Lecture Notes

Introduction to Stochastic Differential Equations

Eckart Schulz

School of Mathematics
Suranaree University of Technology

April, 2551

Contents

Preface

These notes comprise an introductory course in stochastic differential equations given in term 2/2550. The purpose is to introduce an audience of graduate students with no prior knowledge of measure theory or probability to the concepts of stochastic integral, Itô formula, and solutions of linear stochastic differential equations, in order to let them understand what this subject is about, and to foster interest for further study. Thus, severe limitations had to be imposed on the material presented.

The course begins with an introduction to measure theory. Naturally there is no opportunity to prove a some major theorems such as Fubini's theorem or the Radon-Nikodym theorem; however, care is taken to introduce the notions of measure and Lebesgue integral at a slow and detailed pace in order to develop the student's insight and understanding. Then the basic probabilistic concepts such as independence and conditional expectation are introduced. We present simple examples in order to help the student develop probabilistic intuition. The part on stochastic processes and the Itô integral is fairly standard although for the sake simplicity, the Itô integral is discussed for square integrable processes only. With very few exceptions, most theorems on stochastic processes being used are also proved. In particular we present a proof of the general Itô formula, however with a boundedness restriction necessary to avoid the difficult concept of stopping time. The course ends with a brief encounter of reducible stochastic differential equations, applied to linear equations.

As a general rule, proofs are layed out in great detail. While his may have resulted in loss of elegance of presentation, it should make the material better accessible to the intended audience. At any rate, at the end of a course a student can be expected to prepare a summary in his or her own manner and extract the main ideas of the proofs. The text contains numerous exercises which constitute an essential part of the course and should all be attempted by the student.

Chapter 1

Background from Measure Theory

The motivating idea of measure is to generalize the well-known concept of "area" or "volume" of sets in the plane or in three-dimensional space to subsets of general spaces. Intuitively, the volume of subsets of $\mathbb{R}^d$ has the following properties:

1. If A, B are subsets of $\mathbb{R}^d$ whose volumes can be computed and are finite, then
$$\mathrm{vol}(A \cup B) = \mathrm{vol}(A) + \mathrm{vol}(B) - \mathrm{vol}(A \cap B).$$

2. If $\{A_n\}_{n=1}^{\infty}$ is a collection of disjoint subsets of $\mathbb{R}^d$ whose volumes can be computed, then
$$\mathrm{vol}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathrm{vol}(A_n).$$

Now given a general set Ω, we need to

i) describe the class of subsets of Ω whose "volume" we can compute – this will be the concept of *σ-algebra*, and

ii) describe the notion of "volume" on these sets satisfying properties 1. and 2. above – this will be the concept of *measure*.

1.1 Measurable Spaces

Definition 1.1. Let Ω be a non-empty set. A collection $\mathcal{F}$ of subsets of Ω is called a *σ-algebra on* Ω, if the following properties hold:

(S1) $\emptyset \in \mathcal{F}$

(S2) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$

(S3) If $A_1, A_2, A_3, \dots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a *measurable space*. Elements of $\mathcal{F}$ are called *measurable sets*.

Remark 1.1. Let $\mathcal{F}$ be a σ-algebra on Ω.

1. To clarify this concept, elements of a σ-algebra $\mathcal{F}$ are *sets*, namely the sets to whom we want to assign a "volume".

2. Since $\Omega = \emptyset^c$, then by (S1) and (S2), $\Omega \in \mathcal{F}$ always.

3. Let $A_1, A_2, A_3, \dots \in \mathcal{F}$. Then by $(S2)$, $A_1^c, A_2^c, A_3^c, \dots \in \mathcal{F}$ as well. Thus we obtain from (S2) that

$$\text{(S4)} \qquad \bigcap_{n=1}^{\infty} A_n = \Big(\bigcup_{n=1}^{\infty} A_n^c\Big)^c \in \mathcal{F}.$$

4. Obviously, (S3) and (S4) also hold for *finite collections* of elements of $\mathcal{F}$. For let $A_1, A_2, A_3, \dots, A_N \in \mathcal{F}$. Setting $A_n := A_N$ for $n > N$ we obtain

$$\text{(S3')} \qquad \bigcup_{n=1}^{N} A_n = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}, \quad \text{and}$$

$$\text{(S4')} \qquad \bigcap_{n=1}^{N} A_n = \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}.$$

5. Let $A, B \in \mathcal{F}$. Then by (S2) and (S4'),

$$\text{(S5)} \qquad A \backslash B = A \cap B^c \in \mathcal{F}.$$

Thus, a σ-algebra is *closed* under formation of differences, and of countable unions and intersections.

(Here we make the convention that countable means either *finite*, or *countably infinite.*)

Example 1.1. Let us give some simple examples of σ-algebras:

1. Let Ω be a non-empty set. The following are σ-algebras on Ω:

   (a) $\mathcal{F}_1 := \{\emptyset, \Omega\}$. (This is the smallest σ-algebra on Ω.)

   (b) $\mathcal{F}_2 := \mathcal{P}(\Omega)$. (This is the largest σ-algebra on Ω.) Here $\mathcal{P}(\Omega)$, also denoted by 2^Ω, is the power set of Ω.

   (c) $\mathcal{F}_3 := \{\emptyset, E, E^c, \Omega\}$ where E is an arbitrary, fixed subset of Ω. (This is the smallest σ-algebra on Ω containing E.)

   (d) More generally, let $\{E_i\}_{i=1}^N$ (where $N \in \mathbb{N}$ or $N = \infty$) be a countable collection of subsets of Ω satisfying

$$E_i \cap E_j = \emptyset \quad \text{for } i \neq j \qquad \text{and} \qquad \bigcup_{i=1}^{N} E_i = \Omega.$$

   (We call such a collection $\{E_i\}_{i=1}^N$ a *partition* of Ω.) Then

$$\mathcal{F}_4 := \{A \subseteq \Omega : A \text{ is the union of some sets } E_i, \text{ or } A = \emptyset\ \}$$

   is a σ-algebra on Ω. The sets E_i are called its *atoms*.

2. Let Ω be an infinite set.

   (a) First set
   $$\mathcal{F}_5 := \{A \subseteq \Omega : A \text{ is finite or } A^c \text{ is finite}\}.$$
   Then $\mathcal{F}_5$ is *not* a σ-algebra.

   For suppose to the contrary that $\mathcal{F}_5$ is a σ-algebra. Pick a countably infinite subset $\{x_1, x_2, x_3, x_4, \dots\}$ of Ω, and set
   $$A := \{x_2, x_4, x_6, x_8, \dots\}.$$
   Since each singleton $\{x_k\}$ is finite, then $\{x_k\} \in \mathcal{F}_5$ and hence by (S3), $A \in \mathcal{F}_5$. However, A is not a finite set, hence A^c must be finite. On the other hand,
   $$A^c \supseteq \{x_1, x_3, x_5, \dots\}$$
   contradicting finiteness of A^c. Thus, $\mathcal{F}_5$ cannot be a σ-algebra.

   (b) Next set
   $$\mathcal{F}_6 := \{A \subseteq \Omega : A \text{ is countable or } A^c \text{ is countable}\}.$$
   It is an easy exercise to show that $\mathcal{F}_6$ is a σ-algebra.

3. Let $\Omega = \mathbb{R}^d$ and set
   $$\mathcal{O} := \{A \subseteq \mathbb{R}^d : A \text{ is open}\}.$$
   Then $\mathcal{O}$ is *not* a σ-algebra. In fact, if $A \in \mathcal{O}$ then A^c is closed. But nontrivial closed subsets of $\mathbb{R}^d$ are never open, hence $A^c \notin \mathcal{O}$. That is, (S2) does not hold.

   We thus need to work with a larger class of subsets of $\mathbb{R}^d$, which will be introduced at the end of this section. The idea is to use the 'smallest' σ-algebra containing $\mathcal{O}$. First we must clarify what is meant by 'smallest' here.

Exercise 1.1. Prove that $\mathcal{F}_4$ and $\mathcal{F}_6$ are σ-algebras.

Theorem 1.1. *Let $\{\mathcal{F}_\lambda\}_{\lambda\in\Lambda}$ be a family of σ-algebras on a set Ω. Then*
$$\mathcal{F} := \bigcap_{\lambda\in\Lambda} \mathcal{F}_\lambda$$
is also a σ-algebra on Ω.

Proof. Let's simply verify the three axioms for a σ-algebra.

(S1): Since $\emptyset \in \mathcal{F}_\lambda \ \forall\, \lambda$, then $\emptyset \in \bigcap_{\lambda\in\Lambda} \mathcal{F}_\lambda = \mathcal{F}$.

(S2): Let $A \in \mathcal{F}$. Then $A \in \mathcal{F}_\lambda, \forall\, \lambda$. Now as each $\mathcal{F}_\lambda$ satisfies (S2), then $A^c \in \mathcal{F}_\lambda \ \forall\, \lambda$. Hence, $A \in \bigcap_{\lambda\in\Lambda} \mathcal{F}_\lambda = \mathcal{F}$.

(S3): Let $A_1, A_2, A_3, \dots \in \mathcal{F}$. Then $A_n \in \mathcal{F}_\lambda \; \forall \lambda, \; \forall n$. Since each $\mathcal{F}_\lambda$ satisfies (S3), then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_\lambda \; \forall \lambda$. Hence, $\bigcup_{n=1}^{\infty} A_n \in \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda = \mathcal{F}$.

□

Definition 1.2. Let $\mathcal{K}$ be a collection of subsets of Ω. Denote by $\{\mathcal{F}_\lambda\}_{\lambda \in \Omega}$ the collection of *all* σ-algebras on Ω containing $\mathcal{K}$ (that is all σ-algebras $\mathcal{F}_\lambda$ satisfying $\mathcal{K} \subseteq \mathcal{F}_\lambda$). Set

$$\sigma(\mathcal{K}) := \bigcap_{\lambda \in \Lambda} \mathcal{F}_\lambda. \tag{1.1}$$

By theorem 1.1, $\sigma(\mathcal{K})$ is a σ-algebra on Ω. Obviously, $\mathcal{K} \subseteq \sigma(\mathcal{K})$. Observe that if $\mathcal{F}$ is *any* σ-algebra containing $\mathcal{K}$, then $\mathcal{F} = \mathcal{F}_\lambda$ for some λ and hence by (1.1), $\sigma(\mathcal{K}) \subseteq \mathcal{F}$. Hence, $\sigma(\mathcal{K})$ is the *smallest* (in the sense of inclusion) σ-algebra on Ω containing $\mathcal{K}$. We call $\sigma(\mathcal{K})$ the *σ-algebra generated by $\mathcal{K}$*.

Example 1.2. 1. Let Ω be any non-empty set, and $E \subseteq \Omega$. Then

$$\sigma(\{E\}) = \{\emptyset, E, E^c, \Omega\}.$$

2. Let Ω be any infinite set, and let

$$\mathcal{K} = \{A \subseteq \Omega : A \text{ is finite}\}.$$

Then

$$\sigma(\mathcal{K}) = \{A \subseteq \Omega : A \text{ is countable or } A^c \text{ is countable}\}.$$

3. (Generalization of 1.) Let Ω be a non-empty set, and $\{E_i\}_{i=1}^N$ ($N \in \mathbb{N}$ or $N = \infty$) a partition of Ω. Then

$$\sigma(\{E_i\}_{i=1}^N) = \{A \subseteq \Omega : A \text{ is the union of some sets } E_i, \text{ or } A = \emptyset\}.$$

Exercise 1.2. Prove 2. and 3. above.

Definition 1.3. Let X be a topological space (e.g. a metric space) and $\mathcal{O} := \{A \subseteq X : A \text{ is open}\}$ the collection of open subsets. Then

$$\sigma(\mathcal{O}) = \ \sigma\text{-algebra generated by the open sets in } X$$

is called the *Borel σ-algebra on X*, denoted by $\mathcal{B}(X)$. Its elements are called *Borel sets*.

Remark 1.2. 1. There is no easy characterization of Borel subsets of $\mathbb{R}^d$. However, the following are typical Borel sets, as can easily be seen by use of (S2)–(S4):

a) every *open* subset E of $\mathbb{R}^d$ (as $\mathcal{O} \subseteq \mathcal{B}(\mathbb{R}^d)$).

b) every *closed* subset F of $\mathbb{R}^d$ (as $F = E^c$ for some open set E).

c) every set of the form $A = \bigcap_{n=1}^{\infty} E_n$, E_n open in $\mathbb{R}^d$. (Such sets are called *G_δ-sets.*)

d) every set of the form $A = \bigcup_{n=1}^{\infty} F_n$, F_n closed in $\mathbb{R}^d$. (Such sets are called *F_σ-sets.*)

2. We know by Lindelöff's theorem that every open subset E of $\mathbb{R}^d$ is a countable union of open balls, $E = \bigcup_{n=1}^{\infty} B_{\epsilon_n}(x_n)$. (we even may choose x_n to have rational coordinates, and ϵ_n to be rational.) Hence, $\mathcal{B}(\mathbb{R}^d)$ is also the σ-algebra generated by the collection of open balls.

3. In case of $\mathbb{R}$, the Borel σ-algebra is generated by infinite intervals with rational endpoints,

(a)	$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{K}_1)$	where	$\mathcal{K}_1 = \{(-\infty, a) : a \in \mathbb{Q}\}$
(b)	$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{K}_2)$	where	$\mathcal{K}_2 = \{(-\infty, a] : a \in \mathbb{Q}\}$
(c)	$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{K}_3)$	where	$\mathcal{K}_3 = \{(a, \infty) : a \in \mathbb{Q}\}$
(d)	$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{K}_4)$	where	$\mathcal{K}_4 = \{[a, \infty) : a \in \mathbb{Q}\}$

Obviously, in the above we may replace "$\mathbb{Q}$" by "$\mathbb{R}$". Let us prove (a); the remaining identities are proved similarly and left as an exercise.

Since $\mathcal{K}_1 \subseteq \mathcal{O}$ then obviously,

$$\sigma(\mathcal{K}_1) \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\mathbb{R}).$$

For the reverse inclusion, we need to show that $\mathcal{O} \subseteq \sigma(\mathcal{K}_1)$. For then it will follow that

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{O}) \subseteq \sigma(\mathcal{K}_1).$$

Thus, let $E \in \mathcal{O}$, that is, E is an open subset of $\mathbb{R}$.
Case 1: E is a bounded, open interval with rational endpoints, i.e. $E = (a, b)$, with $a, b \in \mathbb{Q}$. Choose N such that $\frac{1}{N} < b - a$ and write

$$(a, b) = \bigcup_{n=N}^{\infty} [a + \frac{1}{n}, b) = \bigcup_{n=N}^{\infty} \big[(-\infty, b)\backslash(-\infty, a + \frac{1}{n})\big].$$

Since $(-\infty, b), (-\infty, a + \frac{1}{n}) \in \mathcal{K}_1$, then $(a, b) \in \sigma(\mathcal{K}_1)$ by (S5) and (S3).
Case 2: If E is an arbitrary open set, then by Lindelöff's theorem, we can write

$$E = \bigcup_{k=1}^{\infty} I_k$$

where each I_k is a bounded, open interval with rational endpoints. By case 1, $I_k \in \sigma(\mathcal{K}_1)$ for all k. Hence by (S3), $E \in \sigma(\mathcal{K}_1)$.

4. Since intervals are Borel sets, it is an easy exercise to show that for every interval I,

$$\mathcal{B}(I) = \{\, E \in \mathcal{B}(\mathbb{R}) : E \subseteq I \,\}.$$

1.2 Measures and Measure Spaces

Definition 1.4. Let $(\Omega, \mathcal{F})$ be a measurable space (that is, $\mathcal{F}$ is a σ-algebra of subsets of Ω). A *measure* on $(\Omega, \mathcal{F})$ is a function

$$\mu : \mathcal{F} \to [0, \infty]$$

satisfying

(M1) $\mu(\emptyset) = 0$,

(M2) If $A_1, A_2, A_3, \ldots \in \mathcal{F}$ is a *countable* collection of pairwise *disjoint* sets, then

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n) \qquad \text{("}\sigma\text{-additivity").}$$

The triple $(\Omega, \mathcal{F}, \mu)$ is called a *measure space.*

Remark 1.3. In the above definition, $[0, \infty] := (0, \infty) \cup \{\infty\}$. Furthermore, if the right-hand series diverges or any of its terms has the value ∞, then the series will be assigned the value ∞.

