_k e_k$ *converges* $\iff (\sum_{k\in\Gamma} |a_k|^2)^{\frac{1}{2}} < \infty$

- Furthermore, if $\sum_{k\in\Gamma} a_k e_k$ *converges*, then $\|\sum_{k\in\Gamma} a_k e_k\|^2 = \sum_{k\in\Gamma} |a_k|^2$.

Proof. First suppose $\sum_{k\in\Gamma} a_k e_k = g$. Then for all $\varepsilon > 0$, there exists Ω such that for all $\Omega \subset \Omega' \subset \Gamma$ holds

$$\|g - \sum_{k\in\Omega'} a_k e_k\| < \epsilon.$$

This implies
$$\|g\| - \varepsilon < \|\sum_{k \in \Omega'} a_k e_k\| < \|g\| + \epsilon.$$

You should note that $\|\sum_{k \in \Omega'} a_k e_k\| = (\sum_{k \in \Omega'} |a_k^2|)^{\frac{1}{2}}$, since Ω' is a finite set. This inquality now implies that $\sum_{k \in \Gamma} |a_k^2|$ is convergent to $\|g\|^2$, in other words,

$$\|\sum_{k \in \Gamma} a_k e_k\| = \|g\| = (\sum_{k \in \Gamma} |a_k^2|)^{\frac{1}{2}}.$$

To prove the other direction of (i), now suppose $\sum_{k \in \Gamma} |a_k^2| < \infty$. Define

$$\sum_{\Gamma \backslash \Omega} |a_k|^2 := \sum_{k \in \Gamma} |a_k|^2 - \sum_{k \in \Omega} |a_k|^2, \text{ where Omega is a finite set.}$$

The definition above makes sense, as you should verify. By definition, there exists an incresing set $\Omega_1 \subset \Omega_2 \subset \cdots$, such that

$$\sum_{k \in \Gamma \backslash \Omega_m} |a_k|^2 < \frac{1}{m^2}.$$

Set $g_m = \sum j \in \Omega_m a_j e_j$, then

$$\|g_n - g_m\|^2 = \sum_{\Omega_n \backslash \Omega_m} |a_j|^2 < \frac{1}{m^2}.(why?)$$

This implies that g_n is a Cauchy sequence in V and thus converges to some element $g \in V$. To show that $\sum_{k \in \Gamma} a_k e_k = g$, suppose $\Omega_m \subset \Omega'$.

$$\begin{aligned} \|g - \sum_{k \in \Omega'} a_k e_k\| &\leq \|g - g_m\| + \|g_m - \sum_{j \in \Omega'} a_j e_j\| \\ &\leq \frac{1}{m} + \frac{1}{m}. \end{aligned}$$

$\square$

Theorem 4.12 (Bassel's inequility)**.**
Suppose $\{e_k\}_{k \in \Gamma}$ is an orthonormal family in an inner product space V and $f \in V$. Then we have
$$\sum_{k \in \Gamma} |\langle f, e_k \rangle|^2 \leq \|f\|^2,$$

and thus $\sum_{k \in \Gamma} \langle f, e_k \rangle e_k$ converges.

Proof. Decomposing f as follows:

$$f \;=\; \sum_{j \in \Omega} \langle f, e_j \rangle e_j + (f - \sum_{j \in \Omega} \langle f, e_j \rangle e_j),$$

where the first part in the right side is orthogonal to the second part, as you should verify. Applying the Pythagorean Theorem to the equation above and we obtain

$$
\begin{aligned}
\|f\|^2 \;&=\; \|\sum_{j \in \Omega} \langle f, e_j \rangle e_j\|^2 + \|(f - \sum_{j \in \Omega} \langle f, e_j \rangle e_j)\|^2 \\
&\geq\; \sum_{k \in \Omega} |\langle f, e_k \rangle|^2.
\end{aligned}
$$

Take supremum, and we have

$$\|f\|^2 \geq \sum_{k \in \Gamma} |\langle f, e_k \rangle|^2.$$

$\square$

Recall that we have already define the span of a family by

$$\mathrm{span}\{e_k\}_{k \in \Gamma} := \{\sum_{k \in \Omega} a_k e_k : \{a_k\}_{k \in \Omega} \text{ is a finite subfamily of } \mathbb{F}\}.$$

And the closure of $\mathrm{span}\{e_k\}_{k \in \Gamma}$ is

$$\overline{\mathrm{span}\{e_k\}_{k \in \Gamma}} = \{g \in V : \forall \varepsilon > 0, B(g, r) \cap \mathrm{span}\{e_k\}_{k \in \Gamma} \neq \emptyset\}.$$

Theorem 4.13 (Closure of the span of an orthonormal family).
Suppose $\{e_k\}_{k \in \Gamma}$ is an orthonormal family in a Hilbert space V. Then

$$\overline{\mathrm{span}\{e_k\}_{k \in \Gamma}} = \left\{ \sum_{k \in \Gamma} a_k e_k : \{a_k\}_{k \in \Gamma} \text{ is a family in } \mathbb{F} \text{ and } \sum_{k \in \Gamma} |a_k|^2 < \infty. \right\}$$

Moreover, for all $f \in \overline{\mathrm{span}\{e_k\}_{k \in \Gamma}}$

$$f = \sum_{k \in \Gamma} \langle f, e_k \rangle e_k.$$

Proof. Note that $\overline{\mathrm{span}\{e_k\}_{k \in \Gamma}}$ is a vector space, as you should verify. The set in the right is well-defined since $\sum_{k \in \Gamma} |a_k|^2 < \infty$. We start our proof with $\left\{ \sum_{k \in \Gamma} a_k e_k : \sum_{k \in \Gamma} |a_k|^2 < \infty. \right\} \subset \overline{\mathrm{span}\{e_k\}_{k \in \Gamma}}$.

First note that for all g in the left set, assume $g = \sum_{k \in \Gamma} a_k e_k$, for all $\varepsilon > 0$, there exists $\Omega \subset \Gamma$, such that

$$\sum_{\Gamma \backslash \Omega} |a_k|^2 < \varepsilon^2,$$

which implies

$$\| \sum_{k \in \Gamma} a_k e_k - \sum_{k \in \Omega} a_k e_k \| = \sum_{\Gamma \backslash \Omega} |a_k|^2 < \varepsilon.$$

In other words,

$$B(g, \varepsilon) \cap \overline{\operatorname{span}\{e_k\}_{k \in \Gamma}} \neq \emptyset,$$

and thus $g \in \overline{\operatorname{span}\{e_k\}_{k \in \Gamma}}$.

To prove the other direction, suppose $f \in \overline{\operatorname{span}\{e_k\}_{k \in \Gamma}}$, and

$$g = \sum_{k \in \Gamma} \langle f, e_k \rangle e_k.$$

The Bassel's inequility now implies g is in the set in the right part, next, by the inclusion we have just proved, $g \in \overline{\operatorname{span}\{e_k\}_{k \in \Gamma}}$ and thus $f - g \in \overline{\operatorname{span}\{e_k\}_{k \in \Gamma}}$. Also,

$$
\begin{aligned}
\langle g, e_j \rangle &= \langle \sum_{k \in \Gamma} \langle f, e_k \rangle e_k, e_j \rangle \\
&= \sum_{k \in \Gamma} \langle f, e_k \rangle \langle e_k, e_j \rangle \\
&= \langle f, e_j \rangle.
\end{aligned}
$$

Hence

$$\langle f - g, e_k \rangle = 0, \ \forall k \in \Gamma.$$

This implies that

$$g - f \in \left(\operatorname{span}\{e_j\}_{j \in \Gamma} \right)^{\perp} = \left(\overline{\operatorname{span}\{e_j\}_{j \in \Gamma}} \right)^{\perp}$$

Since

$$\left(\operatorname{span}\{e_j\}_{j \in \Gamma} \right)^{\perp} \cap \left(\overline{\operatorname{span}\{e_j\}_{j \in \Gamma}} \right)^{\perp} = \{0\},$$

we have $g = f$. □

Definition 4.8 (Orthonormal Basis).
An orthonormal family $\{e_k\}_{k \in \Gamma}$ in a Hilbert space is called an orthonormal family of V if

$$\overline{\operatorname{span}\{e_k\}_{k \in \Gamma}} = V$$

Theorem 4.14 (Parseval's identity)**.** Suppose $\{e_k\}_{k\in\Gamma}$ is an orthonormal basis in a Hilbert space V. For all $f, g \in V$, we have

- $f = \sum_{k\in\Gamma} \langle f, e_k\rangle e_k$

- $\langle f, g\rangle$

- $\langle f, f\rangle = \sum_{k\in\Gamma} |\langle f, e_k\rangle|^2$

Example 4.3. For $k \in \mathbf{Z}$, suppose $e_k : (-\pi, \pi] \to \mathbf{R}$ is defined by

$$
e_k(t) = \begin{cases} \frac{1}{\sqrt{\pi}}\sin(kt) & \text{if } k > 0, \\ \frac{1}{\sqrt{2\pi}} & \text{if } k = 0, \\ \frac{1}{\sqrt{\pi}}\cos(kt) & \text{if } k < 0 \end{cases}
$$

Then $\{e_k\}_{k\in Z}$ is an orthonormal basis of $L^2((-\pi, \pi])$.

Chapter 5

Measure Theory

A measure is a countably additive function from a σ-algebra to $[0, \infty]$. In this chapter, we consider countably additive functions from a σ-algebra to either $\mathbf{R}$ or $\mathbf{C}$. The first section of this chapter shows that these functions, called real measures or complex measures, form an a Banach space with total variation norm.

The second section of this chapter focuses on decomposition theorems that help us understand real and complex measures. Radon-Nikodym theorem is derived and these results will lead to a proof that the dual space of $L^p(\mu)$ can be identified with $L^{p'}(\mu)$.

5.1   Complex Measures

Suppose $(\mathbf{X}, \mathcal{S})$ is a measurable space.

Definition 5.1.

- A function $\nu : S \to F$ is called countably additive if

$$\nu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \nu(E_n).$$

- A real measure on $(\mathbf{X}, \mathcal{S})$ is a countably addtive function: $\nu : \mathcal{S} \to \mathbb{R}$.

- A complex measure on $(\mathbf{X}, \mathcal{S})$ is a countably addtive function: $\nu : \mathcal{S} \to \mathbb{C}$.

Remark 3.

1. We use the terminology (positive) measure to reprensent the function that maps $\mathcal{S} \to [0, \infty]$.

2. A finite measure is a real measure.

3. $\mathcal{M}_{\mathbb{C}}(\mathcal{S}) := \{\nu : \nu \text{ is a complex measure on } (\mathbf{X}, \mathcal{S})\}$

Definition 5.2 (addition and scalar multiplication)**.**
for all $E \in \mathcal{S}$ and $a \in \mathbb{F}$, define

- $\nu + \mu : E \mapsto \nu(E) + \mu(E)$.

- $a\nu : E \mapsto a\nu(E)$.

Theorem 5.1. $\mathcal{M}_{\mathbb{F}}(\mathcal{S})$ is a vector space over $\mathbb{F}$.

Theorem 5.2. Suppose $\nu \in \mathcal{M}_{\mathbb{C}}(\mathcal{S})$, then

1. $\nu(\emptyset) = 0$.

2. $\sum_{n=1}^{\infty} |\nu(E_k)| < \infty$ for all disjoint sequence $E_1, \cdots$ of $\mathcal{S}$.

Proof.

$$\nu(\emptyset) = \nu(\bigcup_{n=1}^{\infty} \emptyset) = \sum_{n=1}^{\infty} \nu(\emptyset).$$

This implies $\nu(\emptyset) = 0$.

On the other hand, suppose F=R, making the following decomposition, we obtain

□

Theorem 5.3.

Suppose μ is a measure on $(\mathbf{X}, \mathcal{S})$ and $h \in \mathcal{L}(\mu)$. Define $\nu : \mathcal{S} \to \mathbb{F}$ by

$$\nu(E) = \int_E h \mathrm{d}\mu$$

Then $\nu \in \mathcal{M}_\mathbb{C}(\mathcal{S})$.

Remark 4.

The ν defined above is often denoted by $h\mathrm{d}\mu$.

Theorem 5.4 (Properties of complex measures)**.**

1. $\nu(E \backslash D) = \nu(E) - \nu(D)$ *for all* $D \subset E$.

2. $\nu(D \cup E) = \nu(D) + \nu(E) - \nu(D \cap E)$.

3. for increasing sequence E_n,

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

4. for decreasing sequence E_n,

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

Definition 5.3.

Total variation measure of ν is the function $|\nu| : \mathcal{S} \to [0, \infty]$ defined by

$$|\nu| = \sup\{\sum_{k=1}^{n} |\nu(E_k)| : E_i \cap E_j = \emptyset, \ \bigcup_{k=1}^{n} E_k \subset E\}.$$

Theorem 5.5 (Properties of total variation measure)**.**

- $|\nu(E)| \leq |\nu|(E)$ for all $E \in \mathcal{S}$.

- $|\nu|(E) = \nu(E)$ if ν is a finite measure.

- $|\nu|(E) = 0$ iff $\forall A \in \mathcal{S},\ \nu(A) = 0$.

- Equivalent definition:

$$|\nu|(E) = \sup\{|\nu(A)| + |\nu(B)| : A \cap B = \emptyset \ and \ A \cup B \subset E\}.$$

Theorem 5.6.
Total variation measure is indeed a complex measure on $(\mathbf{X}, \mathcal{S})$. To be specific,

$$\nu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S}) \text{ implies } |\nu| \text{ is a measure.}$$

Definition 5.4 (Total variation norm)**.**

$$\|\nu\| := |\nu|(\mathbf{X})$$

Theorem 5.7. $(\mathcal{M}_{\mathbb{F}}(\mathcal{S}), \|\|)$ is a Banach space.

We perform the following procedures to prove this theorem.

Proof.

1. $\forall \nu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S}), \|\nu\| = |\nu|(\mathbf{X}) < \infty$.

2. $\|\|$ is a norm on $\mathcal{M}_{\mathbb{F}}(\mathcal{S})$.

3. $(\mathcal{M}_{\mathbb{F}}(\mathcal{S}), \|\|)$ is complete.

S1. First suppose $\mathbb{F} = \mathbb{R}$ and $|\nu|(\mathbf{X}) = \infty$.
Provided that E_n is well-defined such that $|\nu|(E_n) = \infty$ and $|\nu(E_n)| \geq n$. By definition, there exists $A \subset E_n$ such that

$$|\nu(A)| \geq n + 1 + |\nu(E_n)|$$

$$|\nu(E_n \backslash A)| = |\nu(E_n) - \nu(A)| \geq n + 1$$

$$|\nu|(A) + |\nu|(E_n \backslash A) = \infty.$$

Since $|\nu|$ is a measure, we conclude at least one of $\nu|(A)$ and $|\nu|(E_n \backslash A)$ equals ∞. Thus we take

$$E_{n+1} = A \ if \ |\nu|(A) = \infty, \ otherwise \ E_{n+1} = E_n \backslash A.$$

The procedure above induces a decreasing squence E_n with

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

Where the left side is well-defined since ν is a complex measure, while the right side does not exists. This contradicts to our assumption of $|\nu|(X) = \infty$.

S2. The proof of 2 is lefted to the reader.

S3. Suppose ν_n is Cauchy sequence in $\mathcal{M}_{\mathbb{F}}\mathcal{S}$. For each $E \in \mathcal{S}$, we have

$$\begin{aligned}
|\nu_i(E) - \nu_j(E)| &= |(\nu_i - \nu_j)(E)| \\
&\leq |\nu_i - \nu_j|(E) \\
&\leq \|\nu_i - \nu_j\|.
\end{aligned}$$

Which implies $\nu_i(E)$ is a Cauchy sequence in $\mathbb{F}$. Thus we define nu as follows:

$$\nu : \mathcal{S} \to \mathbb{F}, \ E \mapsto \lim_{j\to\infty} \nu_j(E).$$

We make the following claims:

1. $\lim_{n\to\infty} \|\nu_n - \nu\| = 0$.

2. ν is complex measure.

For one thing, suppose $E_1, \cdots, E_n$ are disjoints subsets of $\mathbf{X}$. $\forall \varepsilon > 0, \exists m > 0$, for all i, j $\geq$ m. We have

$$\begin{aligned}
\sum_{l=1}^{n} |\nu - \nu_k|(E_l) &= \lim_{j\to\infty} \sum_{l=1}^{n} |\nu_j - \nu_k|(E_l) \\
&< \lim_{j\to\infty} \|\nu_j - \nu_k\| \\
&\leq \varepsilon.
\end{aligned}$$

The definition of $|\nu - \nu_k|(\mathbf{X})$ implies

$$|\nu - \nu_k|(\mathbf{X}) = \|\nu - \nu_k\| < \varepsilon.$$

On the other hand, also suppose that E_n are disjoint subsets of X, considering that

$$
\begin{aligned}
|\nu(\bigcup_{k=1}^{\infty} E_k) - \sum_{k=1}^{n-1} \nu(E_k)| &= |\lim_{j \to \infty} \nu_j(\bigcup_{k=1}^{\infty} E_k) - \lim_{j \to \infty}(\sum_{k=1}^{n-1} \nu_j(E_k))| \\
&= \lim_{j \to \infty} \sum_{k=n}^{\infty} |\nu_j(E_k)| \\
&= \lim_{j \to \infty} \sum_{k=n}^{\infty} |(\nu_j - \nu_m)(E_k) + \nu_m(E_k)| \\
&\leq \lim_{j \to \infty} \sum_{k=n}^{\infty} |(\nu_j - \nu_m)|(E_k) + \sum_{k=n}^{\infty} |\nu_m(E_k)| \\
&\leq \lim_{j \to \infty} \|\nu_j - \nu_m\| + \sum_{k=n}^{\infty} |\nu_m(E_k)|.
\end{aligned}
$$

Then, we choose .... This completes the prove of Theorem 5.7.

$\square$

5.2   Decomposition Theorem

Theorem 5.8 (Hahn Decomposition Theorem).
Suppose $\nu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S})$. Then there exsits $A, B \in \mathcal{S}$ such that

- $A \cup B = \mathbf{X}$ and $A \cap B = \emptyset$.

- $\nu(E) \geq 0$ for all $E \subset A$.

- $\nu(E) \leq 0$ for all $E \subset B$.