Remark 1.4. If $\mu(\Omega) < \infty$, then μ is called a *finite measure*, and $(\Omega, \mathcal{F}, \mu)$ a *finite measure space.* (In this case, $\mu(E) < \infty$ for all $E \in \mathcal{F}$ by the next theorem.)

If there exists a countable collection $A_1, A_2, A_3, \ldots$ of sets in $\mathcal{F}$ satisfying

$$\Omega = \bigcup_{n=1}^{\infty} A_n \qquad \text{and} \qquad \mu(A_n) < \infty \quad \forall n$$

then μ is called a *σ-finite measure* and $(\Omega, \mathcal{F}, \mu)$ a *σ-finite measure space.*

Remark 1.5. (M2) implies that for any *finite* collection of pairwise *disjoint* sets $A_1, A_2, A_3, \ldots, A_N \in \mathcal{F}$,

(M2')
$$\mu\Big(\bigcup_{n=1}^{N} A_n\Big) = \sum_{n=1}^{N} \mu(A_n) \qquad \text{("additivity").}$$

To see this, simply set$A_n = \emptyset$ for all $n > N$, and apply (M2) using the fact that $\mu(A_n) = \mu(\emptyset) = 0$ for all $n > N$.

Example 1.3. Let Ω be any set, and $\mathcal{F}$ be any σ-algebra on Ω. One can introduce several measures onto $(\Omega, \mathcal{F})$:

1. Two *trivial measures* are given by setting

   i) $\mu(E) = 0$ for all $E \in \mathcal{F}$, or

   ii) $\mu(\emptyset) = 0$ and $\mu(E) = \infty$ for all $E \in F,\ E \neq \emptyset$.

Note that if we set

$$\mu(\emptyset) = 0, \qquad \mu(E) = 1 \quad \forall\, E \in \mathcal{F},\ E \neq \emptyset$$

then in general we do not obtain a measure as (M2) is not satisfied.

2. The *counting measure* is defined by

$$\mu(E) = \begin{cases} \text{card}(E) & \text{if } E \in \mathcal{F} \text{ is finite} \\ \infty & \text{if } E \in \mathcal{F} \text{ is infinite.} \end{cases}$$

Observe that μ is a finite measure $\Leftrightarrow$ Ω is a finite set.
Furthermore, μ is a σ-finite measure $\Leftrightarrow$ Ω is a countable set.

3. Fix a point $\omega \in \Omega$. The *Dirac point measure* δ_w is defined by

$$\delta_w(E) = \begin{cases} 1 & \text{if } \omega \in E \\ 0 & \text{if } \omega \notin E \end{cases}$$

for $E \in \mathcal{F}$. Obviously, δ_w is finite.

Exercise 1.3. Verify that 2. and 3. above are measures.

Example 1.4. Consider the measurable space $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$. Any measure μ on this space is called a *Borel measure.* One can show the following: (see standard text books on measure theory, such as [2] or [3], for example.)

There exists a *unique* measure λ on $\mathcal{B}(\mathbb{R}^d)$ having the property that

$$\lambda(I) = \text{vol}(I) = \prod_{i=1}^{d} (b_i - a_i)$$

for every bounded d-interval $I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$. (The intervals (a_i, b_i) may also be open, or half-open.) λ is called the *Lebesgue measure* on $\mathbb{R}^d$. It has the following additional properties:

1. λ is defined on a σ-algebra $\mathcal{M}_\lambda$ on $\mathbb{R}^d$, with $\mathcal{B}(\mathbb{R}^d) \subsetneq \mathcal{M}_\lambda$.

2. Given $E \in \mathcal{M}_\lambda$ and $x \in \mathbb{R}^d$, set $x + E := \{x + y : y \in E\}$. Then $x + E \in \mathcal{M}_\lambda$ and

$$\lambda(x + E) = \lambda(E). \qquad \text{(translation invariance)}$$

3. Given $E \in \mathcal{M}_\lambda$, set $-E := \{-y : y \in E\}$. Then $-E \in \mathcal{M}_\lambda$ and

$$\lambda(-E) = \lambda(E). \qquad \text{(inversion invariance)}$$

4. λ is compatible with the topology on $\mathbb{R}^d$ in the following sense.

(a) For all $K \subseteq \mathbb{R}^d$ compact, $\lambda(K) < \infty$

(b) For all $E \in \mathcal{M}_\lambda$,

$$\lambda(E) = \inf\{\lambda(U) : U \text{ is open and } E \subseteq U\} \qquad \text{("outer regularity")}$$

(c) For all $E \in \mathcal{M}_\lambda$,

$$\lambda(E) = \sup\{\lambda(K) : K \text{ is compact and } K \subseteq E\} \quad \text{("inner regularity")}$$

Theorem 1.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.*

1. *For all $E, F \in \mathcal{F}$ with $E \subseteq F$ we have*

$$\mu(E) \leq \mu(F). \qquad \text{("monotonicity")}$$

2. *For all $E, F \in \mathcal{F}$ with $E \subseteq F$ and $\mu(F) < \infty$ we have*

$$\mu(F \backslash E) = \mu(F) - \mu(E).$$

3. *For every countable collection $A_1, A_2, A_3, \ldots \in \mathcal{F}$ we have*

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} \mu(A_n) \qquad \text{("}\sigma\text{-subadditivity")}$$

Proof. 1. Write $F = E \cup (F \backslash E)$, a disjoint union. Then

$$\mu(E) \leq \mu(E) + \mu(F \backslash E) \underset{\text{(M2')}}{=} \mu\big(E \cup (F \backslash E)\big) = \mu(F) \tag{1.2}$$

2. If $\mu(F) < \infty$, then by (1.2), $\mu(E) < \infty$ and we can subtract $\mu(E)$ from (1.2) to obtain

$$\mu(F \backslash E) = \mu(F) - \mu(E).$$

3. Let $\{A_n\}_{n=1}^{\infty} \subseteq F$. Set

$$\begin{aligned} B_1 &:= A_1 \in \mathcal{F} \\ B_2 &:= A_2 \backslash A_1 \in \mathcal{F} \\ &\vdots \quad \vdots \end{aligned}$$

and in general

$$B_n := A_n \backslash \bigcup_{i=1}^{n-1} A_i \in \mathcal{F}. \tag{1.3}$$

Observe that for each N,

$$\bigcup_{n=1}^{N} B_n = \bigcup_{n=1}^{N} A_n. \tag{1.4}$$

To see this, note that since $B_n \subseteq A_n$ for all n, then $\bigcup_{n=1}^{N} B_n \subseteq \bigcup_{n=1}^{N} A_n$. For the reverse inclusion, let $\omega \in \bigcup_{n=1}^{N} A_n$. Let n_o be the smallest index n such that $\omega \in A_n$. That is, $\omega \in A_{n_o}$, but $\omega \notin A_i$ for $i = 1, 2, \cdots, n_o - 1$. Then $\omega \in B_{n_o} = A_{n_o} \setminus \bigcup_{i=1}^{n_o-1} A_i$ and hence $\omega \in \bigcup_{n=1}^{N} B_n$. This shows the reverse inclusion, namely that $\bigcup_{n=1}^{N} A_n \subseteq \bigcup_{n=1}^{N} B_n$. It now follows from (1.4) that

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{N=1}^{\infty} \bigcup_{n=1}^{N} A_n = \bigcup_{N=1}^{\infty} \bigcup_{n=1}^{N} B_n = \bigcup_{n=1}^{\infty} B_n. \tag{1.5}$$

Observe also that by definition (1.3), the sets B_n are mutually disjoint. Thus,

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \mu\Big(\bigcup_{n=1}^{\infty} B_n\Big) \underset{(M2)}{=} \sum_{n=1}^{\infty} \mu(B_n) \underset{\text{by 1.}}{\leq} \sum_{n=1}^{\infty} \mu(A_n).$$

This proves the theorem.

□

Remark 1.6. Let $A_1, A_2, \cdots, A_N \in \mathcal{F}$ be a finite collection. Setting $A_n := \emptyset$ for $n > N$, we easily obtain "subadditivity",

$$\mu\Big(\bigcup_{n=1}^{N} A_n\Big) = \mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \underset{\text{by 3.}}{\leq} \sum_{n=1}^{\infty} \mu(A_n). = \sum_{n=1}^{N} \mu(A_n).$$

Definition 1.5. A sequence $\{A_n\}_{n=1}^{\infty}$ of sets is called

1. *increasing* if $A_n \subseteq A_{n+1}$ $\forall n$. We write $\{A_n\}\uparrow$.
2. *decreasing* if $A_n \supseteq A_{n+1}$ $\forall n$. We write $\{A_n\}\downarrow$.

Example 1.5. Given a sequence $\{A_i\}_{i=1}^{\infty}$ of sets, set

$$B_n := \bigcup_{i=1}^{n} A_i \qquad \text{and} \qquad C_n := \bigcap_{i=1}^{n} A_i.$$

Then $\{B_n\}\uparrow$ and $\{C_n\}\downarrow$.

Theorem 1.3. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$.*

1. *If $\{A_n\}\uparrow$, then* $\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mu(A_n)$.

2. *If $\{A_n\}\downarrow$ and $\mu(A_1) < \infty$, then* $\mu\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mu(A_n)$.

Proof. For the proof of part 1., set $B_1 := A_1$ and in general, set

$$B_n := A_n \backslash \bigcup_{i=1}^{n-1} A_i = A_n \backslash A_{n-1}$$

inductively as in the proof of part 3. of theorem 1.2. The right identity holds since $\{A_n\}\uparrow$. As the sets B_n are disjoint, we have

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) \underset{(1.5)}{=} \mu\Big(\bigcup_{n=1}^{\infty} B_n\Big) \underset{(\mathrm{M2})}{=} \sum_{n=1}^{\infty} \mu(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu(B_n)$$
$$\underset{(\mathrm{M2'})}{=} \lim_{N\to\infty} \mu\Big(\bigcup_{n=1}^{N} B_n\Big) \underset{(1.4)}{=} \lim_{N\to\infty} \mu\Big(\bigcup_{n=1}^{N} A_n\Big) \underset{\{A_n\}\uparrow}{=} \lim_{N\to\infty} \mu(A_N).$$

As for the proof of part 2., for each n we set

$$E_n := A_1 \backslash A_n.$$

Then $\{E_n\}\uparrow$ as $\{A_n\}\downarrow$, so that by part 1,

$$\mu\Big(\bigcup_{n=1}^{\infty} E_n\Big) = \lim_{N\to\infty} \mu(E_N) = \lim_{N\to\infty} \mu(A_1 \backslash A_N)$$
$$\underset{\text{thm } 1.2}{=} \lim_{N\to\infty} \big[\mu(A_1) - \mu(A_N)\big] = \mu(A_1) - \lim_{N\to\infty} \mu(A_N). \tag{1.6}$$

Observe that all computations are well defined, since by assumption and monotonicity of the measure, $\mu(A_n) < \infty$ for all n. On the other hand,

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} (A_1 \backslash A_n) = A_1 \backslash \Big(\bigcap_{n=1}^{\infty} A_n\Big).$$

As $\mu(A_1) < \infty$, then by part 2 of theorem 1.2,

$$\mu\Big(\bigcup_{n=1}^{\infty} E_n\Big) = \mu(A_1) - \mu\Big(\bigcap_{n=1}^{\infty} A_n\Big). \tag{1.7}$$

Comparing (1.6) and (1.7) we see that

$$\mu\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{N\to\infty} \mu(A_N).$$

This proves the theorem. □

Remark 1.7. In part 2. of the theorem, one can not omit the assumption that $\mu(A_1) < \infty$. (At the very least, one needs that $\mu(A_n) < \infty$ for *some* n.)

For example, consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Let $A_n = [n, \infty)$. Then $\lambda(A_n) = \infty$ for *all* n, and

$$\lambda\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lambda(\emptyset) = 0 \neq \infty = \lim_{n\to\infty} \lambda(A_n).$$

Corollary 1.4. (Borel-Cantellini Theorem). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ a countable family of measurable sets. If*

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty \tag{1.8}$$

then

$$\mu\Big(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\Big) = 0.$$

Proof. For each n, set $E_n := \bigcup_{i=n}^{\infty} A_i$. Then

1. $\{E_n\}\downarrow$, and

2. $\mu(E_n) = \mu\Big(\bigcup_{i=n}^{\infty} A_i\Big) \underset{\text{thm 1.2}}{\leq} \sum_{i=n}^{\infty} \mu(A_i) \leq \sum_{i=1}^{\infty} \mu(A_i) < \infty$ for all n by assumption.

We can thus apply part 2. of theorem 1.3 to the sets $\{E_n\}$ to obtain

$$\begin{aligned}\mu\Big(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\Big) = \mu\Big(\bigcap_{n=1}^{\infty} E_n\Big) &\underset{\text{thm 1.3}}{=} \lim_{n\to\infty} \mu(E_n) \\ &= \lim_{n\to\infty} \mu\Big(\bigcup_{i=n}^{\infty} A_i\Big) \underset{\text{thm 1.2}}{\leq} \lim_{n\to\infty} \sum_{i=n}^{\infty} \mu(A_i) = 0\end{aligned}$$

by assumption (1.8). □

Remark 1.8. One easily checks that

$$\omega \in \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i \quad \Leftrightarrow \quad \omega \in A_i \text{ for infinitely many } i.$$

So Borel-Cantellini's theorem says that if $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then

$$\{\omega \in \Omega : \omega \in A_n \text{ for } \textit{infinitely} \text{ many } n\}$$

has measure zero.

1.3 Measurable Functions

Recall: A mapping $f : X \to Y$ between topological spaces is continuous

$$\underset{\text{definition}}{\Longleftrightarrow} \quad f^{-1}(U) \text{ is open in } X \ \ \forall U \subseteq Y \text{ open}.$$

For mappings between measurable spaces, we have a similar notion:

Definition 1.6. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. A mapping

$$f : \Omega_1 \to \Omega_2$$

is called $(\mathcal{F}_1, \mathcal{F}_2)$-*measurable* (or often simply *measurable*) if

$$f^{-1}(A) \in \mathcal{F}_1 \qquad \forall\, A \in \mathcal{F}_2. \tag{1.9}$$

(that is, if the pre-image of every measurable set is measurable.)

Remark 1.9. If $(\Omega_2, \mathcal{F}_2) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ then we often call such a function $\mathcal{F}_1$-*measurable.*

If $(\Omega_1, \mathcal{F}_1) = \big(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)\big)$ and $(\Omega_2, \mathcal{F}_2) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, then we call f a *Borel function* or a *Borel-measurable function.*

If $(\Omega_1, \mathcal{F}_1) = \big(\mathbb{R}^m, \mathcal{M}_\lambda\big)$ and $(\Omega_2, \mathcal{F}_2) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, then we call f a *Lebesgue-measurable function.* Every Borel function is Lebesgue-measurable.