Proof. Suppose $a = \sup\{\nu(E) : E \in \mathcal{S}\}$. By definition, we have the following inequality

$$a \leq \|\nu\| < \infty.$$

The sequence E_n is chosen to such that

$$\nu(E_j) \geq a - \frac{1}{2^j}.$$

We claim that

$$A = \bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty} E_j, \ B = X \backslash A.$$

To illustrate why A meets our needs, we will show by induction on n if $n \geq k$ holds

$$\nu(\bigcup_{j=k}^{n} \nu(E_k)) \geq a - \sum_{j=k}^{n} \frac{1}{2^j}.$$

To get started with the induction, note that if $n = k$, the equation clearly holds. Suppose $n \geq k$ hold, then

$$
\begin{aligned}
\nu(\bigcup_{j=k}^{n+1} E_j) &= \nu(\bigcup_{j=k}^{n} E_k) + \nu(E_{n+1}) - \nu(\bigcup_{j=k}^{n} E_k \backslash E_{n+1}) \\
&\geq a - \sum_{j=k}^{n} \frac{1}{2^j} + a - \frac{1}{2^{n+1}} - a \\
&= a - \sum_{j=k}^{n+1} \frac{1}{2^j}.
\end{aligned}
$$

The continuity of measure admits

$$\nu(\bigcup_{j=k}^{\infty} A_j) \geq a - \frac{1}{2^{k-1}}$$

and

$$\nu(\bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} E_j) = a,$$

and hence for all $E \in \mathcal{S}$

$$\nu(E \cap A) = \nu(A) - \nu(A \cap E^c) \geq 0.$$

Since $(E \cap B) \cap A = \emptyset$, we have

$$\nu(E \cap B) = \nu(A \cup E \cap B) - \nu(A) \leq 0.$$

This completes the proof of Theorem 5.8. $\qquad\qquad\qquad\qquad\qquad\qquad \square$

Definition 5.5. Suppose $\nu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S})$. ν is called lives on a set A if for all $E \in \mathcal{S}$, we have

$$\nu(E) = \nu(E \cap A)$$

Remark 5. ν lives on A implies for all $E \in A^c$, $\nu(E) = 0$.

Definition 5.6 (Singular Measures).
Suppose $\nu, \mu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S})$. Then ν and μ are called singular with respect to each other, denoted

$\nu \perp \mu$ if ν and μ live on different set. More precisely, there exists $A, B \in \mathcal{S}$ such that

- $A \cup B = \mathbf{X}$ and $A \cap B = \emptyset$.

- $\nu(E) = \nu(E \cap A)$ and $\mu(E) = \mu(E \cap B)$

Theorem 5.9 (Jordan Decomposition Theorem).
Suppose ν is a real measure on $(\mathbf{X}, \mathcal{S})$. Then there exists unique finite measures ν^+ and ν^- on $(\mathbf{X}, \mathcal{S})$ such that

$$\nu = \nu^+ - \nu^- \text{ and } \nu^+ \perp \nu^-.$$

Moreover, $|\nu| = \nu^+ + \nu^-$.

Proof. Let $\mathbf{X} = A \cup B$ be the Hahn Decomposition of ν. Define

$$\nu^+ : \mathcal{S} \to [0, \infty), \ E \mapsto \nu(E \cap A),$$

and

$$\nu^- : \mathcal{S} \to [0, \infty), \ E \mapsto -\nu(E \cap B).$$

The contably addivity of ν implies ν^+ and ν^- are finite measures on $(\mathbf{X}, \mathcal{S})$ with

$$\nu = \nu^+ - \nu^-, \nu^+ \perp \nu^-. (\nu^+ \text{ lives on } A \text{ while } \nu^- \text{ lives on } B)$$

Let us turn to the proof of $\nu = \nu^+ + \nu^-$. First note that $\nu^+ + \nu^-$ is indeed a measure. On the one hand,

$$
\begin{aligned}
|\nu(E \cap A)| + |\nu(E \cap B)| &= (\nu^+ + \nu^-)(E \cap A) + (\nu^+ + \nu^-)(E \cap B) \\
&= (\nu^+ + \nu^-)(E).
\end{aligned}
$$

Which implies $|\nu|(E) \geq (\nu^+ + \nu^-)(E)$.
On the other hand, Suppose E_1, E_2 are disjoint subsets in E.

$$
\begin{aligned}
|\nu(E_1)| + |\nu(E_2)| &= |(\nu^+ - \nu^-)(E_1)| + |(\nu^+ - \nu^-)(E_2)| \\
&\leq (\nu^+ + \nu^-)(E_1) + (\nu^+ + \nu^-)(E_2) \\
&\leq (\nu^+ + \nu^-)(E).
\end{aligned}
$$

Take supreme of both sides, we obtain $|\nu(E)| \leq (\nu^+ + \nu^-)(E)$. As for the uniqueness, note also that

$$\nu^+ = \frac{|\nu| + \nu}{2}, \ \nu^- = \frac{|\nu| - \nu}{2}.$$

Since $|\nu|$ and ν are well-defined, ν^+ can be uniquely identified by ν. This completes the proof of Theorem 5.9. $\qquad\square$

Remark 6. $\nu^+(\nu^-)$ is called the positive(negative) part of ν.

Definition 5.7 (Integration with respect to real measures). Suppose $\nu \in \mathcal{M}_R(\mathcal{S})$, then the integral of f is defined by

$$\int_{\mathbf{X}} f \mathrm{d}\nu := \int_{\mathbf{X}} f \mathrm{d}\nu^+ - \int_{\mathbf{X}} f \mathrm{d}\nu^-$$

Definition 5.8 (Absolutely continuous; $\ll$). Suppose $\nu \mathcal{M}_{\mathbb{F}}(\mathcal{S})$ and μ is a measure. Then ν is called absolutely continuous with respect to μ denoted by $\nu \ll \mu$ if

$$\mu(E) = 0 \implies \nu(E) = 0.$$

Note that $\{\nu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S}) : \nu \ll \mu\}$ is a subspace.

Theorem 5.10. Suppose $\nu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S})$, then

$$\nu \ll \mu \text{ and } \nu \perp \mu \text{ implies } \nu = 0.$$

Proof. Since nu and μ are singular, there exists disjoints subsets such that $X = A \cup B$ and for all $E \in \mathcal{S}$

$$\begin{aligned}
\nu(E) &= \nu(E \cap A) \\
\mu(E) &= \mu(E \cap B) \\
\mu(E \cap A) &= \mu(E \cap A \cap B) = 0.
\end{aligned}$$

Also note that ν is absolutely continuous with respect to μ, and thus we have

$$\nu(E) = \nu(E \cap A) = 0.$$

Or $\|\nu\| = 0$. $\qquad\square$

Theorem 5.11 (Lebesgue Decomposition Theorem).
Suppose μ is a measure and $\nu \in \mathcal{M}_{\mathbb{F}}(\mathcal{S})$. Then there exsits unique $\nu_a, \nu_b \in \mathcal{M}_{\mathbb{F}}(\mathcal{S})$ such that $\nu = \nu_a + \nu_s$ and

$$\nu_a \ll \mu \text{ and } \nu_s \perp \mu.$$

Proof. Let

$$b = \sup\{|\nu|(B) : B \in \mathcal{S}, \ \mu(B) = 0\},$$

and B_k be such that

$$|\nu|(B_k) \geq b - \frac{1}{k} \text{ and } \mu(B_k) = 0,$$

Taking B as

$$B = \bigcup_{k=1}^{\infty} B_k.$$

then $\mu(B) = 0$ and $|\nu(B)| = b$. Let $A = X \backslash B$ and

$$\nu_a \ : \ \mathcal{S} \to \mathbb{F}, \ E \mapsto \nu(E \cap A)$$
$$\nu_s \ : \ \mathcal{S} \to \mathbb{F}, \ E \mapsto \nu(E \cap B).$$

Clear $\nu = \nu_a + \nu_s$. Note also that

$$\mu(E) = \mu(E \cap A) + \mu(E \cap B) = \mu(E \cap A),$$

which implies μ and ν_s live on different sets and thus mutually singular. At last, to prove $\nu_a \ll \mu$, first suppose $\mu(E) = 0$.

$$\nu_a(E) = \nu(E \cap A) = \nu(E \backslash B)$$

Since $\mu(E) = 0$, we have $\mu(B \cup E) = 0$ and thus

$$b \geq |\nu|(B \cup E) = |\nu|(B) + |\nu|(E \backslash B) = b + |\nu|(E \backslash B).$$

This implies $|\nu|(E \backslash B) = 0$, and hence $\nu(E \backslash B) = 0$. As for the uniqueness, suppose ν_1, ν_2 be another Lebesgue-Decompositon of ν with respect to μ. Then $\nu_a + \nu_s = \nu_1 + \nu_2$ implies

$$\nu_a - \nu_1 = \nu_s - \nu_2.$$

The left part is absolutely continuous to μ while the right part is singuar to μ. Hence we conclude $\nu_1 = \nu_a$ and $\nu_2 = \nu_S$. $\square$

Theorem 5.12 (Radon-Nikodym Theorem)**.**
Suppose μ is a σ-finite measure on a measurable space $(\mathbf{X}, \mathcal{S})$. Suppose ν is a complex measure $(\mathbf{X}, \mathcal{S})$ such $\nu \ll \mu$. Then there exists $h \in \mathcal{L}^1(\mu)$ such that $d\nu = hd\mu$.

Proof.
Step 1. Suppose ν and μ are finite measures on $(\mathbf{X}, \mathcal{S})$.

For all $f \in L^2(\nu + \mu)$, define φ as

$$\varphi : L^2(\nu + \mu) \to \mathbb{R}, \ f \mapsto \int_{\mathbf{X}} f \mathrm{d}\nu.$$

To see why our definition makes sense, note that

$$\int_{\mathbf{X}} |f| \mathrm{d}\nu \leq \int_{\mathbf{X}} |f| \mathrm{d}(\nu + \mu) \leq (\nu(\mathbf{X}) + \mu(\mathbf{X}))^{\frac{1}{2}} \|f\|_{L^2(\nu+\mu)}.$$

This implies $f \in L^2(\nu)$ and $\|\varphi\| \leq (\nu(\mathbf{X}) + \mu(\mathbf{X}))^{\frac{1}{2}}$. Since ν is a bounded linear functional on $f \in L^2(\nu + \mu)$. The Riesz Representation Theorem now implies that there exists unique $g \in L^2(\nu + \mu)$ such that

$$\int_{\mathbf{X}} f \mathrm{d}\nu = \int_{\mathbf{X}} fg \mathrm{d}(\nu + \mu).$$

Equvalently,

$$\int_{\mathbf{X}} f(1-g) \mathrm{d}\nu = \int_{\mathbf{X}} fg \mathrm{d}\mu.$$

Suppose

$$E_1 = \{x \in \mathbf{X} : g(x) \geq 1\}$$
$$E_0 = \{x \in \mathbf{X} : g(x) < 0\}.$$

And

$$f = \chi_{E_1} \implies \mu(E_1) = 0 \implies \nu(E_1) = 0$$
$$f = \chi_{E_0} \implies \mu(E_0) = 0 \implies \nu(E_0) = 0.$$

Thus we can define g as a function with range [0,1), and we define $h : \mathbf{X} \to [0, \infty)$ by

$$h(x) = \frac{g(x)}{1 - g(x)}.$$

Then suppose $E \in \mathcal{S}$, and let f_k be

$$f_k(x) = \frac{\chi_E(x)}{1 - g(x)} I_{\frac{\chi_E(x)}{1-g(x)} < k}(x).$$

Easy to verify that f_k is an increasing function with

$$\int_{\mathbf{X}} f_k(x)^2 \mathrm{d}(\nu + \mu) \leq k^2 (\nu(\mathbf{X}) + \mu(\mathbf{X})).$$

Thus $f_k \in L^2(\nu + \mu)$ and hence

$$\int_{\mathbf{X}} f_k(1 - g)\mathrm{d}\nu = \int_{\mathbf{X}} f_k g \mathrm{d}\mu$$

The Monotone Convergence Theorem now implies

$$\int_{\mathbf{X}} \chi_E \mathrm{d}\nu = \int_{\mathbf{X}} \chi_E h \mathrm{d}\mu.$$

Step 2. Suppose ν is a measure while μ is a σ-finite measure.

Now relex the assumption on μ to the hypothesis that μ is a σ-finite measure. Thus there exsits an increasing sequence sets $\mathbf{X}_1 \subset \mathbf{X}_2 \subset \cdots$ such that $\bigcup_{j=1}^{\infty} \mathbf{X}_j = \mathbf{X}$ and $\mu(\mathbf{X}_j) < \infty$. Let ν_k, μ_k be the restriction of ν, μ on $\mathbf{X}_k$. Since ν_k and μ_k are finite measures, step1 implies there exists $h_k \in L^1(\mu)$ such that $\mathrm{d}\nu_k = h_k \mathrm{d}\mu_k$. Next, for $j < k$ and for all $E \in \mathcal{S}$, we have

$$\begin{aligned}
\int_{E \cap \mathbf{X}_j} h_j \mathrm{d}\mu_j &= \int_{E \cap \mathbf{X}_j} h_j \mathrm{d}\mu \\
&= \nu_j(E \cap \mathbf{X}_j) \\
&= \nu_k(E \cap \mathbf{X}_j) \\
&= \int_{E \cap \mathbf{X}_j} h_k \mathrm{d}\mu.
\end{aligned}$$

This implies

$$\mu(\{x \in \mathbf{X}_k : h_j(x) \neq h_k(x)\}) = 0.$$

We conclude there exists $h : \mathbf{X} \to [0, \infty)$ such that

$$\mu(\{x \in \mathbf{X}_k : h(x) \neq h_k(x)\}) = 0$$

for all $k > 0$. Suppose otherwise ...

$$\nu(E \cap \mathbf{X}_k) = \int_{E \cap \mathbf{X}_k} h \mathrm{d}\mu$$

Let $k \to \infty$, and complete the proof.

Step 3. Suppose ν is a real measure while μ is a σ-finite measure.

Note that $|\nu| + \nu$ and $|\nu| - \nu$ are finite measures. By the case proved last page, there exists $h_+, h_- \in L^1(\mu)$ such that

$$\mathrm{d}(|\nu| + \nu) = h_+ \mathrm{d}\mu, \ \mathrm{d}(|\nu| - \nu) = h_- \mathrm{d}\mu.$$

and
$$d\nu = \frac{h_+ + h_-}{2} d\mu.$$

$\square$

Remark 7. The construction of h shows that it is undistinguishable on a set of zero measure with respect μ,

Theorem 5.13. Suppose ν is a complex measure and μ, λ are σ-finite measures on a measurable space $(\mathbf{X}, \mathcal{S})$,

1. If $\nu \ll \mu$, then $g \in L^1(\mathbf{X}, \nu)$ implies $g\frac{d\nu}{d\mu} \in L^1(\mathbf{X}, \mu)$ and $\int_X g d\nu = \int_X g\frac{d\nu}{d\mu} d\mu$.

2. Chain rule: If $\nu \ll \mu$, $\mu \ll \lambda$, then $\nu \ll \lambda$ and $\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$.

Proof. Suppose $E \in \mathcal{S}$, and $f = \chi_E$.

$$\int_{\mathbf{X}} g d\nu = \nu(E) = \int_E \frac{d\nu}{d\mu} d\mu = \int_{\mathbf{X}} g\frac{d\nu}{d\mu} d\mu.$$

Other cases are obtained by fact that the simple function is dense $L^1(\mathbf{X}, \nu)$.

$\square$

Example 5.1. Suppose P is a probabilty measure on $(\Omega, \mathcal{F})$ and $B \in \mathcal{F}$ with $P(B) > 0$, then
$$P_B : E \mapsto \frac{P(E \cap B)}{P(B)}$$
is a probability measure on $(\Omega, \mathcal{F})$, and $P_B \ll P$.

$$\frac{dP_B}{dP} = \frac{I_B}{P(B)}.$$

Thus if $X \in L^1(\Omega, \mathcal{F}, P_B)$, we have

$$E(X|B) := E^{P_B}(X) = \frac{E(XI_B)}{P(B)}.$$

5.3 Product Measures

$$x \in \bigcup_{n=1}^{\infty} A_n \iff \exists n, x \in A_n$$

$$\bigcup_{n=1}^{\infty} A_n := \{x : \exists n, x \in A_n\}$$

$$f : \mathbb{R} \to \mathbb{R}, \ x \mapsto x^2$$

Chapter 6

Integration

Chapter 7

Probability Theory

Probability theory has become increasingly important in multiple parts of science. Getting deeply into probability theory requires a full book, not just a chapter. For readers who intend to pursue further studies in probability theory, this chapter gives you a good head start. For readers not intending to delve further into probability theory, this chapter gives you a taste of the subject.

Modern probability theory makes major use of measure theory. As we will see, a probability measure is simply a measure such that the measure of the whole space equals 1 . Thus a thorough understanding of the chapters of this book dealing with measure theory and integration provides a solid foundation for probability theory. However, probability theory is not simply the special case of measure theory where the whole space has measure 1 . The questions that probability theory investigates differ from the questions natural to measure theory. For example, the probability notions of independent sets and independent random variables, which are introduced in this chapter, do not arise in measure theory.

Even when concepts in probability theory have the same meaning as well-known concepts in measure theory, the terminology and notation can be quite different. Thus one goal of this chapter is to introduce the vocabulary of probability theory. This difference in vocabulary between probability theory and measure theory occurred because the two subjects had different historical developments, only coming together in the first half of the twentieth century.

7.1   Probability Space

Definition 7.1. Suppose $(\Omega, \mathcal{F})$ is a measurable space, we say P is a probabilty measure define on $(\Omega, \mathcal{F})$ if

1. P is a measure on $(\Omega, \mathcal{F})$,

2. range$P \in [0, 1]$,

3. $P(\Omega) = 1$.

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Definition 7.2 (Distribution Function).
Suppose $H : \mathbb{R} \to [0, 1]$ is a function. Then H is called a distribution function if the following conditions are all satisfied:

1. H is an increasing function;

2. $\lim_{t \to -\infty} H(t) = 0$;

3. $\lim_{t \to -\infty} H(t) = 1$;

4. $\lim_{t \to s^+} H(t) = H(s)$. In other words, H is right continuous.

Theorem 7.1. If H is a distribution function, then there exists a probability space $(\Omega, \mathcal{F}, P)$, and a random variable $X \in L^1(P)$ such that $H = F_X$.