In general, it suffices to verify (1.9) for generating sets of the σ-algebra $\mathcal{F}_2$:

Theorem 1.5. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces, and $\mathcal{F}_2 = \sigma(\mathcal{K})$, the σ-algebra generated by $\mathcal{K}$. Then*

$$f : \Omega_1 \to \Omega_2 \text{ is measurable} \quad \Leftrightarrow \quad f^{-1}(A) \in \mathcal{F}_1 \qquad \forall\, A \in \mathcal{K}.$$

Proof. $\Rightarrow$: This is obvious as $\mathcal{K} \subseteq \mathcal{F}_2$.
$\Leftarrow$: Suppose, $f^{-1}(A) \in \mathcal{F}_1 \quad \forall\, A \in \mathcal{K}$. Set

$$\mathcal{E} := \{A \in \mathcal{F}_2 : f^{-1}(A) \in \mathcal{F}_1\}.$$

Then obviously, $\mathcal{K} \subseteq \mathcal{E} \subseteq \mathcal{F}_2$. We need to show that $\mathcal{E} = \mathcal{F}_2$. For this we claim that $\mathcal{E}$ is a σ-algebra. In fact,

i) Since $\emptyset \in \mathcal{F}_2$ and $f^{-1}(\emptyset) = \emptyset \in \mathcal{F}_1$, it follows that $\emptyset \in \mathcal{E}$.

ii) Let $A \in \mathcal{E}$. That is, $A \in \mathcal{F}_2$ and $f^{-1}(A) \in \mathcal{F}_1$. Now as $\mathcal{F}_2$ is a σ-algebra, then $A^c \in \mathcal{F}_2$ as well. Also, $f^{-1}(A^c) = \big[f^{-1}(A)\big]^c \in \mathcal{F}_1$ as $f^{-1}(A) \in \mathcal{F}_1$ and $\mathcal{F}_1$ is a σ-algebra. Hence, $A^c \in \mathcal{E}$.

iii) Let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{E}$. Then $A_n \in \mathcal{F}_2$ and $f^{-1}(A_n) \in \mathcal{F}_1$ for each n. Since $\mathcal{F}_2$ is a σ-algebra, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_2$ as well. Also, $f^{-1}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \bigcup_{n=1}^{\infty} f^{-1}(A_n) \in \mathcal{F}_1$ as $f^{-1}(A_n) \in \mathcal{F}_1$ and $\mathcal{F}_1$ is a σ-algebra. Hence, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{E}$.

This proves the claim. Now as $\sigma(\mathcal{K})$ is the smallest σ-algebra containing $\mathcal{K}$ it follows that $\mathcal{F}_2 = \sigma(\mathcal{K}) \subseteq \mathcal{E}$. Hence, $\mathcal{E} = \mathcal{F}_2$ and the theorem is proved. □

Corollary 1.6. *Every continuous function $f : \mathbb{R}^m \to \mathbb{R}^d$ is Borel-measurable.*

Proof. Here we are dealing with the measurable spaces $(\Omega_1, \mathcal{F}_1) = \big(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m)\big)$ and $(\Omega_2, \mathcal{F}_2) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. Now recall that $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{O})$, where

$$\mathcal{O} = \{U \subseteq \mathbb{R}^d : U \text{ is open }\}.$$

Now as f is continuous, $f^{-1}(U)$ is open in $\mathbb{R}^m$ for all $U \subseteq \mathbb{R}^d$ open, and hence, $f^{-1}(U) \in \mathcal{B}(\mathbb{R}^m)$. Thus by theorem 1.5, f is $\big(\mathcal{B}(\mathbb{R}^m), \mathcal{B}(\mathbb{R}^d)\big)$-measurable. □

Corollary 1.7. *Let $(\Omega, \mathcal{F})$ be a measurable space, and $f : \Omega \to \mathbb{R}$ a function. Then the following are equivalent:*

1. *f is $\mathcal{F}$-measurable,*
2. $\{\omega \in \Omega : f(x) < a\} \in \mathcal{F} \quad \forall\, a \in \mathbb{Q}$.
3. $\{\omega \in \Omega : f(x) \le a\} \in \mathcal{F} \quad \forall\, a \in \mathbb{Q}$.
4. $\{\omega \in \Omega : f(x) > a\} \in \mathcal{F} \quad \forall\, a \in \mathbb{Q}$.
5. $\{\omega \in \Omega : f(x) \ge a\} \in \mathcal{F} \quad \forall\, a \in \mathbb{Q}$.

Furthermore, in the above, "$\forall\, a \in \mathbb{Q}$" can be replaced by "$\forall\, a \in \mathbb{R}$".

Proof. Recall that by remark 1.2,

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{K}_1) = \sigma(\mathcal{K}_2) = \sigma(\mathcal{K}_3) = \sigma(\mathcal{K}_4)$$

where

$$\begin{aligned}\mathcal{K}_1 &= \{\ (-\infty, a) : a \in \mathbb{Q}\ \}\\ \mathcal{K}_2 &= \{\ (-\infty, a] : a \in \mathbb{Q}\ \}\\ \mathcal{K}_3 &= \{\ (a, \infty) : a \in \mathbb{Q}\ \}\\ \mathcal{K}_4 &= \{\ [a, \infty) : a \in \mathbb{Q}\ \}.\end{aligned}$$

The corollary is proved using this characterization together with theorem 1.5.

Let us prove 1. ⇔ 2. By theorem 1.5,

$$\begin{aligned} f \text{ is } \mathcal{F}\text{-measurable} &\underset{\text{thm 1.5}}{\iff} f^{-1}(A) \in \mathcal{F} \qquad \forall A \in \mathcal{K}_1\\ &\underset{\text{def. of } \mathcal{K}_1}{\iff} f^{-1}\big((-\infty, a)\big) \in \mathcal{F} \qquad \forall a \in \mathbb{Q}\\ &\iff \{\omega \in \Omega : f(\omega) < a\} \in \mathcal{F} \qquad \forall a \in \mathbb{Q}.\end{aligned}$$

The remaining equivalences are proved similarly. □

The composition of measurable functions is measurable:

Theorem 1.8. *Let $(\Omega, \mathcal{F}_1)$, $(\Omega, \mathcal{F}_2)$ and $(\Omega, \mathcal{F}_3)$ be measurable spaces. If*

$$f : \Omega_1 \to \Omega_2 \qquad \textit{is } (\mathcal{F}_1, \mathcal{F}_2)\textit{-measurable}$$

and

$$g : \Omega_2 \to \Omega_3 \qquad \textit{is } (\mathcal{F}_2, \mathcal{F}_3)\textit{-measurable}$$

then

$$g \circ f : \Omega_1 \to \Omega_3 \qquad \textit{is } (\mathcal{F}_1, \mathcal{F}_3)\textit{-measurable.}$$

Proof. Let $A \in \mathcal{F}_3$ be arbitrary. Since g is $(\mathcal{F}_2, \mathcal{F}_3)$-measurable, then $g^{-1}(A) \in \mathcal{F}_2$. Now as f is $(\mathcal{F}_1, \mathcal{F}_2)$-measurable, then

$$(g \circ f)^{-1}(A) = f^{-1}\big(\underbrace{g^{-1}(A)}_{\in \mathcal{F}2}\big) \in \mathcal{F}_1.$$

This shows that $g \circ f$ is $(\mathcal{F}_1, \mathcal{F}_3)$-measurable. □

Remark 1.10. We will usually make use of this theorem in case $\mathcal{F}_2$ and $\mathcal{F}_3$ are Borel σ-algebras. Then the statement of the theorem becomes:

Let $(\Omega, \mathcal{F})$ be a measurable space. If $f : \Omega \to \mathbb{R}^m$ is $\mathcal{F}$-measurable and $g : \mathbb{R}^m \to \mathbb{R}^d$ is Borel measurable then $g \circ f : \Omega \to \mathbb{R}^d$ is $\mathcal{F}$-measurable.

Often we want to give a set Ω the smallest σ-algebra making a function $f : \Omega \to \mathbb{R}^d$, or each $f_\lambda : \Omega \to \mathbb{R}^d$ in a family $\{f_\lambda\}_{\lambda \in \Lambda}$ measurable:

Definition 1.7. Let Ω be a set, and $\{f_\lambda\}_{\lambda \in \Lambda}$ a family of functions $f_\lambda : \Omega \to \mathbb{R}^d$. Set

$$\mathcal{K} := \{f_\lambda^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^d),\ \lambda \in \Lambda\}.$$

Then $\sigma(\mathcal{K})$ is called the *σ-algebra generated by* $\{f_\lambda\}$, also denoted by $\sigma\big(\{f_\lambda\}\big)$.

Observe that if $\mathcal{F}$ is any σ-algebra on Ω in which the functions f_λ are measurable, then

$$f_\lambda^{-1}(A) \in \mathcal{F} \qquad \forall A \in \mathcal{B}(\mathbb{R}^d),\ \lambda \in \Lambda$$

and hence $\mathcal{K} \subseteq \mathcal{F}$. It follows that

$$\sigma\big(\{f_\lambda\}\big) = \sigma(\mathcal{K}) \subseteq \mathcal{F}.$$

Obviously, every f_λ is $\sigma\big(\{f_\lambda\}\big)$-measurable. So $\sigma\big(\{f_\lambda\}\big)$ is the smallest σ-algebra on Ω making all functions f_λ measurable.

If we have a single generator $f : \Omega \to \mathbb{R}$, then we simply denote $\sigma\big(\{f\}\big)$ by $\sigma(f)$. We have a simple description of this σ-algebra.

Theorem 1.9. $\sigma(f) = \{f^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^d)\}$.

Proof. Set

$$\mathcal{K} := \{f^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^d)\}.$$

Since $\sigma(f) = \sigma(\mathcal{K})$ is the smallest σ algebra containing $\mathcal{K}$, it is enough to show that $\mathcal{K}$ itself is a σ-algebra.

i) As $\emptyset \in \mathcal{B}(\mathbb{R}^d)$ then $\emptyset = f^{-1}(\emptyset) \in \mathcal{K}$.

ii) Let $E \in \mathcal{K}$ be given. Then $E = f^{-1}(A)$ for some $A \in \mathcal{B}(\mathbb{R}^d)$. As $\mathcal{B}(\mathbb{R}^d)$ is a σ-algebra, then $A^c \in \mathcal{B}(\mathbb{R}^d)$, Now $E^c = \big[f^{-1}(A)\big]^c = f^{-1}(A^c)$; hence $E^c \in \mathcal{K}$ as well.

iii) Let $\{E_n\}_{n=1}^{\infty} \subseteq \mathcal{K}$. Then for each n there exists as set $A_n \in \mathcal{B}(\mathbb{R}^d)$ such that $E_n = f^{-1}(A_n)$. As $\mathcal{B}(\mathbb{R}^d)$ is a σ-algebra, then $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{B}(\mathbb{R}^d)$. Now

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} f^{-1}(A_n) = f^{-1}\Big(\bigcup_{n=1}^{\infty} A_n\Big);$$

hence $\bigcup\limits_{n=1}^{\infty} E_n \in \mathcal{K}$ as well.

This shows that $\mathcal{K}$ is a σ-algebra and proves the theorem. □

Definition 1.8. Given functions $f, g : \Omega \to \mathbb{R}$, define

$$\begin{aligned} \max(f,g) : \Omega \to \mathbb{R} \quad &\text{by} \quad \max(f,g)(\omega) = \max\big(f(\omega), g(\omega)\big) \\ \min(f,g) : \Omega \to \mathbb{R} \quad &\text{by} \quad \min(f,g)(\omega) = \min\big(f(\omega), g(\omega)\big) \end{aligned}$$

for every $\omega \in \Omega$. We also set

$$f^+ := \max(f, 0) \qquad \text{and} \qquad f^- = \max(-f, 0).$$

Observe that $f^+, f^- \geq 0$ (by this we mean that $f^+(\omega), f^-(\omega) \geq 0$ for all $\omega \in \Omega$), and

$$f = f^+ - f^- \qquad \text{and} \qquad |f| = f^+ + f^-.$$

Theorem 1.10. *Let $(\Omega, \mathcal{F})$ be a measurable space, and $f, g : \Omega \to \mathbb{R}$ measurable functions. Then*

$$f+g, \quad \alpha f, \quad fg, \quad \frac{1}{g}, \quad \max(f,g), \quad \min(f,g), \quad f^+, \quad f^-, \quad |f|$$

are also measurable. Here, α is any real number. Furthermore, in the case of $\frac{f}{g}$ we require that $g(\omega) \neq 0$ for all ω.

Proof. We omit the proof for αf, fg and $\frac{1}{g}$. This missing part is left as an exercise. (In the case of fg, prove first that f^2 is measurable, and then write $fg = \frac{(f+g)^2-(f-g)^2}{4}$.)

$\boxed{f+g}$: We claim that for each $a \in \mathbb{Q}$,

$$\{\omega \in \Omega : (f+g)(\omega) < a\} = \bigcup_{r \in \mathbb{Q}} \Big[\underbrace{\{\omega \in \Omega : f(\omega) < r\}}_{\in \mathcal{F} \text{ as } f \text{ is measurable}} \cap \underbrace{\{\omega \in \Omega : g(\omega) < a - r\}}_{\in \mathcal{F} \text{ as } g \text{ is measurable}} \Big]. \tag{1.10}$$

In fact, let ω be such that $(f+g)(\omega) < a$. Pick $\epsilon > 0$ such that $f(\omega) + g(\omega) + \epsilon < a$. Now pick $r \in \mathbb{Q}$ such that $f(\omega) < r < f(\omega) + \epsilon$. Then

$$g(\omega) < a - \big[f(\omega) + \epsilon\big] < a - r.$$

That is, $f(\omega) < r$ and $g(\omega) < a - r$ which shows that "$\subseteq$" holds in (1.10).

Conversely, if $f(\omega) < r$ and $g(\omega) < a - r$ for some $r \in \mathbb{Q}$, then obviously, $f(\omega) + g(\omega) < r + (a - r) = a$. This shows that "$\supseteq$" holds in (1.10), and the claim now follows.

Now since the right-hand sets in (1.10) belong to $\mathcal{F}$, it follows by corollary 1.7 that $f + g$ is $\mathcal{F}$-measurable.

$\boxed{\max(f,g),\ \min(f,g)}$: For each $a \in \mathbb{Q}$ we have

$$\{\omega \in \Omega : \max(f,g)(\omega) < a\} = \underbrace{\{\omega \in \Omega : f(\omega) < a\}}_{\in\mathcal{F} \text{ as } f \text{ is measurable}} \cap \underbrace{\{\omega \in \Omega : g(\omega) < a\}}_{\in\mathcal{F} \text{ as } g \text{ is measurable}} \in \mathcal{F}.$$

$$\{\omega \in \Omega : \min(f,g)(\omega) < a\} = \underbrace{\{\omega \in \Omega : f(\omega) < a\}}_{\in\mathcal{F} \text{ as } f \text{ is measurable}} \cup \underbrace{\{\omega \in \Omega : g(\omega) < a\}}_{\in\mathcal{F} \text{ as } g \text{ is measurable}} \in \mathcal{F}.$$

Hence by corollary 1.7, $\max(f,g)$ and $\min(f,g)$ are $\mathcal{F}$-measurable.

$\boxed{f^+, f^-, |f|}$: Measurability of these functions follows from the identities

$$f^+ = \max(f,0), \qquad f^- = \min(f,0), \qquad |f| = f^+ + f^-,$$

together with what has already been proved and the fact that constant functions are always measurable. $\square$

Exercise 1.4. Complete the proof by showing that αf, fg and $\frac{1}{g}$ are $\mathcal{F}$-measurable.

Exercise 1.5. Let $f : \mathbb{R} \to \mathbb{R}$ be increasing (to be precise, non-decreasing). Show that f is a Borel function.

Exercise 1.6. Let $(\Omega, \mathcal{F})$ be a measurable space and $A \subseteq \Omega$. The *characteristic function* χ_A of A is defined by

$$\chi_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Show: $\quad \chi_A$ is $\mathcal{F}$-measurable $\quad \Leftrightarrow \quad A \in \mathcal{F}$.

Theorem 1.11. *Let $(\Omega, \mathcal{F})$ be a measurable space, and $\{f_n\}_{n=1}^{\infty}$ a sequence of $\mathcal{F}$-measurable functions, $f_n : \Omega \to \mathbb{R}$. Then*

$$\sup_n f_n, \qquad \inf_n f_n, \qquad \limsup_n f_n, \qquad \liminf_n f_n, \qquad \lim_{n\to\infty} f_n$$

are all $\mathcal{F}$-measurable, provided they exist and are finite valued.

Proof. We will again make use of corollary 1.7.

$\boxed{\sup_n f_n}$: Recall that $\sup_n f_n$ is defined pointwise by

$$(\sup_n f_n)(\omega) = \sup_n f_n(\omega) \qquad \forall\, \omega \in \Omega.$$

Note also that for every $M \subseteq \mathbb{R}$ and $a \in \mathbb{R}$,

$$\sup M > a \quad \Leftrightarrow \quad \exists\, y \in M \text{ with } y > a. \tag{1.11}$$

Hence for all $a \in \mathbb{Q}$,

$$\begin{aligned}\{\omega \in \Omega : (\sup_n f_n)(\omega) > a\} &= \{\omega \in \Omega : \sup_n f_n(\omega) > a\} \\ &\underset{(1.11)}{=} \bigcup_{n=1}^{\infty} \underbrace{\{\omega \in \Omega : f_n(\omega) > a\}}_{\in \mathcal{F} \text{ as } f_n \text{ is measurable}} \in \mathcal{F}\end{aligned}$$

as $\mathcal{F}$ is a σ-algebra. Hence by corollary 1.7, $\sup_n f_n$ is $\mathcal{F}$-measurable.

$\boxed{\inf_n f_n}$: Arguing as above, we have that for all $a \in \mathbb{Q}$,

$$\{\omega \in \Omega : (\inf_n f_n)(\omega) < a\} = \bigcup_{n=1}^{\infty} \underbrace{\{\omega \in \Omega : f_n(\omega) < a\}}_{\in \mathcal{F} \text{ as } f_n \text{ is measurable}} \in \mathcal{F}.$$

Hence by corollary 1.7, $\inf_n f_n$ is $\mathcal{F}$-measurable.

$\boxed{\limsup_n f_n}$: Recall that if $\{y_n\}$ is a sequence of real numbers, then

$$\limsup_n y_n = \inf_n \big(\sup_{k \geq n} y_k\big). \tag{1.12}$$

Now $\limsup_n f_n$ is defined pointwise, hence for all $\omega \in \Omega$,

$$\begin{aligned}(\limsup_n f_n)(\omega) &= \limsup_n f_n(\omega) \underset{(1.12)}{=} \inf_n \big(\sup_{k \geq n} f_k(\omega)\big) \\ &= \inf_n \big((\sup_{k \geq n} f_k)(\omega)\big) = \big(\inf_n \sup_{k \geq n} f_k\big)(\omega).\end{aligned} \tag{1.13}$$

Now as already proved, $g_n := \sup_{k \geq n} f_k$ is $\mathcal{F}$-measurable for each n and similarly, $\limsup_n f_n \underset{(1.13)}{=} \inf_n \sup_{k \geq n} f_k = \inf_n g_n$ is $\mathcal{F}$-measurable.

$\boxed{\liminf_n f_n}$: In a similar way one shows that

$$\liminf_n f_n = \sup_n \inf_{k \geq n} f_k.$$

Now as already proved, $h_n := \inf_{k \geq n} f_k$ is $\mathcal{F}$-measurable for each n and similarly, $\liminf_n f_n = \sup_n h_n$ is $\mathcal{F}$-measurable.

$\boxed{\lim_{n\to\infty} f_n}$: Recall that if $\{y_n\}$ is a sequence of real numbers and $y \in \mathbb{R}$, then

$$\lim_{n\to\infty} y_n = y \quad \Leftrightarrow \quad \limsup_n y_n = \liminf_n y_n = y.$$

Since $\lim_{n\to\infty} f_n$, $\limsup_n f_n$ and $\liminf_n f_n$ are defined pointwise, one easily deduces that for a function $f : \Omega \to \mathbb{R}$,

$$\lim_{n\to\infty} f_n = f \quad \Leftrightarrow \quad \limsup_n f_n = \liminf_n f_n = f.$$

So if $f_n \to f$, then $f = \limsup_n f_n$; hence f is $\mathcal{F}$-measurable as shown above. □

Remark 1.11. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. One can show that if $f, g : \Omega \to \mathbb{R}^d$ are $\mathcal{F}$-measurable vector valued functions, then so are $f + g$ and αf for each real number α. Similarly, if $\{f_n\}_{n=1}^{\infty}$ is a sequence of $\mathcal{F}$-measurable vector valued functions, $f_n : \Omega \to \mathbb{R}^d$, which converges, say $f_n(\omega) \to f(\omega)$ $\forall \omega \in \Omega$, then the limit function f is also $\mathcal{F}$-measurable. We will not prove these facts, but make use of them freely.

1.4 Almost Everywhere

Definition 1.9. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A set $A \in \mathcal{F}$ is called a *null set*, or *set of measure zero*, if $\mu(A) = 0$.

Example 1.6. Consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.

1. Every *singleton* $\{x\}$ is a null set. In fact,

$$\lambda(\{x\}) = \lambda\Big(\bigcap_{n=1}^{\infty} [x-\frac{1}{n}, x+\frac{1}{n}]\Big) \underset{\text{thm 1.3}}{=} \lim_{n\to\infty} \lambda\Big([x-\frac{1}{n}, x+\frac{1}{n}]\Big) = \lim_{n\to\infty} \frac{2}{n} = 0.$$

2. Every *countable* subset A of $\mathbb{R}$ is a null set. In fact, write the elements of A as a (possibly finite) sequence, $A = \{x_n\}_{n=1}^{\infty}$. Then by σ-additivity,

$$\lambda(A) = \lambda\Big(\bigcup_{n=1}^{\infty} \{x_n\}\Big) \underset{\text{(M2)}}{=} \sum_{n=1}^{\infty} \lambda(\{x_n\}) = \sum_{n=1}^{\infty} 0 = 0.$$

3. There exist *uncountable* subsets of $\mathbb{R}$ which are null sets. (e.g. the Cantor set, see [2].)

Example 1.7. Consider the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ where μ denotes the counting measure. Thus for $A \in \mathcal{F}$ we have

$$A \text{ is a null set} \quad \Leftrightarrow \quad \mu(A) = 0 \quad \Leftrightarrow \quad \text{card}(A) = 0 \quad \Leftrightarrow \quad A = \emptyset.$$

Note that if $A = \{x\}$ is a singleton, then $\mu(A) = 1$.

Remark 1.12. 1. Let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ be a collection of null sets. Then by σ-additivity,

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n-1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} 0 = 0.$$

That is, the countable union of null sets is again a null set.

2. Let $A, B \in \mathcal{F}$ with $B \subseteq A$. If $\mu(A) = 0$, then by monotonicity of the measure, $\mu(B) = 0$. That is measurable subsets of null sets are again null sets.

Definition 1.10. A measure space $(\Omega, \mathcal{F}, \mu)$ is called *complete*, if whenever $A \in \mathcal{F}$ is a null set and $B \subseteq A$, then $B \in \mathcal{F}$. (That is, all subsets of null sets are measurable.)

Example 1.8.
1. Let Ω be any set, $\mathcal{F} = \mathcal{P}(\Omega)$ and μ any measure on $(\Omega, \mathcal{F})$. Then $(\Omega, \mathcal{F}, \mu)$ is complete.
2. One can show (see [2], [3]) that
   (a) $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ is not complete. However,
   (b) $(\mathbb{R}^d, \mathcal{M}_\lambda, \lambda)$ is complete.

Definition 1.11. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $E \in \mathcal{F}$ and (P) a statement about the elements of E. We say that (P) *holds almost everywhere ("a.e.") on* E if there exists a *null set* $A \in \mathcal{F}$ such that

$$B := \{\omega \in E : \text{(P) is not valid at } \omega\} \subseteq A.$$

If $E = \Omega$, then we say that *property (P) holds a.e.*

Remark 1.13. In this definition, the set B itself is not required to be measurable. However, property (P) must hold for all ω outside of the null set A.

In case that $(\Omega, \mathcal{F}, \mu)$ is *complete*, then B will also be measurable and by monotonicity of the measure, $\mu(B) \leq \mu(A) = 0$. That is, B will also be a null set, and the above definition can be restated as

$$\text{(P) holds a.e. on } E \quad \Leftrightarrow \quad \mu\big(\{\omega \in E : \text{(P) does not hold}\}\big) = 0.$$

Example 1.9. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

1. Let $f, g : \Omega \to \mathbb{R}$ be two functions. The statement "$f(x) = g(x)$ a.e." means: There exists a null set $A \in \mathcal{F}$ such that

   $$\{w \in \Omega : f(\omega) \neq g(\omega)\} \subseteq A.$$

   Equivalently, $\omega \notin A$ implies that $f(\omega) = g(\omega)$.

   If $(\Omega, \mathcal{F}, \mu)$ is complete, then

   $$\text{"}f(x) = g(x) \text{ a.e."} \quad \Leftrightarrow \quad \{w \in \Omega : f(\omega) \neq g(\omega)\} \text{ is a null set.}$$

2. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions, $f_n : \Omega \to \mathbb{R}^n$. The statement "$f_n(x)$ converges a.e." means: There exist a function $f : \Omega \to \mathbb{R}^n$ and a null set $A \in \mathcal{F}$ such that

   $$f_n(\omega) \to f(\omega) \qquad \forall\, \omega \notin A.$$

   If $(\Omega, \mathcal{F}, \mu)$ is complete, then

   $$\text{"}f_n(x) \to f \text{ a.e."} \quad \Leftrightarrow \quad \{w \in \Omega : f_n(\omega) \not\to f(\omega)\} \text{ is a null set.}$$

Theorem 1.12. *Let $(\Omega, \mathcal{F}, \mu)$ be a* complete *measure space, and $f, g : \Omega \to \mathbb{R}^d$. Suppose that*

1. *f is $\mathcal{F}$-measurable,*
2. *$f(x) = g(x)$ a.e.*

Then g is also $\mathcal{F}$-measurable.

Proof. Let
$$A := \{\omega \in \Omega : f(\omega) \neq g(\omega)\}.$$
Since $f = g$ a.e. and $(\Omega, \mathcal{F}, \mu)$ is complete, then $A \in \mathcal{F}$ and $\mu(A) = 0$. Now for all $E \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} g^{-1}(E) &= \{\omega \in \Omega : g(\omega) \in E\} \\ &= \big[\{\omega \in \Omega : g(\omega) \in E\} \cap A^c\big] \cup \big[\{\omega \in \Omega : g(\omega) \in E\} \cap A\big] \\ &= \big[\{\omega \in \Omega : f(\omega) \in E\} \cap A^c\big] \cup \big[\underbrace{\{\omega \in \Omega : g(\omega) \in E\} \cap A}_{=:B}\big] \\ &= \big[f^{-1}(E) \cap A^c\big] \cup B \end{aligned}$$

since $g(\omega) = f(\omega)$ on A^c. Now $f^{-1}(E) \in \mathcal{F}$ because f is $\mathcal{F}$-measurable. Also, since $B \subseteq A \in \mathcal{F}$ and $(\Omega, \mathcal{F}, \mu)$ is complete, then $B \in \mathcal{F}$ as well. Thus, $g^{-1}(E) \in \mathcal{F}$. As E was an arbitrary Borel set, it follows that g is $\mathcal{F}$-measurable. □

Remark 1.14. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, not necessarily complete.

1. Let $f : \Omega \to \mathbb{R}^d$ be $\mathcal{F}$-measurable. Given a null set $A \in \mathcal{F}$, let us modify the values of f on A to be constant by setting
$$\tilde{f}(\omega) := \begin{cases} f(\omega) & \text{if } \omega \notin A \\ 0 & \text{if } \omega \in A. \end{cases}$$
(We could replace 0 by any constant of choice.)

   Claim: $\tilde{f}$ is also $\mathcal{F}$-measurable.

   In fact, for all $E \in \mathcal{B}(\mathbb{R}^d)$,
$$\begin{aligned} \tilde{f}^{-1}(E) &= \{\omega \in \Omega : \tilde{f}(\omega) \in E\} \\ &= \big[\{\omega \in \Omega : \tilde{f}(\omega) \in E\} \cap A^c\big] \cup \big[\{\omega \in \Omega : \tilde{f}(\omega) \in E\} \cap A\big] \\ &= \big[\{\omega \in \Omega : f(\omega) \in E\} \cap A^c\big] \cup \{\omega \in A : \tilde{f}(\omega) \in E\} \\ &= \begin{cases} f^{-1}(E) \cap A^c \in \mathcal{F} & \text{if } 0 \notin E \\ \big[f^{-1}(E) \cap A^c\big] \cup A \in \mathcal{F} & \text{if } 0 \in E \end{cases} \end{aligned}$$
as f is $\mathcal{F}$-measurable and $\tilde{f} = f$ on A^c. This proves the claim.

2. Next let $\{f_n\}_{n=1}^{\infty}$ be a sequence of $\mathcal{F}$-measurable functions, $f_n : \Omega \to \mathbb{R}^d$, converging a.e. That is, there exist a null set $A \in \mathcal{F}$ and $f : \Omega \to \mathbb{R}^d$ such that
$$f_n(\omega) \to f(\omega) \qquad \forall \omega \in A^c.$$
For each n, set
$$\tilde{f}_n(\omega) := \begin{cases} f_n(\omega) & \text{if } \omega \notin A \\ 0 & \text{if } \omega \in A \end{cases}$$
and also set
$$\tilde{f}(\omega) := \begin{cases} f(\omega) & \text{if } \omega \notin A \\ 0 & \text{if } \omega \in A. \end{cases}$$
Then each $\tilde{f}_n$ is $\mathcal{F}$-measurable as shown in part 1. and
$$\tilde{f}_n(\omega) \to \tilde{f}(\omega) \qquad \forall \omega \in \Omega.$$
It follows from theorem 1.11 (respectively remark 1.11) that $\tilde{f}$ is also $\mathcal{F}$-measurable. Furthermore, since
$$f_n(\omega) \to \tilde{f}(\omega) \quad \text{a.e.} \quad (\text{that is } \forall \omega \notin A)$$
and $\tilde{f}(\omega) = f(\omega)$ for $\omega \notin A$, we may replace f by $\tilde{f}$, that is, take $\tilde{f}$ as the a.e. limit of $\{f_n\}$. Doing so, we obtain:

If f_n is $\mathcal{F}$-measurable for all n, and $f_n \to f$ a.e., then (by choosing an appropriate f) f is $\mathcal{F}$-measurable.

3. Now suppose that $(\Omega, \mathcal{F}, \mu)$ is *complete*. In this case, since $\tilde{f}$ is $\mathcal{F}$-measurable and $\tilde{f}(\omega) = f(\omega)$ a.e., then by theorem 1.12, f is also $\mathcal{F}$-measurable. We thus obtain:

If f_n is $\mathcal{F}$-measurable for all n, and $f_n \to f$ a.e., then f is $\mathcal{F}$-measurable.

4. Similar arguments apply to $\inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$ and $\limsup_n f_n$ in case of real valued $\mathcal{F}$-measurable functions $f_n : \Omega \to \mathbb{R}$.

1.5 The Lebesgue Integral

Throughout this section, $(\Omega, \mathcal{F}, \mu)$ will denote a fixed measure space. Thus, the words "measurable function" will mean "$\mathcal{F}$-measurable function.". We are now ready to define the Lebesgue integral
$$\int f \, d\mu$$
of a measurable function $f : \Omega \to \mathbb{R}$. Although one usually defines the Lebesgue integral for complex valued functions as well, we will not do so because in probability one deals mostly with real valued functions. The integral is defined in several steps.

Recall from exercise 1.6 that the *characteristic function* of a set $A \subseteq \Omega$ is given by

$$\chi_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

and that χ_A is an $\mathcal{F}$-measurable function iff $A \in \mathcal{F}$.