Proof. We will show this theorem by construction. Let $\Omega = (0, 1)$, $\mathcal{F} = B((0, 1))$ and P is the Lesbesgue measure on $(\Omega, \mathcal{F})$. Definea random varable X by

$$X(\omega) = \sup\{t \in R : H(t) < \omega\}.$$

$\square$

7.2   Random Variables and Distribution

In this course, a random variable is a function from Ω to $\mathbb{R}$, often denoted by X, Y, Z. It does no harm to assume that $X \in L^1(\mathcal{F}, P)$.

Definition 7.3 (Probabilty Distribution Measure). .   Suppose $X \in L^1(\mathcal{F}, P)$, then the probabilty distribution measure of X, denoted by μ_X, is a set function defined by

$$\mu_X := P \circ X^{-1} : B(\mathbb{R}) \to [0, 1], \ B \mapsto P \circ X^{-1}(B)$$

The function μ_X defined above is indeed a probability measure on $(\mathbb{R}, B(\mathbb{R}))$, as you should verify.

Remark 8. 123

Definition 7.4. (CDF)
Suppose $X \in L^1(\mathcal{F}, P)$, then the cdf of X is the function F_X defined by

$$F_X : \mathbb{R} \to [0, 1], \ x \mapsto \mu_X((-\infty, x]).$$

Proposition 7.2.1. Suppose λ is the Lebesgue measure on $(\mathbb{R}, B(\mathbb{R}))$ and μ_X is the probability distribution defined on $(\mathbb{R}, B(\mathbb{R}))$. If $\mu_X \ll \lambda$, then there exists unique $f_X \in L^1(\mathbb{R}, B(\mathbb{R}), \lambda)$, such that for

$$\mu_X(A) = \int_A f_X \mathrm{d}\lambda.$$

Proof. A direct consequence of Radon-Nikodym Theorem. $\qquad\qquad\square$

Remark 9. f_X is often called the density function of X.

Theorem 7.2 (Expectation)**.** Suppose g is a Borel-measurable function, then

$$E(g(X)) = \int_\Omega g \circ X \mathrm{d}P = \int_\mathbb{R} g \circ I \mathrm{d}\mu_X = \int_\mathbb{R} g(x)\mathrm{d}\mu_X(x).$$

Where $I : \mathbb{R} \to \mathbb{R}, \ x \mapsto x$.

Proof. First suppose $g \circ X$ is a simple function that in $L^1(\Omega, \mathcal{F}, P)$, that is

$$g \circ X = \sum_{k=1}^n c_k I_{E_k}.$$

$$
\begin{aligned}
gXX^{-1} &= \sum_{k=1}^n c_k I_{E_k} X^{-1}. \\
&= \sum_{k=1}^n c_k I_{X(E_k)}
\end{aligned}
$$

On the one hand,

$$\int_\Omega g \circ X \mathrm{d}P = \sum_{k=1}^n c_k P(E_k).$$

On the other hand,

$$
\begin{aligned}
\int_{\mathbb{R}} g \circ Id\mu_X \;&=\; \int_{\mathbb{R}} \sum_{k=1}^{n} c_k I_{X(E_k)} d\mu_X \\
&=\; \sum_{k=1}^{n} c_k \mu_X X E_k \\
&=\; \sum_{k=1}^{n} c_k P X^{-1} X E_k \\
&=\; \sum_{k=1}^{n} c_k P(E_k).
\end{aligned}
$$

The other case is obtained by the density argument. $\square$

Proposition 7.2.2. If $\mu_X \ll \lambda$, then

$$
E(g(X)) = \int_{\Omega} g \circ X dP = \int_{\mathbb{R}} g(x) f(x) d\lambda(x)
$$

Definition 7.5 (Random Vector)**.**
Suppose $X_1, X_2 \in L^1(\Omega, \mathcal{F}, P)$. Then the joint probabilty measure of X_1 and X_2 is a measure on $(\mathbb{R}^2, B(\mathbb{R}^2))$, defined by

$$
\mu_{X_1, X_2} : B(\mathbb{R}^2) \to [0, 1], \; B_1 \times B_2 \mapsto P \circ (X_1, X_2)^{-1}(B_1 \times B_2)
$$

where

$$
(X_1, X_2)^{-1} : (B_1 \times B_2) \mapsto X^{-1}(B_1) \cap X^{-1}(B_2)
$$

Further, if $\mu_{X_1, X_2} \ll \lambda^2$, the Radon-Nikodym Theorem implies there exists unique $f_{X_1, X_2} \in L^2(\mathbb{R}^2, B(\mathbb{R}^2), \lambda^2)$, such that for all $E \in B(\mathbb{R}^2)$

$$
\mu_{X_1, X_2}(E) = \int_E f_{X_1, X_2} d\lambda^2.
$$

Theorem 7.3. Suppose g is a Borel-function on from $\mathbb{R}^2$ to $\mathbb{R}$. Then

$$
E(g(X, Y)) = \int_{\Omega} g \circ (X, Y) dP = \int_{\mathbb{R}^2} g(x, y) d\mu_{X, Y}(x, y)
$$

Further, if $\mu_{X,Y} \ll \lambda$, we have

$$
E(g(X, Y)) = \int_{\mathbb{R}^2} g(x, y) d\mu_{X,Y}(x, y) = \int_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) d\lambda^2(x, y).
$$

Since $f_{X,Y} \in L^2(\lambda^2)$, we have

$$f_Y : y \mapsto \int_{\mathbb{R}} f_{X,Y}(x,y) \mathrm{d}\lambda(x).$$

The Fubini's Theorem now implies $f_Y \in L^1(\lambda)$. You should verify that

$$f_Y = \frac{\mathrm{d}\mu_Y}{\mathrm{d}\lambda}.$$

f_y defined above if often the marginal density function of Y. Some facts that you should verify

- $\mu_Y(B) = \mu_{X,Y}(\mathbb{R}, B)$

-

Definition 7.6 (Conditional density function).
Suppose (X,Y) is a random vector and $\mu_{X,Y} \ll \lambda^2$. For $x \in \mathrm{supp}(f_X)$, we define the conditional density function of Y with respect to X by

$$f_{Y|X=x} : y \mapsto \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}\lambda(y)} = f(y|x)$$

Example 7.1 (Conditional Expectation I).
Suppose (X,Y) is a random vector and $\mu_{X,Y} \ll \lambda^2$. Then $E(Y|X = x)$ is defined by

$$E(Y|X = x) := \int_{\mathbb{R}} y f(y|x) \lambda(y)$$

Example 7.2 (Conditional Expectation II).
Suppose (X,Y) is a random vector and $\mathrm{range} X$ is countable. Then for all $x_k \in \mathrm{range} X$

$$E(Y|X = x_k) := E^{P_{x_k}}(Y) = \frac{E(Y I_{X=x_k})}{P(X = x_k)}$$

Example 7.3 (Conditional Expectation III).
Suppose (X,Y) is a random vector and suppose $\mathrm{range} X$ is countable.

Theorem 7.4.
Since we have not define the conditional expectation over a zero measure set. Thus we first suppose $E(Y|X = x) = h(x)$ is a Borel-measurale function, then we have

$$E(Y|X) = h(X).$$

First note that the left side of the equation above is well-defined(Conditional expectation of a sigma-algebra) while the right side is a compound function,

Proof. First assume (X, Y) is a random vector and $\mu_{X,Y} \ll \lambda^2$. On the one hand, $h(X)$ is $\mathcal{F}$-measurable. On the other hand, for all $A \in \sigma(X)$

$$
\begin{aligned}
\int_A h \circ X \mathrm{d}P &= \int_\Omega I_A h(X) \mathrm{d}P \\
&= \int_{\mathbb{R}} I_{X(A)}(x) h(x) \mathrm{d}\mu_X(x) \\
&= \int_{\mathbb{R}} I_{X(A)}(x) h(x) f_X(x) \mathrm{d}\lambda(x) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} I_{X(A)}(x) f_X(x) \frac{y f_{X,Y}(x,y)}{\int_{\mathbb{R}} f_{X,Y}(x,y) \mathrm{d}\lambda(y)} \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} I_{X(A)}(x) f_{X,Y}(x,y) y \mathrm{d}\lambda(x) \mathrm{d}\lambda(y) \\
&= \int_{\mathbb{R}^2} I_{X(A)}(x) y \mathrm{d}\mu_{X,Y}(x,y) \\
&= \int_\Omega I_A Y \mathrm{d}P
\end{aligned}
$$

This complete the proof of special case.

Case 2 : Assume range$X$ is countable. $\square$

7.3   Limit Theory

7.4   Convergence of Random Variables

7.5   Conditional Expectations

Suppose $(\Omega, \mathcal{F}, P)$ is a probability sapce, $\mathcal{G}$ is sub-sigma algebra and $X \in L^1(P)$.

Definition 7.7 (Conditional Expectation).
The random variables Y is called the conditional expectation of X if the following holds

1. $Y \in L^1(\Omega, \mathcal{G}, P_\mathcal{G})$.

2. $\forall A \in \mathcal{G}, \ \int_A X \mathrm{d}P = \int_A Y \mathrm{d}P$.