Definition 1.12. A function of the form

$$\varphi = \sum_{k=1}^{N} c_k \, \chi_{A_k}, \qquad c_k \in \mathbb{R}, \; A_k \subseteq \Omega, \; N \in \mathbb{N} \tag{1.14}$$

is called a *simple function.*

A few observations are in order:

Remark 1.15. 1. Obviously, the range of a simple function is a finite set, as

$$\text{range}(\varphi) \subseteq \left\{ \sum_{k=1}^{N} (c_k)^{m_k}, \quad m_1, \ldots, m_k \in \{0,1\} \right\}$$

and the right-hand set has cardinality $\leq 2^N$. Conversely, if $\varphi : \Omega \to \mathbb{R}$ has finite range, say $\text{range}(\varphi) = \{c_1, \ldots, c_N\}$, then

$$\varphi = \sum_{k=1}^{N} c_k \, \chi_{A_k} \qquad \text{where } A_k = \varphi^{-1}(\{c_k\})$$

and hence φ is a simple function. That is, simple functions are precisely the finite-range functions.

2. A simple function φ may have several representations of form (1.14). If $\text{range}(\varphi) \backslash \{0\} = \{c_1, \ldots, c_N\}$ then we can write

$$\varphi = \sum_{k=1}^{N} c_k \, \chi_{A_k} \qquad \text{where } A_k = \varphi^{-1}(\{c_k\}) \tag{1.15}$$

which is called the *canonical representation* of φ. Obviously, $A_j \cap A_k = \emptyset$ if $j \neq k$, that is, the sets A_k are disjoint. (Sometimes it will be convenient to allow $c_k = 0$ in the canonical representation.)

3. Let φ be a simple function as defined by (1.14). If $A_k \in \mathcal{F}$ for each k, then by exercise 1.6 and theorem 1.10, φ will be $\mathcal{F}$-measurable.

   Conversely, let φ be an $\mathcal{F}$-measurable simple function with canonical representation (1.15). A simple application of corollary 1.7 shows that each set $A_k = \varphi^{-1}(\{c_k\})$ is $\mathcal{F}$-measurable. We thus obtain:

   A simple function φ is measurable $\Leftrightarrow$ each set A_k in its canonical representation (1.15) is measurable.

4. It is an easy exercise to show that

$$V_s := \{\varphi : \Omega \to \mathbb{R} : \varphi \text{ is simple and } \mathcal{F}\text{-measurable}\}$$

is a real vector space.

Example 1.10. Let $(\Omega, \mathcal{F}) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$. Then

$$\varphi = \sum_{k=1}^{N} c_k \chi_{I_k} \qquad (I_k \text{ an interval})$$

is called a *step function.* As intervals are Borel sets, it follows that every step function is Borel-measurable.

The next theorem says that every non-negative measurable function is the a.e. limit of an increasing sequence of simple, measurable functions. It is the crucial ingredient in the definition of the Lebesgue integral.

Theorem 1.13. (Structure Theorem for Measurable Functions) *Let* $f : \Omega \to \mathbb{R}$ *be an* $\mathcal{F}$*-measurable function. There exists a sequence* $\{\varphi_n\}_{n=1}^{\infty}$ *of* simple, $\mathcal{F}$-measurable *functions with*

$$0 \le |\varphi_n(\omega)| \le |f(\omega)| \qquad \forall \omega \in \Omega,\ \forall n$$

such that

$$\varphi_n(\omega) \to f(\omega) \qquad \forall \omega \in \Omega.$$

If $f \ge 0$*, then we may choose* $0 \le \varphi_n(\omega) \le \varphi_{n+1}(\omega)\ \ \forall \omega \in \Omega, \forall n$ *(we write* $\{\varphi_n\}\uparrow$*).*

Proof. 1. Assume first that $f \ge 0$. Set

$$\varphi_1(\omega) = \begin{cases} 0 & \text{if } 0 \le f(\omega) < \frac{1}{2} \\ \frac{1}{2} & \text{if } \frac{1}{2} \le f(\omega) < 1 \\ 1 & \text{if } f(\omega) \ge 1. \end{cases}$$

and for general n, set

$$\varphi_n(\omega) = \begin{cases} \frac{i-1}{2^n} & \text{if } f(\omega) < n \text{ and } \frac{i-1}{2^n} \le f(\omega) < \frac{i}{2^n} \text{ for some } 1 \le i \le n\,2^n \\ n & \text{if } f(\omega) \ge n. \end{cases} \tag{1.16}$$

Note that

$$\varphi_n = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{A_i} + n\chi_{A_0}$$

where we have set

$$A_i = A_i^{(n)} := \Big\{\omega \in \Omega : \frac{i-1}{2^n} \le f(\omega) < \frac{i}{2^n}\Big\} \qquad (1 \le i \le n2^n), \quad \text{and}$$
$$A_0 = A_0^{(n)} := \big\{\omega \in \Omega : f(\omega) \ge n\big\}.$$

Since f is measurable, then $A_i \in \mathcal{F}$ for all i, $0 \leq i \leq n2^n$. Hence, each φ_n is an $\mathcal{F}$-measurable simple function, and $0 \leq \varphi_n \leq f$ by definition.

We now show that $\{\varphi_n\}\uparrow$. In fact let n be fixed, and $\omega \in \Omega$.
a) Suppose, $0 \leq f(\omega) < n$. Then there exists i, $1 \leq i \leq n2^n$, so that

$$\frac{i-1}{2^n} \leq f(\omega) < \frac{i}{2^n},$$

and $\varphi_n(\omega) = \frac{i-1}{2^n}$. Now compute $\varphi_{n+1}(\omega)$. The above inequality gives

$$\frac{2i-2}{2^{n+1}} \leq f(\omega) < \frac{2i}{2^{n+1}}.$$

If $\frac{2i-2}{2^{n+1}} \leq f(\omega) < \frac{2i-1}{2^{n+1}}$, then by (1.16),

$$\varphi_{n+1}(\omega) = \frac{2i-2}{2^{n+1}} = \varphi_n(\omega),$$

while if $\frac{2i-1}{2^{n+1}} \leq f(\omega) < \frac{2i}{2^{n+1}}$, then

$$\varphi_{n+1}(\omega) = \frac{2i-1}{2^{n+1}} \geq \frac{2i-2}{2^{n+1}} = \varphi_n(\omega).$$

b) Suppose, $n \leq f(\omega) < n+1$. Then $\varphi_n(\omega) = n$, and

$$\varphi_{n+1}(\omega) = \frac{i-1}{2^{n+1}}$$

where i is the unique positive integer satisfying $\frac{i-1}{2^{n+1}} \leq f(\omega) < \frac{i}{2^{n+1}}$ Note that $\frac{i-1}{2^{n+1}} \geq n$ since $f(\omega) \geq n$. Hence,

$$\varphi_{n+1}(\omega) = \frac{i-1}{2^{n+1}} \geq n\varphi_n(\omega).$$

c) Finally, suppose that $f(\omega) \geq n+1$. then

$$\varphi_{n+1}(\omega) = n+1 > n = \varphi_n(\omega).$$

This shows that $\varphi_{n+1}(\omega) \geq \varphi_n(\omega)$ for all ω.

Next we show that $\varphi_n \to f$. Let $\omega \in \Omega$ be given. By construction (1.16) of φ_n, we have

$$0 \leq f(\omega) - \varphi_n(\omega) \leq \frac{1}{2^n} \to 0$$

as $n \to \infty$. Hence, $\varphi_n(\omega) \to f(\omega)$.

2. Next let f be arbitrary, $\mathcal{F}$-measurable. Then f^+ and f^- are $\mathcal{F}$-measurable. Let $\{\varphi_n\}\uparrow$, $\{\psi_n\}\uparrow$ be sequences of simple, $\mathcal{F}$-measurable functions constructed in 1., with

$$0 \leq \varphi_n \leq f^+, \qquad 0 \leq \psi_n \leq f^-$$

for all n, and

$$\varphi_n \to f^+, \quad \psi_n \to f^-.$$

Then $\{\varphi_n - \psi_n\}_{n=1}^{\infty}$ is a sequence of $\mathcal{F}$-measurable simple functions satisfying

1. $|\varphi_n - \psi_n| \leq |\varphi_n| + |\psi_n| = \varphi_n + \psi_n \leq f^+ + f^- = |f|$, and
2. $(\varphi_n - \psi_n)(\omega) = \varphi_n(\omega) - \psi_n(\omega) \to f^+(x) - f^-(x) = f(\omega)$ for all ω.

This proves the theorem. □

Definition 1.13. (Lebesgue integral of simple, non-negative functions). Let

$$\varphi = \sum_{k=1}^{N} c_k \, \chi_{A_k} \tag{1.17}$$

be an $\mathcal{F}$-measurable, non-negative (i.e. $c_k > 0 \ \forall k$) simple function in *canonical form.* We define its integral by

$$\int \varphi \, d\mu := \sum_{k=1}^{N} c_k \, \mu(A_k). \tag{1.18}$$

Remark 1.16. 1. It may be possible that $\mu(A_k) = \infty$ for some k. In this case, as $c_k > 0$, we can make the convention that $c_k \cdot \infty = \infty$, so that $\int \varphi \, d\mu = \infty$. In order to also permit $c_k = 0$ in the canonical representation (1.17) and hence in the definition of the integral (1.18), we make the convention that $0 \cdot \infty = 0$.

2. Suppose we have a representation of φ of the form

$$\varphi = \sum_{j=1}^{M} a_j \, \chi_{B_j} \qquad a_j \geq 0, \ B_j \in \mathcal{F} \ \ \forall j \tag{1.19}$$

with the sets B_j mutually disjoint. Then for each j, $a_j \in \text{range}(\varphi)$, that is, $a_j = c_{k(j)}$ for some $k = k(j)$ and hence $B_j \subseteq A_{k(j)}$. We can thus rearrange (1.19) as

$$\varphi = \sum_{k=1}^{N} \sum_{\{j:k(j)=k\}} c_k \, \chi_{B_j}.$$

Observe that

$$A_k = \varphi^{-1}(\{c_k\}) = \bigcup_{\{j:k(j)=k\}} B_j$$

and hence

$$\begin{aligned}
\sum_{j=1}^{M} a_j \, \mu(B_j) &= \sum_{k=1}^{N} \sum_{\{j:k(j)=k\}} a_j \, \mu(B_j) = \sum_{k=1}^{N} \sum_{\{j:k(j)=k\}} c_k \, \mu(B_j) \\
&= \sum_{k=1}^{N} c_k \sum_{\{j:k(j)=k\}} \mu(B_j) = \sum_{k=1}^{N} c_k \, \mu\Big(\bigcup_{\{j:k(j)=k\}} B_j \Big) \\
&= \sum_{k=1}^{N} c_k \, \mu(A_k) = \int \varphi \, d\mu.
\end{aligned}$$

That is, in the definition of the integral (1.18) we don't require that (1.17) be the canonical representation of φ, but only that the sets A_k are disjoint.

3. Using 2. above, it is not difficult to show:

   Let φ, ψ be simple, non-negative $\mathcal{F}$-measurable functions and $\alpha > 0$. Then

   (a) $\int(\varphi+\psi)\,d\mu = \int\varphi\,d\mu + \int\psi\,d\mu$ *("additivity")*

   (b) $\int\alpha\varphi\,d\mu = \alpha\int\varphi\,d\mu$ *("positive homogeneity")*

   (c) *If* $\varphi \leq \psi$ *then* $\int\varphi\,d\mu \leq \int\psi\,d\mu$ *("monotonicity")*

   The proof is left as an exercise. (The idea is to choose representations of φ and ψ with identical sets A_k.)

4. Let

$$\varphi = \sum_{j=1}^{M} a_j\,\chi_{B_j} \qquad a_j \geq 0,\ B_j \in \mathcal{F}\ \ \forall j$$

   be *any* representation of. φ. Then by part 3. and definition of the integral of a characteristic function,

$$\int \varphi\,d\mu = \sum_{j=1}^{\infty} a_j \int \chi_{B_j}\,d\mu = \sum_{j=1}^{\infty} a_j\,\mu(B_j).$$

   That is, in the definition of the integral (1.18) we may even allow the sets A_k to overlap.

Example 1.11. Consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. If φ is a step function,

$$\varphi = \sum_{k=1}^{N} c_k \chi_{I_k}$$

where $c_k \geq 0$ and each I_k is an interval with endpoints a_k and b_k, $a_k \leq b_k$, then

$$\int \varphi\,d\lambda = \sum_{k=1}^{N} c_k\,\lambda(I_k) = \sum_{k=1}^{N} c_k(b_k - a_k).$$

Definition 1.14. (Lebesgue integral of non-negative functions). Let $f : \Omega \to [0, \infty)$ be $\mathcal{F}$-measurable. Then by theorem 1.3, there exists an increasing sequence $\{\varphi_n\}\uparrow$ of simple, non-negative $\mathcal{F}$-measurable functions such that

$$\varphi_n(\omega) \to f(\omega) \qquad \forall\omega \in \Omega.$$

By monotonicity of the integral, then

$$\left\{\int \varphi_n\,d\mu\right\}_{n=1}^{\infty}$$

is an increasing sequence in $[0, \infty)$ which either converges, or diverges to infinity. We can thus set

$$\int f\,d\mu := \lim_{n\to\infty} \int \varphi_n\,d\mu. \tag{1.20}$$

Remark 1.17. 1. Obviously by definition, $\int f\,d\mu \in [0,\infty]$.

2. One needs to show that this integral is *well defined*. By this we mean that the limit in (1.20) is independent of the sequence $\{\varphi_n\}$ chosen. So one must show that if $\{\psi_n\}\uparrow$ is another increasing sequence of simple, non-negative $\mathcal{F}$-measurable functions such that $\psi_n(\omega)\to f(\omega)$, then

$$\lim_{n\to\infty}\int \psi_n\,d\mu = \lim_{n\to\infty}\int \varphi_n\,d\mu.$$

   This is left as an exercise.

3. It is also left as an exercise to prove:
   Let f, g be non-negative $\mathcal{F}$-measurable functions and $\alpha > 0$. Then

   (a) $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$ *("additivity")*

   (b) $\int \alpha f\,d\mu = \alpha \int f\,d\mu$ *("positive homogeneity")*

   (c) If $f \le g$ *then* $\int f\,d\mu \le \int g\,d\mu$ *("monotonicity")*

Definition 1.15. (Lebesgue integral of arbitrary functions). Let $f:\Omega\to\mathbb{R}$ be $\mathcal{F}$-measurable. Then by theorem 1.10, $f^+, f^-:\Omega\to[0,\infty)$ are also $\mathcal{F}$-measurable. Since $f = f^+ - f^-$ we define

$$\int f\,d\mu := \int f^+\,d\mu - \int f^-\,d\mu \tag{1.21}$$

provided that the right-hand side is not of the form $\infty - \infty$.

We say that f *is integrable* if $\int f^+$ and $\int f^-$ are both finite (i.e. $\neq\infty$). Thus, f is integrable iff $\int f\,d\mu$ is defined and is finite.

Remark 1.18. Let

$$\mathcal{L}^1 := \{f:\Omega\to\mathbb{R} : f \text{ is } \mathcal{F}\text{-measurable and integrable}\}.$$

One can show:

1. *$\mathcal{L}^1$ is a vector space, and $f \mapsto \int f\,d\mu$ is a linear functional on $\mathcal{L}^1$. That is, if $f, g \in \mathcal{L}^1$ and $\alpha, \beta \in \mathbb{R}$, then*

   (a) $\alpha f + \beta g \in \mathcal{L}^1$ *and*

   (b) $\displaystyle\int (\alpha f + \beta g)\,d\mu = \alpha\int f\,d\mu + \beta\int g\,d\mu.$

2. *The integral is monotone: If $f, g \in \mathcal{L}^1$ with $f \le g$, then*

$$\int f\,d\mu \le \int g\,d\mu.$$

   The proof of 1. and 2. is left as an exercise.