To illustrate that our definition makes sense, first note that the partial average $\nu(A) = \int_A X \mathrm{d}P$ forms a real measure on $(\Omega, \mathcal{G})$ and $\nu \ll P_\mathcal{G}$. Thus Radon-Nikodym Theorem ensures that there exists unique $Y \in L^1(\Omega, \mathcal{G}, P_\mathcal{G})$ such that

$$\int_A X \mathrm{d}P = \nu(A) = \int_A Y \mathrm{d}P_\mathcal{G} = \int_A Y \mathrm{d}P$$

Remark 10.

1. The conditional expectation forms a linear map from $L^1(\Omega, \mathcal{F}, P)$ to $L^1(\Omega, \mathcal{G}, P_\mathcal{G})$, denoted by $E_\mathcal{G}(E(\cdot|\mathcal{G}))$ and $\|E_\mathcal{G}\| \leq 1$.

2. Y is undistinguishable on a set with zero measure.

Theorem 7.5 (Properties of $E_\mathcal{G}$).

Theorem 7.6 (Conditional Expectation defined on $L^2(\Omega, \mathcal{F}, P)$).
We have already shown that $L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space. Note also that $L^2(\Omega, \mathcal{G}, P_\mathcal{G})$ is a complete subspace of $L^2(\Omega, \mathcal{F}, P)$, as you should verify. Then we have

$$L^2(\Omega, \mathcal{F}, P) = L^2(\Omega, \mathcal{G}, P_\mathcal{G}) \oplus (L^2(\Omega, \mathcal{G}, P_\mathcal{G}))^\perp.$$

That is for all $X \in L^2(\Omega, \mathcal{F}, P)$, then conditional expectation of X with respect to $\mathcal{G}$ is $E_\mathcal{G} X$. Also note that for all $A \in \mathcal{G}$, we have $I_A \in L^2(\Omega, \mathcal{G}, P_\mathcal{G})$, and hence

$$\langle X - E_\mathcal{G} X, I_A \rangle = 0,$$

which implies

$$\int_A X \mathrm{d}P = \int_A E_\mathcal{G} X \mathrm{d}P.$$

Remark 11. The Theorem above ensures that an element of $L^2(\Omega, \mathcal{F}, P)$ can be orthogonaly decomposed as $h + g$.

Definition 7.8 (Conditional Expectations with respect to set with nozero measure). 123

Definition 7.9 (Conditional Expectations with respect to set with zero measure). 123

Chapter 8

Fourier Analysis

This chapter uses Hilbert space theory to motivate the introduction of Fourier coefficients and Fourier series. The classical setting applies these concepts to functions defined on bounded intervals of the real line. However, the theory becomes easier and cleaner when we instead use a modern approach by considering functions defined on the unit circle of the complex plane.

The first section of this chapter shows how consideration of Fourier series leads us to harmonic functions and a solution to the Dirichlet problem. In the second section of this chapter, convolution becomes a major tool for the L^p theory.

The third section of this chapter changes the context to functions defined on the real line. Many of the techniques introduced in the first two sections of the chapter transfer easily to provide results about the Fourier transform on the real line. The highlights of our treatment of the Fourier transform are the Fourier Inversion Formula and the extension of the Fourier transform to a unitary operator on $L^2(\mathbf{R})$.

The vast field of Fourier analysis cannot be completely covered in a single chapter. Thus this chapter gives readers just a taste of the subject. Readers who go on from this chapter to one of the many book-length treatments of Fourier analysis will then already be familiar with the terminology and techniques of the subject.

Definition 8.1. ∂D is the unit circle in then complex plane:

$$\partial D = \{z \in \mathbb{C} : |z| = 1\}$$

Definition 8.2 (Measurable subsets of ∂D)**.**

- A subset set of ∂D is measurable $\{t \in (-\pi, \pi] : e^{it} \in E\}$ is a Borel subset of $\mathbb{R}$. The collection of all measurable subset if denoted by $\mathbf{M}$. ($\mathbf{M}$ is sigma-algebra, as you should verify)

- σ is the measure on $(\partial D, \mathbf{M})$ obtained by transferring Lebesgue measure from .

$$\sigma(E) = \frac{\lambda(\{t \in (-\pi, \pi] : e^{it} \in E\})}{2\pi} = \frac{\lambda \circ h^{-1}(E)}{2\pi}$$

Theorem 8.1. $\mathbf{M}$ is a sigma-algebra and σ is indeed a measure on $(\partial D, \mathbf{M})$

Proof. In fact, $\mathbf{M}$ can be written as

$$\left\{E : h^{-1}(E) \in B(\mathbb{R})\right\}.$$

Suppose $A \in \mathbf{M}$, then
$$h^{-1}(\partial D \backslash A) = h^{-1}(\partial D) \backslash h^{-1}(A),$$

which implies $h^{-1}(\partial D \backslash A) \in B(\mathbb{R})$. Also   □

Lemma 8.0.1. Suppose f is a map from $\mathbb{R}$ to $\mathbb{C}$, E is an element in $\mathbf{M}$, then

$$\chi_E \circ f = \chi_{f^{-1}(E)},$$

and

$$\chi_E \circ f^{-1} = \chi_{f(E)}.$$

Proof. Considering that $\chi_E \circ f(t) = 1$ if $f(t) \in E$. Also, $\chi_{f^{-1}(E)}(t) = 1$ if $t \in f^{-1}(E)$. In other words, $f(t) \in E$.   □

Theorem 8.2. Suppose $f \in L^1(\partial D)$

$$\int_{\partial D} f \mathrm{d}\sigma = \int_{(-\pi, \pi]} f(e^{it}) \frac{\mathrm{d}\lambda(t)}{2\pi}$$

Proof. Suppose $f = \chi_E$, where E is **M**-measurable. Then

$$\int_{\partial D} \chi_E \mathrm{d}\sigma = \sigma(E).$$

On the other hand, suppose $h : (-\pi, \pi] \to \partial D, \ t \mapsto \mathrm{e}^{it}$, then we have

$$
\begin{aligned}
\int_{(-\pi,\pi]} \chi_E \circ h \frac{\mathrm{d}\lambda}{2\pi} &= \int_{(-\pi,\pi]} \chi_{h^{-1}(E)} \frac{\mathrm{d}\lambda}{2\pi} \\
&= \frac{\lambda \circ h^{-1}(E)}{2\pi}.
\end{aligned}
$$

$\square$

Remark 12.

Chapter 9

Markov Process

9.1 Multi-dimensional Normal Distribution

At the start of this chapter, we will make a brief review of Linear Algebra. By the spectrum theorem of operator on $\mathbb{R}$, we conclude that a there exists a orth0normal basis such that the matrix of T is an diagonal matrix. To be specific, a positive defined matrix can be viewed as a positive self-adjoint operator and thus

Definition 9.1 (Random Vector).
Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. A random vector $\mathbf{X}$ is a map from Ω to $\mathbb{R}^n$, and its $k - th$ slot is in $L^1(\mathcal{F})$. More precisely,

$$\mathbf{X}(\omega) := \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_p(\omega) \end{bmatrix}.$$

Definition 9.2 (Mean, Corvariance).

$$E(\mathbf{X}) := \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix}, \ \Sigma_{\mathbf{X}} = Cov(\mathbf{X}, \mathbf{X}) := E(\mathbf{X} - E(\mathbf{X})(\mathbf{X} - E(\mathbf{X}))'$$

Theorem 9.1. For the mean and covariance defined above, we have the following properties:

- $E(A\mathbf{X} + B\mathbf{Y}) = AE(\mathbf{X}) + BE(\mathbf{Y})$.

- $E(\mathbf{X}C + \mathbf{Y}D) = E(\mathbf{X})C + E(\mathbf{Y})D$.

- $Cov(A\mathbf{X} + b, A\mathbf{X} + b) = ACov(\mathbf{X}, \mathbf{X})A'.$

- $E(\mathbf{X}\prime A\mathbf{X}) = \text{tr}(A\Sigma) + \mu'A\mu$

Example 9.1. Suppose $X_1, X_2, \cdots, X_p$ are i.i.d. $N(0,1)$. Then the joint density function of them is given by:

$$f_{\mathbf{X}}(x) = \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}x'x}.$$

Suppose $Y_1, Y_2, \cdots, Y_q$ are given by:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix} = A \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} + \mu,$$

where A is a bijection from $\mathbb{R}^p$ to $\mathbb{R}^p$. Then the joint density function of $Y_1, \cdots, Y_p$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{Y}-\mu)'(A^{-1})'A^{-1}(\mathbf{Y}-\mu)},$$

where $\Sigma = AA', \Sigma^{-1} = (A')^{-1}A^{-1}$. You should also note that

$$\{\mathbf{y} \in \mathbb{R}^p : \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{y}-\mu)'(A^{-1})'A^{-1}(\mathbf{y}-\mu)} = c\}$$

is a hyper-ellipse with parameter $\ln((2\pi)^p|\Sigma|c)$ on $\mathbb{R}^p$. By the spectrum theorem of a positive self-adjoint operator, we have

$$\Sigma = \lambda_1 e_1 e_1' + \cdots + \lambda_p e_p e_p',$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_p$. Note that

$$d(\mathbf{y}, \mu)^2 = \langle \Sigma(\mathbf{y} - \mu), \mathbf{y} - \mu \rangle.$$

Suppose $\mathbf{y} = \mu + a * e_1$. Then

$$d(\mathbf{y}, \mu) = \ln((2\pi)^p|\Sigma|c)$$

implies $a = \frac{1}{\sqrt{\lambda_1}} \ln((2\pi)^p|\Sigma|^p c)$.