3. Since
$$0 \le f^+, f^- \le f^+ + f^- = |f|$$
we have by linearity and monotonicity of the integral that
$$\int |f|\,d\mu < \infty \quad \Leftrightarrow \quad \int f^+\,d\mu < \infty \quad \text{and} \quad \int f^-\,d\mu < \infty.$$
Thus by definition of "integrable",
$$\int |f|\,d\mu < \infty \quad \Leftrightarrow \quad f \text{ is integrable.}$$
In particular, if $\mu(\Omega) < \infty$ and f is bounded, say $|f(\omega)| \le M$ for all ω, then by monotonicity of the integral,
$$\int |f|\,d\mu \le \int M\,d\mu = M \cdot \mu(\Omega) < \infty.$$
That is, f is integrable.

4. Let f be $\mathcal{F}$-measurable and $g \in \mathcal{L}^1$ with $|f| \le g$. Then by monotonicity of the integral,
$$\int |f|\,d\mu \le \int g\,d\mu < \infty.$$
That is, every bounded and measurable function f on a finite measure space is integrable.

5. Let f be integrable. Since
$$-|f| \le f \le |f|$$
then by linearity and monotonicity of the integral,
$$-\int |f|\,d\mu = \int -|f|\,d\mu \le \int f\,d\mu \le \int |f|\,d\mu.$$
From $-a \le b \le a \Leftrightarrow |b| \le a$, for $a > 0, b \in \mathbb{R}$ we obtain
$$\left|\int f\,d\mu\right| \le \int |f|\,d\mu.$$

6. One uses various notations for the integral. For example
$$\int_\Omega f\,d\mu \qquad \int_\Omega f \qquad \int f \qquad \int_\Omega f(\omega)\,d\mu(\omega) \qquad \int_\Omega f(\omega)\,d\omega$$
all denote $\int f\,d\mu$.

Exercise 1.7. Let $f, g : \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable. Show:
If $f(\omega) = g(\omega)$ a.e., then
$$\int f\,d\mu \text{ is defined} \Leftrightarrow \int g\,d\mu \text{ is defined.}$$
Furthermore, if these integrals are defined, then they are equal.

1.6 Convergence Theorems

The following three convergence theorems constitute one of the strengths of the Lebesgue integral and we will need to use them as tools throughout. We will omit their proofs which can be found in [2] or [3], for example. Throughout, $(\Omega, \mathcal{F}, \mu)$ will be a measure space, and $\{f_n\}_{n=1}^{\infty}$ will denote a sequence of $\mathcal{F}$-measurable functions, $f_n : \Omega \to \mathbb{R}$. The theorems answer the question: Is $\int \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int f_n$?

The first theorem deals with increasing sequences of functions. Thus, for each $\omega \in \Omega$ the sequence $\{f_n(\omega)\}$ either converges, or diverges to ∞. The implicit assumption here is that $\{f_n(\omega)\}$ converges a.e. $(\omega \in \Omega)$. Observe that the functions here need not be integrable, i.e. some of the integrals below may be infinite.

Theorem 1.14. (Monotone Convergence Theorem, MCT). *If $\{f_n\}\uparrow$ and $f_n \geq 0$ for all n, and $f_n(\omega) \to f(\omega)$ a.e., then*

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

In the second theorem, the sequence $\{f_n\}$ need not converge. Since all functions involved are non-negative, some of the integrals may again be infinite. This theorem is often used to show that $\int f\, d\mu$ is finite.

Theorem 1.15. (Fatou's Lemma). *If $f_n \geq 0$ for all n, and $f = \liminf_n f_n$ a.e., then*

$$\int f\, d\mu \leq \liminf_n \int f_n\, d\mu.$$

In the third theorem, the functions involved may take negative values. This requires the condition that all functions be dominated by an integrable function.

Theorem 1.16. (Lebesgue Dominated Convergence Theorem, LDCT). *Suppose*

1. $f_n(\omega) \to f(\omega)$ *a.e.*
2. *There exists $g \in \mathcal{L}^1$ such that $|f_n(\omega)| \leq g(\omega)$ a.e., for all n.*

Then $f \in \mathcal{L}^1$ (i.e. f is integrable) and

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

Exercise 1.8. Consider the measure space $(\mathbb{R}, \mathcal{M}_\lambda, \lambda)$.

1. Let

$$f(x) = \begin{cases} \frac{1}{n} & x \in [n, n+1) \\ 0 & \text{else} \end{cases} \quad \text{and} \quad g(x) = \begin{cases} \frac{(-1)^n}{n} & x \in [n, n+1) \\ 0 & \text{else.} \end{cases}$$

Show that $\int f\, d\lambda = \infty$ while $\int g\, d\lambda$ is undefined. (So both functions are not integrable, but for different reasons.)

2. Let
$$h(x) = \begin{cases} (-1)^n & \text{if } x \in [n, n+\frac{1}{n^3}) \\ 0 & \text{else.} \end{cases}$$
Show that h is unbounded, but integrable.

Next we define the integral over a measurable subset of Ω.

Definition 1.16. (Lebesgue integral over a measurable set). Let $A \in \mathcal{F}$. We define
$$\int_A f\, d\mu := \int f\chi_A\, d\mu$$
provided that the right-hand integral is defined.

Remark 1.19. 1. If $f \geq 0$ and $A, B \in \mathcal{F}$ with $A \subseteq B$, then as $\chi_A \leq \chi_B$ we have by monotonicity of the integral,
$$\int_A f\, d\mu = \int f\chi_A\, d\mu \leq \int f\chi_B\, d\mu = \int_B f\, d\mu.$$

2. If $f \in \mathcal{L}^1$ and $\{A_k\}_{k=1}^{\infty}$ is a *disjoint* collection of subsets of $\mathcal{F}$, let us set $B_n := \bigcup_{k=1}^{n} A_k$ and $B := \bigcup_{k=1}^{\infty} A_k$. Now as
$$(f\chi_{B_n})(\omega) \to (f\chi_B)(\omega) \quad \forall w \qquad \text{and} \qquad |f|\chi_{B_n} \leq |f|\chi_B \leq |f| \in \mathcal{L}^1$$
we have by linearity of the integral and the LDCT,
$$\begin{aligned} \int_B f\, d\mu &= \int f\chi_B\, d\mu = \int \lim_{n\to\infty} f\chi_{B_n}\, d\mu = \lim_{n\to\infty} \int f\chi_{B_n}\, d\mu \\ &= \lim_{n\to\infty} \int f\Big(\sum_{k=1}^{n} \chi_{A_k}\Big)\, d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} \int f\chi_{A_k}\, d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} \int_{A_k} f\, d\mu \end{aligned}$$
That is,
$$\int_{\bigcup_{k=1}^{\infty} A_k} f\, d\mu = \sum_{k=1}^{\infty} \int_{A_k} f\, d\mu.$$

3. Using remark 1.18 one easily checks that $f \mapsto \int_A f\, d\mu$ is also linear and monotone.

Exercise 1.9. One can prove the following theorem:

If $f : [a,b] \to \mathbb{R}$ *is* Riemann integrable, *then* f *is Lebesgue-measurable and Lebesgue-integrable on* $[a,b]$*, and Riemann and Lebesgue integrals coincide,*
$$R - \int_a^b f(x)\, dx = \int_{[a,b]} f\, d\lambda.$$
(Here the left integral denotes the Riemann integral).

This is no longer true for *improper* Riemann integrals. In fact, show:

1. $R - \int_0^\infty \frac{\sin x}{x}\,dx$ converges, but $\int\limits_{[0,\infty)} \frac{\sin x}{x}\,d\lambda$ is undefined.

2. If $f \geq 0$ and f is Riemann integrable on each bounded interval $[0,t]$, $t > 0$, and if the improper Riemann integral
$$R - \int_0^\infty f(x)\,dx$$
converges, then f is Lebesgue integrable on $[0,\infty)$ and
$$R - \int_0^\infty f(x)\,dx = \int\limits_{[0,\infty)} f\,d\lambda.$$

Recall the Fundamental Theorem of Calculus: Let $f : [a,b] \to \mathbb{R}$ be Riemann integrable. Fix $x_o \in (a,b)$ and set
$$F(x) = \int_{x_o}^x f(t)\,dt \qquad (x \in [a,b]).$$
If f is continuous at x_o, then F is differentiable at x_o and $f(x_o) = F'(x_o)$. That is,
$$f(x_o) = \lim_{h\to 0} \frac{F(x_o+h) - F(x_o)}{h} = \lim_{h\to 0} \frac{1}{h} \int_{x_o}^{x_o+h} f(t)\,dt.$$
Given $\epsilon > 0$, setting $h = \epsilon$ and $h = -\epsilon$, respectively, we obtain
$$f(x_o) = \lim_{\epsilon\to 0^+} \frac{1}{\epsilon} \int_{x_o}^{x_o+\epsilon} f(t)\,dt$$
and
$$f(x_o) = \lim_{\epsilon\to 0^+} \frac{1}{-\epsilon} \int_{x_o}^{x_o-\epsilon} f(t)\,dt = \lim_{\epsilon\to 0^+} \frac{1}{\epsilon} \int_{x_o-\epsilon}^{x_o-} f(t)\,dt.$$
Averaging both equations,
$$f(x_o) = \frac{1}{2} \lim_{\epsilon\to 0^+} \frac{1}{\epsilon} \int_{x_o-e}^{x_o+\epsilon} f(t)\,dt = \lim_{\epsilon\to 0^+} \frac{1}{\lambda(x_o-\epsilon, x_o+\epsilon)} \int_{x_o-e}^{x_o+\epsilon} f(t)\,dt$$
where $\lambda(x_0-\epsilon, x_o+\epsilon)$ denotes the length(=Lebesgue measure) of this interval.

This generalizes to the Lebesgue integral for the proof see [3]:

Theorem 1.17. (Lebesgue Theorem). *Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lebesgue integrable. Then for a.e. $x \in \mathbb{R}^n$,*
$$f(x) = \lim_{\epsilon\to 0^+} \frac{1}{\lambda\big(B(x,\epsilon)\big)} \int_{B(x,\epsilon)} f\,d\lambda \tag{1.22}$$
where $B(x,\epsilon)$ denotes the open (or closed) ball with center x and radius ϵ. (In case $d = 1$ we may replace the open balls by half-intervals $(x-\epsilon,\epsilon]$, respectively $[x, x+\epsilon$.)

.

1.7 L^p-spaces

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and p be fixed, $1 \leq p < \infty$. Set

$$L^p = L^p(\Omega) = L^p(\Omega, \mathcal{F}, \mu) := \{f : \Omega \to \mathbb{R} : f \text{ is } \mathcal{F}\text{-measurable and } \int |f|^p \, d\mu < \infty\}.$$

(To be precise, we identify functions which are equal a.e. Thus, elements of $L^p(\Omega)$ are equivalence classes of functions, but for simplicity, we identify each equivalence class with a member function. As a consequence, any property of a function depending on the elements ω of Ω can hold only a.e. on Ω. However, this identification does not pose a problem for the integral; exercise 1.7 shows that functions in the same equivalence class have identical integrals.) Observe that $f \in L^p \Leftrightarrow |f|^p \in L^1$.

The proofs of the following two theorems can be found in any textbook on measure theory, such as [2] or [3], for example.

Theorem 1.18. *$L^p(\Omega)$ is a Banach space with norm*

$$\|f\|_p := \Big[\int |f|^p \, d\mu\Big]^{1/p}.$$

In particular, for all $f, g \in L^p$,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \qquad \text{("Minkowski's inequality")}$$

Theorem 1.19. *Let $1 < p < \infty$ and let q be such that $\frac{1}{p} + \frac{1}{q} = 1$. ($q$="conjugate of p", $1 < q < \infty$.) Then for all $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$, we have*

1. *fg is integrable (i.e. $fg \in L^1(\Omega)$)*

2. $\displaystyle\int |fg| \, d\mu \leq \|f\|_p \|g\|_q.$ *("Hölder's inequality")*

Remark 1.20. If $p = 2$ then $q = 2$ as well, and one easily verifies that

$$< f, g > := \int fg \, d\mu$$

defines an inner product on $L^2(\Omega)$ making it a *real Hilbert space*. (For this reason, Hölders's inequality is also called the "Cauchy-Schwarz inequality" in case $p = 2$.)

The Lebesgue Dominated Convergence Theorem has a version for the L^p-spaces:

Theorem 1.20. *Let $1 \leq p < \infty$ and let $\{f_n\}_{n=1}^{\infty}$ a sequence of $\mathcal{F}$-measurable functions such that*

1. *$f_n(\omega) \to f(\omega)$ a.e.*

2. *There exists $g \in L^p(\Omega)$ such that $|f_n(\omega)| \leq |g(\omega)|$ a.e.*

Then $f_n, f \in L^p(\Omega)$ and $f_n \xrightarrow{\|\cdot\|_p} f$.

Proof. By assumption,

$$|f_n(\omega)|^p \to |f(\omega)|^p \text{ a.e.} \qquad \text{and} \qquad |f_n(\omega)|^p \leq |g(\omega)|^p \text{ a.e.}$$

As $|g|^p \in L^1(\Omega)$, then by the LDCT,

$$|f_n|^p, |f|^p \in L^1(\Omega) \qquad \text{that is} \qquad |f_n|, |f| \in L^p(\Omega).$$

Now since $L^p(\Omega)$ is a vector space then $|g| + |f| \in L^p(\Omega)$ which is equivalent to $[|g| + |f|]^p \in L^1(\Omega)$. From

$$|f_n(\omega) - f(\omega)|^p \to 0 \text{ a.e.}$$

and

$$|f_n(\omega) - f(\omega)|^p \leq [|f_n(\omega)| + |f(\omega)|]^p \leq [|g(\omega)| + |f(\omega)|]^p \text{ a.e.}$$

it follows by the LDCT that

$$\lim_{n\to\infty} \|f_n - f\|_p^p = \lim_{n\to\infty} \int |f_n - f|^p \, d\mu \underset{\text{LDCT}}{=} \int 0 \, d\mu = 0$$

which proves the theorem. □

Integrals over "small" sets are small:

Theorem 1.21. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $g \in L^1(\Omega)$ be fixed. Then for each $\epsilon > 0$ there exists $\delta > 0$ so that*

$$\int_E |g| \, d\mu < \epsilon \qquad \forall E \in \mathcal{F} \text{ with } \mu(E) < \delta.$$

Proof. Let $\epsilon > 0$ be given.
1. Assume first that g is bounded, say $|g(\omega)| \leq M$ for all $\omega \in \Omega$, where $M > 0$. Set $\delta = \frac{\epsilon}{M}$. Then for all measurable sets E with $\mu(E) < \delta$ we have

$$\int_E |g| \, d\mu = \int_\Omega |g|\chi_E \, d\mu \leq \int_\Omega M\chi_E \, d\mu = M\mu(E) < M \cdot \frac{\epsilon}{M} = \epsilon.$$

2. Next let g be arbitrary. For each $n \in \mathbb{N}$, set

$$A_n := \{\omega \in \Omega : |g(\omega)| \leq n\}.$$

Then $A_n \in \mathcal{F}$. So if we set $g_n := g\chi_{A_n}$, then g_n is an $\mathcal{F}$-measurable function. Furthermore for all ω,

1. $|g_n(\omega)| \leq |g(\omega)| \in L^1(\Omega)$, and
2. $g_n(\omega) \to g(\omega)$.

Hence by theorem 1.20,

$$\| g - g_n \|_1 \to 0 \qquad \text{as } n \to \infty.$$

In particular, there exists n such that

$$\| g - g_n \|_1 < \frac{\epsilon}{2}.$$

Applying part 1. to g_n, there exists $\delta > 0$ such that

$$\int_E |g_n| \, d\mu < \frac{\epsilon}{2} \qquad \text{whenever } \mu(E) < \delta.$$

Thus for all measurable sets E with $\mu(E) < \delta$ we have

$$\begin{aligned}\int_E |g| \, d\mu &\leq \int_E |g - g_n| \, d\mu + \int_E |g_n| \, d\mu \\ &< \int_\Omega |g - g_n| \, d\mu + \frac{\epsilon}{2} = \|g - g_n\|_1 + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon\end{aligned}$$

which proves the theorem. □

Theorem 1.22. (Chebychev inequality). *Let $1 \leq p < \infty$ and $f \in L^p(\Omega)$. Then for each $M > 0$,*

$$\mu\big(\{\omega \in \Omega : |f(\omega)| \geq M\} \big) \leq M^{-p} \|f\|_p^p. \tag{1.23}$$

Proof. Set

$$E_M := \{\omega \in \Omega : |f(\omega)| \geq M\}.$$

Then by monotonicity of the integral,

$$\|f\|_p^p = \int_\Omega |f|^p \, d\mu \geq \int_{E_M} |f|^p \, d\mu \geq \int_{E_M} M^p \, d\mu = M^p \int_\Omega \chi_{E_M} \, d\mu = M^p \cdot \mu(E_M).$$

Dividing by $M^p > 0$ we obtain the assertion (1.23). □

Exercise 1.10. Let $(\Omega, \mathcal{F}, \mu)$ be a *finite* measure space, and $1 \leq p < q < \infty$. Show:

1. $L^q(\Omega) \subseteq L^p(\Omega)$ and

$$\|f\|_p \leq \mu(\Omega)^{\frac{1}{p} - \frac{1}{q}} \|f\|_q \qquad \forall f \in L^q(\Omega).$$

   (So in particular, if $\mu(\Omega) = 1$ then $\|f\|_p \leq \|f\|_q$.)

2. $L^p(\mathbb{R}) \not\subseteq L^q(\mathbb{R})$ and $L^q(\mathbb{R}) \not\subseteq L^p(\mathbb{R})$.
   (In particular, 1. need not hold if $\mu(\Omega) = \infty$.)

1.8 Modes of Convergence

Throughout this section, fix a measure space $(\Omega, \mathcal{F}, \mu)$. Let $\{f_n\}$ be a sequence of $\mathcal{F}$-measurable functions, $f_n : \Omega \to \mathbb{R}$, and let $f : \Omega \to \mathbb{R}$. We are already familiar with the following *modes of convergence:*

1. *pointwise convergence*:

$$f_n \longrightarrow f \quad \Leftrightarrow \quad \lim_{n\to\infty} f_n(\omega) = f(\omega) \quad \forall w \in \Omega.$$

   In this case, f is also $\mathcal{F}$measurable by theorem 1.11

2. *a.e. convergence*:

$$f_n \longrightarrow f \ \text{ a.e.} \quad \Leftrightarrow \quad \exists A \in \mathcal{F},\ \mu(A) = 0 \ \text{ s.t. } \lim_{n\to\infty} f_n(\omega) = f(\omega) \quad \forall w \in A^c.$$

   In this case, f is equal a.e. to a measurable function (so we may assume f to be measurable) by remark 1.14.

3. *uniform convergence*:

$$f_n \Longrightarrow f \quad \Leftrightarrow \quad \lim_{n\to\infty} \sup_{\omega\in\Omega} |f_n(\omega) - f(\omega)| = 0.$$

   Obviously, 3. $\Rightarrow$ 1. $\Rightarrow$ 2.

4. *convergence in the p-th mean*: Suppose, $f_n, f \in L^p(\Omega)$. Then

$$f_n \overset{\|\cdot\|_p}{\longrightarrow} f \quad \Leftrightarrow \quad \lim_{n\to\infty} \|f_n - f\|_p = 0.$$

Definition 1.17. Suppose, f is also $\mathcal{F}$-measurable. We say that $\{f_n\}$ *converges to f in measure* and write

$$f_n \overset{\text{meas}}{\longrightarrow} f$$

if for each $\epsilon > 0$,

$$\lim_{n\to\infty} \mu\big(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\}\big) = 0.$$

That is,

$$f_n \overset{\text{meas}}{\longrightarrow} f \quad \Leftrightarrow \quad \lim_{n\to\infty} \mu\big(E_{n,\epsilon}\big) = 0 \quad \forall \epsilon > 0$$

where $E_{n,\epsilon} = \{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\}$.

Exercise 1.11. Show:

1. We may replace "$|f_n(\omega) - f(\omega)| > \epsilon$" by "$|f_n(\omega) - f(\omega)| \geq \epsilon$" in the above definition.

2. The limit in measure is essentially unique: If f, g are $\mathcal{F}$-measurable and
$$f_n \xrightarrow{\text{meas}} f \qquad \text{and} \qquad f_n \xrightarrow{\text{meas}} g$$
then $f(\omega) = g(\omega)$ a.e.

3. Let f_n, f, g_n, g be $\mathcal{F}$-measurable, and α, β real numbers. If
$$f_n \xrightarrow{\text{meas}} f \qquad \text{and} \qquad g_n \xrightarrow{\text{meas}} g$$
then
$$\alpha f_n + \beta g_n \xrightarrow{\text{meas}} \alpha f + \beta g.$$

Example 1.12. Consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.

1. Let $f_n = \frac{1}{n}\chi_{(0,n)}$.
   (a) Since for all $x \in \mathbb{R}$, $|f_n(x) - 0| \leq \frac{1}{n} \to 0$ then $f_n \Longrightarrow 0$ on $\mathbb{R}$.
   (b) Obviously, $f_n \in L^p(\mathbb{R})$ for all $1 \leq p < \infty$. Now for all $m > n$,
$$\begin{aligned}\|f_m - f_n\|_1 &= \int \left|\frac{1}{m}\chi_{(0,m)} - \frac{1}{n}\chi_{(0,n)}\right| d\lambda \\ &= \int \left|\frac{1}{m} - \frac{1}{n}\right| \chi_{(0,n]}\, d\lambda + \int \frac{1}{m}\chi_{(n,m)}\, d\lambda \\ &= \left|\frac{n-m}{nm}\right| n + \frac{1}{m}(m-n) = \frac{m-n}{m} + \frac{m-n}{m} = 2 - 2\frac{n}{m}\end{aligned}$$
   which shows that
$$\|f_m - f_n\|_1 \geq 1$$
   whenever $m \geq 2n$. Hence, $\{f_n\}$ is not Cauchy and thus does not converge in $L^1(\mathbb{R})$.
   On the other hand, if $p > 1$ then
$$\|f_n - 0\|_p^p = \int \left|\frac{1}{n}\chi_{(0,n)}\right|^p d\lambda = \frac{1}{n^p} \int \chi_{(0,n)}\, d\lambda = \frac{1}{n^p} \cdot n = \frac{1}{n^{p-1}} \to 0$$
   as $n \to \infty$. This shows that $f_n \xrightarrow{\|\cdot\|_p} 0$ in $L^p(\mathbb{R})$ for all $1 < p < \infty$.
   (c) For fixed $\epsilon > 0$, set
$$E_{n,\epsilon} = \{\omega \in \Omega : |f_n(\omega) - 0| > \epsilon\}.$$
   Let us choose N such that $\frac{1}{N} < \epsilon$. Then for all $n \geq N$ and $x \in \mathbb{R}$ we have
$$|f_n(x)| \leq \frac{1}{n} \leq \frac{1}{N} < \epsilon,$$
   that is, $E_{n,\epsilon} = \emptyset$. Hence
$$\lim_{n\to\infty} \lambda\big(E_{n,\epsilon}\big) = \lim_{n\to\infty} \lambda(\emptyset) = \lim_{n\to\infty} 0 = 0.$$
   which shows that $f_n \xrightarrow{\text{meas}} 0$.

2. Now let $f_n = \chi_{(n,n+1)}$.

   (a) For $m \neq n$ we have
   $$\sup_{x\in\mathbb{R}} |f_m(x) - f_n(x)| = 1$$
   which shows that $\{f_n\}$ is not uniformly Cauchy, hence does not converge uniformly.

   (b) On the other hand, given $x \in \mathbb{R}$, pick $n \in \mathbb{N}$ such that $x < N$. Then for all $n \geq N$ we have $x \notin (n, n+1)$, that is, $f_n(x) = 0$. This shows that $f_n(x) \to 0$ pointwise on $\mathbb{R}$ (and hence trivially, $f_n(x) \to 0$ a.e on $\mathbb{R}$).

   (c) Next let p be arbitrary, $1 \leq p < \infty$. We have for all $m \neq n$,
   $$\begin{aligned} \|f_m - f_n\|_p &= \left[\int \left|\chi_{(m,m+1)} - \chi_{(n,n+1)}\right|^p d\lambda\right]^{1/p} \\ &= \left[\int \left[\chi_{(m,m+1)} + \chi_{(n,n+1)}\right] d\lambda\right]^{1/p} = \sqrt[p]{2}. \end{aligned}$$
   Hence, $\{f_n\}$ is not Cauchy and thus does not converge in $L^p(\mathbb{R})$.

   (d) Now let ϵ be given, $0 < \epsilon < 1$. For each n set
   $$E_{n,\epsilon} = \{\omega \in \Omega : |f_n(\omega) - 0| > \epsilon\}.$$
   Then $E_{n,\epsilon} = (n, n+1)$ by definition of the functions f_n, and
   $$\lim_{n\to\infty} \lambda\big(E_{n,\epsilon}\big) = \lim_{n\to\infty} \lambda\big((n, n+1)\big) = \lim_{n\to\infty} 1 = 1.$$
   This shows that $f_n \overset{\text{meas}}{\not\to} 0$. It will follow from theorem 1.23 below that if this sequence f_n converges in measure to any f, then $f(\omega) = 0$ on each bounded interval; thus $\{f_n\}$ does not converge in measure at all.

The second example above shows that even if $f_n \to f$ a.e., then f_n need not converge in measure. However, for finite measure spaces we have:

Theorem 1.23. *Suppose, $\mu(\Omega) < \infty$. If $f_n \overset{a.e.}{\longrightarrow} f$ then $f_n \overset{meas}{\longrightarrow} f$.*

Proof. Suppose, $f_n \to f$ a.e. Then (replacing f by an appropriate function if necessary as in remark 1.14) f is also $\mathcal{F}$-measurable.

Now given $\epsilon > 0$, let
$$E_n = E_{n,\epsilon} := \{\omega \in \Omega : |f_n(\omega) - f| > \epsilon\}.$$
We need to show that $\lim_{n\to\infty} \mu(E_n) = 0$.

To do so, for each k we set
$$A_k := \bigcup_{n=k}^{\infty} E_n \qquad \big(= \{\omega : |f_n(\omega) - f(\omega)| > \epsilon \text{ for some } n \geq k\big).$$

Observe that $\{A_k\}\downarrow$ and

$$\begin{aligned} \omega \in \bigcap_{k=1}^{\infty} A_k \underset{\text{rem. 1.8}}{\iff} & \ \omega \text{ is contained in infinitely many } E_n \\ \iff & \ |f_n(\omega) - f(\omega)| > \epsilon \quad \text{for infinitely many } n \\ \implies & \ f_n(\omega) \not\to f(\omega). \end{aligned}$$

Thus,

$$\bigcap_{k=1}^{\infty} A_k \subseteq \{\omega : f_n(\omega) \not\to f(\omega)\}.$$

By assumption, the right-hand set is contained in some null set, hence

$$\mu\Big(\bigcap_{k=1}^{\infty} A_k\Big) = 0.$$

Now as $\mu(\Omega) < \infty$ we have by theorem 1.3 that

$$\lim_{k\to\infty} \mu(A_k) = \mu\Big(\bigcap_{k=1}^{\infty} A_k\Big) = 0.$$

Since $E_k \subseteq A_k$ for all k we can apply the Sandwich theorem to obtain that

$$\lim_{k\to\infty} \mu(E_k) = 0$$

as well. This proves the theorem. □

The next example shows that the converse statement of this theorem is wrong. In fact, it shows that if $f_n \xrightarrow{\text{meas}} f$ then $\{f_n\}$ need not converge a.e.

Example 1.13. Define a sequence of measurable functions $f_n : [0,1) \to \mathbb{R}$ as follows.

For each $n \in \mathbb{N}$, there exist a unique pair $k = k(n) \in \mathbb{N}_0$ and $m = m(n) \in \mathbb{N}$, $0 \le m < 2^k$, such that

$$n = 2^k + m.$$

Set

$$f_n := \chi_{\left[\frac{m}{2^k}, \frac{m+1}{2^k}\right)}.$$

For example,

$$f_1 = \chi_{[0,1)},\quad f_2 = \chi_{[0,\frac{1}{2})},\quad f_3 = \chi_{[\frac{1}{2},1)},\quad f_4 = \chi_{[0,\frac{1}{4})},\quad f_5 = \chi_{[\frac{1}{4},\frac{2}{4})},\quad f_6 = \chi_{[\frac{2}{4},\frac{3}{4})},$$
$$f_7 = \chi_{[\frac{3}{4},1)},\quad f_8 = \chi_{[0,\frac{1}{8})},\quad f_9 = \chi_{[\frac{1}{8},\frac{2}{8})},\quad f_{10} = \chi_{[\frac{2}{8},\frac{3}{8})},\quad f_{11} = \chi_{[\frac{3}{8},\frac{4}{8})},\quad \ldots$$

Observe that $k(n) \to \infty$ as $n \to \infty$. We now consider various modes of convergence for the sequence $\{f_n\}$.

1. *Claim:* : $f_n \overset{\text{meas}}{\longrightarrow} 0$.
   In fact, for each ϵ, $0 \leq \epsilon < 1$, we have

$$E_{n,\epsilon} := \{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\} = \Big[\frac{m}{2^k}, \frac{m+1}{2^k}\Big).$$

   Hence,

$$\lambda(E_{n,\epsilon}) = \lambda\Big(\Big[\frac{m}{2^k}, \frac{m+1}{2^k}\Big)\Big) = \frac{1}{2^{k(n)}} \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty.$$

   This proves the claim.

2. *Claim:* : $f_n \overset{\|\cdot\|_p}{\longrightarrow} 0$ for all $1 \leq p < \infty$.
   In fact,

$$\|f_n - 0\|_p^p = \int |f_n|^p \, d\lambda = \int \chi_{\left[\frac{m}{2^k}, \frac{m+1}{2^k}\right)} \, d\lambda = \lambda\Big(\Big[\frac{m}{2^k}, \frac{m+1}{2^k}\Big]\Big) = \frac{1}{2^{k(n)}} \longrightarrow 0$$

   as $n \longrightarrow \infty$. This proves the claim.

3. *Claim:* : $f_n \overset{\text{a.e.}}{\not\longrightarrow} 0$.
   (Observe that if $f_n \longrightarrow f$ a.e., then by theorem 1.23 and part 1, $f = 0$ a.e.)
   Let $x \in [0, 1)$ be arbitrary. Now for each k,

$$\left\{\frac{0}{2^k}, \frac{1}{2^k}, \frac{1}{2^k}, \ldots, \frac{2^K}{2^k}\right\}$$

   is a partition of $[0, 1]$ into subintervals of equal length. Thus for each k there exists a unique m_k, $0 \leq m_k < 2^k$ with $x \in \left[\frac{m_k}{2^k}, \frac{m_k+1}{2^k}\right)$. Set

$$n_k := 2^k + m_k \qquad (k = 1, 2, 3, \ldots).$$

   Then

$$f_{n_k}(x) = \chi_{\left[\frac{m_k}{2^k}, \frac{m_k+1}{2^k}\right)}(x) = 1$$

   for all k, that is,

$$\lim_{k \to \infty} f_{n_k}(x) = 1.$$

   We have shown that for *each* $x \in [0, 1)$, $f_n(x) \not\longrightarrow 0$, hence the claim follows.

Observe that in the above example, one can construct various subsequences of $\{f_n\}$ converging to 0 a.e. For example, if we choose $n_k = 2^k$ then we have $f_{n_k}(x) = \chi_{\left[0, \frac{1}{2^k}\right)}(x) \longrightarrow 0$ for all $x \neq 0$.