Definition 9.3 (Moment Generating Function)**.**

Then moment generating function of a random vector is a map from $\mathbb{R}^p$ to $\mathbb{R}$:

$$M_{\mathbf{X}}(\mathbf{t}) := E(e^{\langle \mathbf{t}, \mathbf{X} \rangle}).$$

Theorem 9.2 (Properties of mgf).

$$M_{A\mathbf{X}+\mu}(\mathbf{t}) = e^{\langle \mathbf{t}, \mu \rangle} M_{\mathbf{X}}(A^{\star}\mathbf{t})$$

Definition 9.4. Suppose $X_1, \cdots, X_p$ are i.i.d. $N(0,1)$, then

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\frac{1}{2}\langle \mathbf{t}, \mathbf{t} \rangle}.$$

Also, if $\mathbf{Y} \sim N(\mu, \Sigma)$, where $\Sigma = AA'$, then

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\langle \mathbf{t}, \mu \rangle + \frac{1}{2}\langle A'\mathbf{t}, A'\mathbf{t} \rangle} = e^{\langle \mathbf{t}, \mu \rangle + \frac{1}{2}\langle \Sigma \mathbf{t}, \mathbf{t} \rangle}.$$

Definition 9.5. A random vector $\mathbf{X} = (X_1, X_2, \cdots, X_p)$ has the multivarite normal distribution $N(\mu, \Sigma)$, if its characteristic function is

$$E(e^{i\langle \mathbf{u}, \mathbf{X} \rangle}) = \exp\left(i\langle \mathbf{u}, \mathbf{X} \rangle - \frac{1}{2}\mathbf{u}'\Sigma\mathbf{u}\right).$$

Theorem 9.3 (Conditional Distribution of Multinormal distribution).
Suppose $\mathbf{X} = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}$ follows a multi-dimensional distribution $N(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{(1)(1)} & \Sigma_{(1)(2)} \\ \Sigma_{(2)(1)} & \Sigma_{(2)(2)} \end{bmatrix})$.
Then $X^{(1)}_{|x^{(2)}=a} \sim N(\mu^{(1)} + \Sigma_{(1)(2)}\Sigma^{-1}_{(2)(2)}(a - \mu^{(2)}), \Sigma_{(1)(1)} - \Sigma_{(1)(2)}\Sigma^{-1}_{(2)(2)}\Sigma_{(2)(1)})$

Proof. See Mulcond for more details. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 13. In that, $X^{(1)}_{|X^{(2)}} - \mu_1$ is the projection of ...

Theorem 9.4. Suppose $X \sim N(\mu, \Sigma)$, $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$. Then

$$AX + \mathbf{b} \sim N(A\mu + \mathbf{b}, A\Sigma A')$$

9.2   Sample Geometry

Definition 9.6. A stochastic process $X = \{X_t, t \geq 0\}$ is a famlily of random variables

$$X_t : \Omega \to \mathbb{R}, \omega \mapsto X_t(\omega)$$

defined on a probability space $(\Omega, \mathcal{F}, P)$.

Note that a famlily is function from ...

Definition 9.7. We can also define a stochastic process on a product measure space $(\Omega \times \mathbb{R}_+, \mathcal{F} \otimes B(\mathbb{R}_+), P \times \lambda)$. The stochastic process is a measurable function defined on $\Omega \times \mathbb{R}_+$,

$$X : \Omega \times \mathbb{R}_+ \to \mathbb{F}, \ (t, \omega) \mapsto X(t, \omega).$$

Definition 9.8. The finite-dimensional marginal distributions of process X is defined by

$$\mu_{t_1, \cdots, t_n} = P \circ (X_{t_1}, \cdots, X_{t_n})^{-1}.$$

Where $(X_{t_1}, \cdots, X_{t_n})^{-1}$ is a map from $B(\mathbb{R})^n$ to $\mathcal{F}$,

$$(X_{t_1}, \cdots, X_{t_n})^{-1} : \ (B_1, \cdots, B_n) \mapsto \bigcap_{i=1}^{n} X_{t_i}^{-1}(B_i)$$

Theorem 9.5. $P_{t_1, \cdots, t_n}$ is a probability measure on $(\mathbb{R}^n, B(\mathbb{R}^n))$.

Definition 9.9. For every $\omega \in \Omega$, the mapping

$$X(\cdot, \omega) : \ t \mapsto X(t, \omega)$$

is called a trajectory of the process X.

Remark 14. Easy to verify that the trajectory of the process X is a measurable function on $(\mathbb{R}, B(\mathbb{R}))$.

Definition 9.10. Two processed X and Y are equivalent (or X is a version of Y) if for all $t \geq 0$,

$$P\{X_t = Y_t\} = 1$$

indistinguishable.

Theorem 9.6 (Kolmogorov's contiuity theorem)**.**

9.3   Brownian Motion

Definition 9.11 (Brownian Motion)**.**
A stochastic process B on $(\Omega \times \mathbb{R}_+, \mathcal{F} \otimes B(\mathbb{R}_+))$ is called a Brownian motion if

1. $B(0, \cdot) = 0$ almost surely.

2. B have independent and stationary increments.

3. The increment of B has the normal distribution.

4. $B(\cdot, \omega)$ is continuous for a.s. $\omega \in \Omega$.

Definition 9.12 (Filtration generated by Bronian motion.)**.**

$$\sigma(B_t) = \sigma(\bigcup_{B \in B(\mathbb{R})} \{\omega : B_t(\omega) \in B\})$$

Clearly $\sigma(B_t) \subset \mathcal{F}$.

$$\mathcal{F}_t := \sigma(\bigcup_{s \leq t} \sigma(B_t))$$

Theorem 9.7 (Construction of Brownian Motion)**.**

9.4 Other topics

Definition 9.13 (Filtration)**.**
A filtration $\{\mathcal{F}_t \subset \mathcal{F}, t \geq 0\}$ is an increasing family σ-algebras.

Definition 9.14 (Adapted Process)**.**
A stochastic process X is called an adapted if $X(t, \cdot)$ is $\mathcal{F}_t$-measurable.

Definition 9.15 (Markov process)**.**
An adapted process X is called a Markov process with respect to $\{\mathcal{F}_t\}$ if for all $s, t > 0$ and f (f is Borel-measurable function),

$$E(f(X_{s+t})|\mathcal{F}_t) = E(f(X_{s+t})|X_s)$$

Definition 9.16 (Transition Probabilities)**.**

$$P(t, x, t + s, B) = P(X_{t+s} \in B|X_s = x),$$

where B is a Borel subset.

Definition 9.17 (Filtration generated by a process)**.**
Suppose X is a stochastic process. Then the filtration generated by X is defined by

$$\mathcal{F}_t^X := \{\}$$

Chapter 10

Stochastic Calculus

10.1 Integral with respect to Brownian Motion

In this chapter, we will the define the Ito integral of some special stochastic process(Progessively measurable).

Recall that we define the stochastic process $X : X \times [0, \infty) \to \mathbb{R}$ to be a $\mathcal{F} \otimes B([0, \infty)) - measurable$ function.

B is a Brownian motion on $(\Omega \times \mathbb{R}_+, \mathcal{F} \otimes B(\mathbb{R}_+))$, and $\mathcal{F}_t$ is the filtration generated by Brownian motion.

Definition 10.1. X is called processively measurable if for all $t \geq 0$, the restriction of X to $\Omega \times [0, t]$ is $\mathcal{F}_t \otimes B([0, t]) - measurable$.

$\mathcal{P}$ is defined to be a σ-algebra of Set $A \subset \mathbb{R}^+ \times \Omega$ such that I_A is progressively measurable.

Example 10.1. Step function $\varphi(t, \omega) = \sum_{j=0}^{m-1} a_j \chi_{(t_j, t_{j+1}]}(t)$ is progressive measurable. Moreover, φ is $\mathcal{P} - measurable$.

Example 10.2. Simple process $\varphi(t, \omega) = \sum_{j=0}^{m-1} \phi_j \chi_{(t_j, t_{j+1}]}(t)$ is progressive measurable, if ϕ_j is $\mathcal{F}_{t_j} - measurable$. Moreover, φ is $\mathcal{P} - measurable$.

Proof. for every $B \in B(\mathbb{R}^+)$, we have

$$
\begin{aligned}
\{(s, \omega) : \varphi(s, \omega) I_{(0, t]} \in B\} &= \bigcup_{i=0}^{m-1} ((t_i \wedge t, t_{i+1} \wedge t] \times \phi_i^{-1}(B)) \\
&= (t_0 \wedge t, t_m \wedge t] \times \bigcup_{i=0}^{m-1} \phi_i^{-1}(B). X
\end{aligned}
$$

The set above is clearly $B(0, t] \otimes \mathcal{F}_t - measurable$.

$\square$

Example 10.3. B is progressively measurable, where B is a Brownian motion. Moreover, if $f : \mathbb{R} \to R$ is a borel measurable function, then $f \circ B$ is progressively measurable.

Proof.
$$\{(s, \omega) : B(s, \omega) I_{(0,t](s)} \in D\} = (0, t] \times \bigcap_{r \leq t} B_r^{-1}(D)$$

and

$$\bigcap_{r \leq t} B_r^{-1}(D) \in \mathcal{F}_t$$

$\square$

Definition 10.2.
$$\mathcal{S} := \{\phi : \phi \text{ is simple process}\}$$

Definition 10.3.
$$L^2(\Omega \times \mathbb{R}_+, \mathcal{P}, P \times \lambda) = L^2(\mathcal{P})$$

and L^2 norm is defined by

$$\|X\| = \int_{\Omega \times B(\mathbb{R}^+)} X^2 \, dP \times \lambda.$$

and Fubini's theorem yields

$$\|X\| = \int_{\Omega} \int_{B(\mathbb{R}^+)} X^2 \, d\lambda dP$$

Definition 10.4 (Ito Integral of simple process)**.**
Suppose $u \in \mathcal{S}$, $u = \sum_{j=0}^{m-1} \phi_j \chi_{(t_j, t_{j+1}]}$. The Ito integral of of u is defined by

$$I(u) = \int_{\mathbb{R}^+} u_t dB_t := \sum_{j=0}^{m-1} \phi_j (B_{t_{j+1}} - B_{t_j})$$

Remark 15. A quick scan of I is

$$I(\chi_{(a,b]}) = B_b - B_a$$

$$I(F\chi_{(a,b]}) = F(B_b - B_a)$$

if $F \in L^2(\Omega)$ and $\mathcal{F}_a - measurable$

Ito integral is a isometry from $\mathcal{S}$ to $L^2(\Omega, \mathcal{F}, P)$.

Theorem 10.1 (Properities of Ito Integral on $\mathcal{S}$)**.**
Suppose $u \in \mathcal{S}$, then

- I is a linear map from S to $L^2(\Omega)$.

- zero mean: $E(I(u)) = 0$

- Isometry Property: $\|u\| = \|I(u)\|$ and thus $\|I\| = 1$.

- Isometry Property II: $\langle u, v \rangle = \langle I(u), I(v) \rangle$

Proof. Suppose $u = \sum_{j=0}^{m-1} \phi_j \chi_{(t_j, t_{j+1}]}$, then

$$
\begin{aligned}
E(I(u)) &= E(\sum_{j=0}^{m-1} \phi_j (B_{t_{j+1}} - B_{t_j})) \\
&= E(E(\sum_{j=0}^{m-1} \phi_j (B_{t_{j+1}} - B_{t_j}) | \mathcal{F}_{t_j})) \\
&= 0
\end{aligned}
$$

Also note that,
$$
\|u\| = \int_{B(\mathbb{R}^+) \times \Omega} u \, d\lambda \times P
$$

$$
\begin{aligned}
\|u\|^2 &= \int_{B(\mathbb{R}^+) \times \Omega} u^2 \, d\lambda \times P \\
&= \int_{B(\mathbb{R}^+) \times \Omega} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \phi_i \phi_j I_{(t_i, t_{i+1}]} I_{(t_j, t_{j+1}]} \, d\lambda \times P \\
&= \sum_{i=0}^{m-1} \int \phi_i^2 I_{(t_i, t_{i+1}]} \\
&= \sum_{i=0}^{m-1} E(\phi_i^2)(t_{i+1} - t_i)
\end{aligned}
$$

$$\begin{aligned}
\|I(u)\|^2 &= E(\sum_{j=0}^{m-1} \phi_j(B_{t_{j+1}} - B_{t_j}))^2 \\
&= E(\sum_{i=0}^{m-1}\sum_{j=0}^{m-1} \phi_i\phi_j(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})) \\
&= \sum_{i=0}^{m-1} E(\phi_i^2)(t_{i+1} - t_i)
\end{aligned}$$

At last,

$$\begin{aligned}
2\langle I(u), I(v)\rangle &= \langle I(u) + I(v), I(u) + I(v)\rangle - \langle I(u), I(u)\rangle - \langle I(v), I(v)\rangle \\
&= \langle u + v, u + v\rangle - \langle u, u\rangle - \langle v, v\rangle \\
&= 2\langle u, v\rangle
\end{aligned}$$

$\square$

Theorem 10.2. $\mathcal{S}$ is dense in $L^2(\mathcal{P})$.

Remark 16. A consequence of theorem 2.2 is for all $u \in L^2(\mathcal{P})$, there exists $u_n : u_n \in \mathcal{S}$, such that

$$\lim_{n \to \infty} u_n = u.$$

Now consider $I(u_n)$.

$$\|I(u_n) - I(u_m)\|_2 = \|u_n - u_m\|_2$$

and thus $I(u_n)$ is a cauchy sequence in $L^2(\Omega)$. The completeness of Hilbert space admits there exists $I(u)$, such that

$$\lim_{n \to \infty} \|I(u_n) - I(u)\|_2 = 0$$

Theorem 10.3 (Properties of Integral defined on $L^2(\mathcal{P})$)**.**

- I is a linear map from S to $L^2(\Omega)$.

- zero mean: $E(I(u)) = 0$

- Isometry Property I: $\|u\| = \|I(u)\|$

- Isometry Property II: $\langle u, v\rangle = \langle I(u), I(v)\rangle$

Proof. Recall that the continuity is equivalent to boundness for linear maps from a normed vector space to another. Thus $u_n \to u$ implies

$$\lim_{n \to \infty} I(u_n) = I(u).$$

For thm 1,

$$E(I(u)) = E(\lim_{n \to \infty} I(u_n)) = \lim_{n \to \infty} E(I(u_n)) = 0,$$

where the second equality follows from Dominated Convergence Theorem.

For thm 2,

$$\|I(u)\|^2 = \|\lim_{n \to \infty} I(u_n)\| = \lim_{n \to \infty} \|(I(u_n)\| = \lim_{n \to \infty} \|u_n\| = \|u\|,$$

A direct consequence of thm 2 yields thm 3. $\square$

Definition 10.5 (Indefinite Ito Integtal). For $u \in L^2(\mathcal{P})$

$$I(u, a, b) = \int_a^b u_s \mathrm{d}B_s := I(uI_{(a,b]})$$

$$I(u, t) = \int_0^t u_s \mathrm{d}B_s := I(uI_{(0,t]})$$

Theorem 10.4 (Properties of Indefinite Integral).

By definition,

- Additivity

- Factorization $I(Fu, a, b) = FI(u, a, b)$ if F is $\mathcal{F}_a - measurable$.

- Martingale Property

Proof. First suppose $u \in \mathcal{S}$, $u = \sum_{j=0}^{m-1} \phi_j \chi_{(t_j, t_{j+1}]}$, F is $\mathcal{F}_a - measurable$. Then

$$FuI_{(a,b]} = \sum_{j=0}^{m-1} F\phi_j \chi_{(a \vee t_j \wedge b, a \vee t_{j+1} \wedge b]}$$

is a simple process. Thus the inregral of $FuI_{(a,b]}$ is defined by

$$
\begin{aligned}
I(Fu, a, b) &= \sum_{j=0}^{m-1} F\phi_j (B_{a \vee t_{j+1} \wedge b} - B_{a \vee t_j \wedge b}) \\
&= F \sum_{j=0}^{m-1} \phi_j (B_{a \vee t_{j+1} \wedge b} - B_{a \vee t_j \wedge b}) \\
&= FI(u, a, b)
\end{aligned}
$$

$$\begin{aligned}
I(u,t) &= I(uI_{(0,t]}) \\
&= \sum_{j=0}^{m-1} I(\phi_j \chi_{(t_j \wedge t, t_{j+1} \wedge t]}) \\
&= \sum_{j=0}^{m-1} \phi_j I(\chi_{(t_j \wedge t, t_{j+1} \wedge t]}) \\
&= \sum_{j=0}^{m-1} \phi_j (B_{t_{j+1} \wedge t} - B_{t_j \wedge t}).
\end{aligned}$$

$$\begin{aligned}
E(I(u,t)|\mathcal{F}_s) &= \sum_{j=0}^{m-1} E(E(\phi_j(B_{t_{j+1} \wedge t} - B_{t_j \wedge t})|\mathcal{F}_{t_j \vee s})|\mathcal{F}_s) \\
&= \sum_{j=0}^{m-1} E(\phi_j E((B_{t_{j+1} \wedge t} - B_{t_j \wedge t})|\mathcal{F}_{t_j \vee s})|\mathcal{F}_s) \\
&= \sum_{j=0}^{m-1} \phi_j (B_{t_{j+1} \wedge s} - B_{t_j \wedge s}) \\
&= I(u,s)
\end{aligned}$$

Where the last two equation follows from $(t_j \wedge t) \wedge (t_j \vee s) \wedge s = t_j \wedge s$ and $(t_{j+1} \wedge t) \wedge (t_j \vee s) \wedge s = t_j \wedge s$.

Since the conditional expectation and the limit operation are commutative, we conclude that $I(u,t)$ is a martingale with respected to the Filtration generated by Brownian Motion for all $u \in L^2(\mathcal{P})$. $\square$

Theorem 10.5. $I(u,t)$ is a square integrable martingale with respect to the filtration generated by Brownian Moton and admits a continuous version.

Proof. The martingale property of $I(u,t)$ follows from 10.4, and the Isometry property admits that $I(u,t)$ is square integrable when fixing t. $\square$

Theorem 10.6. I(u,t)

Definition 10.6. 123

10.2   Integral with respect to local martingales