In fact, we have in general:

Theorem 1.24. *If $f_n \overset{meas}{\longrightarrow} f$ then there exists a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ such that $f_{n_k} \overset{a.e.}{\longrightarrow} f$.*

Proof. By assumption, for each $\epsilon > 0$ we have

$$\lim_{n\to\infty} \mu(E_{n,\epsilon}) = 0 \qquad \text{where} \quad E_{n,\epsilon} := \{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\}.$$

That is, given $\delta > 0$ there exists $N = N(\epsilon, \delta)$ such that

$$\mu\Big(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\} \Big) < \delta \qquad \forall n \geq N. \tag{1.24}$$

Now we extract a subsequence of $\{f_n\}$ inductively. By (1.24), choosing $\epsilon = 1$ and $\delta = \frac{1}{2}$, there exists $n_1 \in \mathbb{N}$ such that

$$\mu\Big(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > 1\} \Big) < \frac{1}{2} \qquad \forall n \geq n_1.$$

Next choosing $\epsilon = \frac{1}{2}$ and $\delta = \frac{1}{4}$, there exists $n_2 > n_1$ such that

$$\mu\Big(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \frac{1}{2}\} \Big) < \frac{1}{2} \qquad \forall n \geq n_2.$$

Suppose we have picked a positive integer n_k such that

$$\mu\Big(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \frac{1}{k}\} \Big) < \frac{1}{2^k} \qquad \forall n \geq n_k. \tag{1.25}$$

Then by (1.24), choosing $\epsilon = \frac{1}{k+1}$ and $\delta = \frac{1}{2^{k+1}}$, there exists $n_{k+1} > n_k$ such that

$$\mu\Big(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \frac{1}{k+1}\} \Big) < \frac{1}{2^{k+1}} \qquad \forall n \geq n_{k+1}.$$

We thus obtain a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ satisfying

$$\mu\Big(\underbrace{\{\omega \in \Omega : |f_{n_k}(\omega) - f(\omega)| > \frac{1}{k}\}}_{=:A_k} \leq \frac{1}{2^k} \qquad \forall k.$$

Set

$$A := \bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} A_k.$$

Since

$$\sum_{k=1}^{\infty} \mu(A_k) \leq \sum_{k=1}^{\infty} 2^k = 1$$

then by the Borel-Cantellini Theorem, $\mu(A) = 0$.

Now let $x \in A^c$ (i.e. outside of a null set). Then $x \notin \bigcup_{k=j}^{\infty} A_k$ for some j, that is, $x \notin A_k$ for all $k \geq j$. Equivalently,

$$|f_{n_k}(\omega) - f(\omega)| \leq \frac{1}{k} \qquad \forall k \geq j.$$

It follows that $|f_{n_k}(\omega) - f(\omega)| \to 0$ as $k \to \infty$. We have shown that

$$f_{n_k}(\omega) \to f(\omega) \qquad \forall x \notin A$$

and thus proved the theorem. □

Theorem 1.25. *Let $1 \leq p < \infty$ and suppose $f_n, f \in L^p(\Omega, \mathcal{F}, \mu)$ for all n. If $f_n \xrightarrow{\|\cdot\|_p} f$ then $f_n \xrightarrow{meas} f$.*

Proof. Let $\epsilon > 0$ be arbitrary. Then be Chebychev's inequality,

$$\mu\Big(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \epsilon\} \Big) = \mu\Big(\{\omega \in \Omega : |(f_n - f)(\omega)| > \epsilon\} \Big)$$
$$\leq e^{-p} \|f_n - f\|_p^p \to 0$$

as $n \to \infty$ by assumption. This shows that $f_n \xrightarrow{\text{meas}} f$. □

Exercise 1.12. Suppose, $f_n \Longrightarrow f$ on Ω. Show:

1. $f_n \xrightarrow{\text{meas}} f$
2. If $\mu(\Omega) < \infty$ then $f_n \xrightarrow{\|\cdot\|_p} f$.
3. If $\mu(\Omega) = \infty$ then $\{f_n\}$ need not converge to f in $\|\cdot\|_p$.

1.9 Product Spaces and Fubini's Theorem

Given two sets X and Y which carry a measure space structure, we want to make $X \times Y$ into a measure space compatible with the structures on X and Y. We just give a brief overview over the main ideas here, proofs of the theorems can be found in [2] or [3], for example.

First let $(X, \mathcal{F})$ and $(Y, \mathcal{E})$ be two measurable spaces. We introduce a σ-algebra onto $X \times Y$ a follows.

Start with sets of the form

$$A \times B \qquad\qquad A \in \mathcal{F},\ B \in \mathcal{E}.$$

Such sets are called *measurable rectangles.* Then we define

$$\mathcal{F} \otimes \mathcal{E} := \sigma\Big(\{A \times B : A \times B \text{ is a measurable rectangle}\} \Big).$$

Thus, $\mathcal{F} \times \mathcal{E}$ is the σ-algebra generated by all measurable rectangles. We call $\mathcal{F} \times \mathcal{E}$ the *product σ-algebra* on $X \times Y$.

Remark 1.21. In a similar way, if $(X_1, \mathcal{F}_1)$, $(X_2, \mathcal{F}_2)$, …,$(X_n, \mathcal{F}_n)$ is a finite collection of measurable spaces, we define the product σ-algebra by

$$\mathcal{F}_1 \times \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n := \sigma\Big(\{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathcal{F}_i,\ i = 1 \ldots n\} \Big).$$

Example 1.14. Consider the measurable spaces $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. One can show that

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R}).$$

(For the proof, see [3].) So the Borel σ-algebra on $\mathbb{R}^d$ is generated by measurable rectangles. Similarly,

$$\mathcal{B}(\mathbb{R}^{d+m}) = \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^m).$$

However, if $\mathcal{M}_\lambda(\mathbb{R}^d)$ denotes the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}^d$, then

$$\mathcal{M}_\lambda(\mathbb{R}^d) \supsetneq \mathcal{M}_\lambda(\mathbb{R}) \otimes \mathcal{M}_\lambda(\mathbb{R}) \otimes \mathcal{M}_\lambda(\mathbb{R}) \otimes \cdots \otimes \mathcal{M}_\lambda(\mathbb{R}).$$

Example 1.15. Let $(X, \mathcal{F})$ and $(Y, \mathcal{E})$ be measurable spaces, let $f : X \to \mathbb{R}$ be $\mathcal{F}$-measurable and $g : Y \to \mathbb{R}$ be $\mathcal{E}$-measurable. Then f and g define functions $\hat{f}, \hat{g} : X \times Y$ which are constant with respect to one variable by

$$\hat{f}(x, y) = f(x) \qquad \text{and} \qquad \hat{g}(x, y) = g(y).$$

Claim: $\hat{f}$ is $\mathcal{F} \otimes E$-measurable. In fact, for each $a \in \mathbb{Q}$ we have

$$\begin{aligned} \{(x, y) \in X \times Y : \hat{f}(x, y) < a\} &= \{(x, y) \in X \times Y : f(x) < a\} \\ &= \underbrace{\{x \in X : f(x) < a\}}_{\in \mathcal{F} \text{ by cor. 1.7}} \times Y, \end{aligned}$$

a measurable rectangle in $\mathcal{F} \otimes \mathcal{E}$. By corollary 1.7, the claim follows.

One shows in a similar way that $\hat{g}$ is $\mathcal{F} \otimes E$-measurable. It follows that

$$h(x, y) := f(x)g(y) = \hat{f}(x, y)\hat{g}(x, y)$$

is a product of $\mathcal{F} \otimes E$-measurable functions, and hence is measurable.

Theorem 1.26. *Let $(X, \mathcal{F})$ and $(Y, \mathcal{E})$ be measurable spaces.*

1. *Let $E \in \mathcal{F} \otimes \mathcal{E}$. Given $x \in X$ and $y \in Y$ arbitrary but fixed, set*

$$\begin{aligned} E_x &:= \{y \in Y : (x, y) \in E) && \text{("x-section")} \\ E^y &:= \{x \in X : (x, y) \in E) && \text{("y-section")} \end{aligned}$$

   Then $E_x \in \mathcal{E}$ and $E_y \in \mathcal{F}$.

2. *Let $f : X \times Y \to \mathbb{R}$ be $\mathcal{F} \otimes \mathcal{E}$-measurable. Given $x \in X$ and $y \in Y$ arbitrary but fixed, define*

$$\begin{aligned} f_x &: Y \to \mathbb{R} && \text{by} && f_x(y) = f(x, y) && \text{("x-section")} \\ f^y &: X \to \mathbb{R} && \text{by} && f^y(x) = f(x, y) && \text{("y-section").} \end{aligned}$$

   Then f_x is $\mathcal{E}$-measurable and f^y is $\mathcal{F}$-measurable.

Now let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{E}, \nu)$ be measure spaces. Given a measurable rectangle $A \times B$ we set

$$(\mu \times \nu)(A \times B) := \mu(A)\nu(B)$$

and hope to be able to extend $\mu \times \nu$ to a measure on $\mathcal{F} \otimes \mathcal{E}$.

Theorem 1.27. *Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{E}, \nu)$ be* σ-finite *measure spaces. There exists a unique measure on $(X \times Y, \mathcal{F} \otimes \mathcal{E})$ denoted by $\mu \times \nu$ and satisfying*

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B) \qquad \forall A \in \mathcal{F},\ B \in \mathcal{E}.$$

(Consequently, we call the measure space $(X \times Y, \mathcal{F} \otimes \mathcal{E}, \mu \times \nu)$ the *product measure space* of $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{E}, \nu)$.)

Theorems 1.26 and 1.27 generalize to products of more than two spaces, as does theorem 1.28 below.

Example 1.16. Consider

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R}).$$

One can show that the Lebesgue measure λ_d on $\mathcal{B}(\mathbb{R}^d)$ is the product of the Lebesgue measures λ on $\mathcal{B}(\mathbb{R})$,

$$\lambda_d = \lambda \times \lambda \times \lambda \times \cdots \times \lambda.$$

Theorem 1.28. *Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{E}, \nu)$ be* σ-finite *measure spaces and let $f : X \times Y \to \mathbb{R}$ be $\mathcal{F} \otimes \mathcal{E}$-measurable.*

1. (Tonelli's Theorem). *Suppose, $f \geq 0$. Then the function*

$$g(x) := \int_Y f_x(y)\, d\nu = \int_Y f(x, y)\, d\nu(y) \qquad \text{is } \mathcal{F}\text{-measurable}$$
$$h(y) := \int_X f_y(x)\, d\mu = \int_X f(x, y)\, d\mu(x) \qquad \text{is } \mathcal{E}\text{-measurable}$$

and the double integral can be written as an iterated integral,

$$\int_{X \times Y} f(x, y)\, d(\mu \times \nu) = \int_X \underbrace{\int_Y f(x, y)\, d\nu}_{=g(x)}\, d\mu = \int_Y \underbrace{\int_X f(x, y)\, d\mu}_{=h(y)}\, d\nu. \tag{1.26}$$

2. (Fubini's Theorem). *Suppose, $f \in L^1(X \times Y, \mathcal{F} \otimes \mathcal{E}, \mu \times \nu)$. Then*

$$f_x \in L^1(Y, \mathcal{E}, \nu) \quad a.e.\ x$$
$$f^y \in L^1(X, \mathcal{F}, \mu) \quad a.e.\ y$$

and

$$\begin{aligned} g(x) &:= \int_Y f_x(y)\, d\nu = \int_Y f(x, y)\, d\nu(y) \in L^1(X, \mathcal{F}, \mu) \\ h(y) &:= \int_X f_y(x)\, d\mu = \int_X f(x, y)\, d\mu(x) \in L^1(Y, \mathcal{E}, \nu) \end{aligned} \tag{1.27}$$

and identity (1.26) holds.

Remark 1.22. 1. In Tonelli's theorem, it may happen that $g(x) = \infty$ or $h(y) = \infty$. For the outer integrals in (1.26) to make sense, one thus needs to introduce the concept of measurability as well as the Lebesgue integral for *extended real valued functions* $f : \Omega \to [0, \infty]$. Measurability can be defined using the equivalent properties of corollary 1.7. Since the structure theorem 1.13 holds for such functions as well, definition 1.14 can be used to define the integral of such functions.

We have not introduced extended real valued functions earlier, since we usually do not encounter them in probability.

2. In Fubini's theorem, Since $f_x, f^y \in L^1$ a.e. only, the functions $f(x)$ and $g(y)$ are defined a.e. only. However, as usual by suitably modifying f_x on null subsets of X, and f^y on null subsets of Y, the integrals in (1.27) can be made to make sense.

3. Given a $\mathcal{F} \times \mathcal{E}$-measurable function $f : X \times Y \to \mathbb{R}$, one usually first applies Tonelli's theorem to $|f|$ in order to check whether $f \in L^1(X \times Y)$. Then one can use Fubini's theorem to express the double integral of f as an iterated integral as in (1.26). The next example illustrates this idea.

Example 1.17. Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{E}, \nu)$ be *σ-finite* measure spaces, $f \in L^1(X)$ and $g \in L^1(Y)$. Set $h(x, y) := f(x)g(y)$. Then by example 1.15, h is $\mathcal{F} \times \mathcal{E}$-measurable.

Claim: : h is integrable, and

$$\int_{X \times Y} h(x, y)\, d(\mu \times \nu) = \Big[\int_X f(x)\, d\mu\Big] \Big[\int_Y g(y)\, d\nu\Big].$$

In fact, we have

$$\begin{aligned}
\int_{X \times Y} |h(x, y)|\, d(\nu \times \mu) &= \int_X \int_Y |h(x, y)|\, d\nu\, d\mu \qquad \text{(by Tonelli)} \\
&= \int_X \int_Y |f(x)|\, |g(y)|\, d\nu\, d\mu = \int_X |f(x)| \Big[\int_Y |g(y)|\, d\nu\Big] d\mu \\
&= \int_X |f(x)|\, \|g\|_1\, d\mu = \|g\|_1 \int_X |f(x)|\, d\mu = \|g\|_1\, \|f\|_1 < \infty.
\end{aligned}$$

Hence, $h \in L^1(X \times Y)$ and we can apply Fubini's theorem to repeat essentially the same computations,

$$\begin{aligned}
\int_{X \times Y} h(x, y)\, d(\nu \times \mu) &= \int_X \int_Y h(x, y)\, d\nu\, d\mu \qquad \text{(by Fubini)} \\
&= \int_X \int_Y f(x)\, g(y)\, d\nu\, d\mu \\
&= \int_X f(x) \Big[\int_Y g(y)\, d\nu\Big] d\mu = \Big[\int_X f(x)\, d\mu\Big] \Big[\int_Y g(y)\, d\nu\Big].
\end{aligned}$$

This proves the claim.

1.10 The Radon-Nikodym Theorem

Definition 1.18. Let μ and ν be two measures on a measurable space $(\Omega, \mathcal{F})$. If for all $E \in \mathcal{F}$,

$$\mu(E) = 0 \;\Rightarrow\; \nu(E) = 0$$

then we write "$\nu \prec \mu$" and say that ν *is absolutely continuous with respect to* μ.

Note that $\nu \prec \mu$ does *not* mean that $\nu(E) \leq \mu(E)$ for all $E \in \mathcal{F}$. Instead, it means that every null set of μ is also a null set of ν.

We have all the tools available to prove:

Exercise 1.13. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and fix an $\mathcal{F}$-measurable function $h : \Omega \to [0, \infty)$. Set

$$\nu(E) := \int_E h \, d\mu \qquad \forall E \in \mathcal{F}.$$

Show:

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

2. Let $f : \Omega \to \mathbb{R}$ be any $\mathcal{F}$-measurable function. Then for *every* null set $E \in \mathcal{F}$,

$$\int_E f \, d\mu = 0.$$

   (Hint: Use exercise 1.7).

3. $\mu \prec \nu$.

4. If $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable, and either $f \geq 0$ or $f \in L^1(\Omega\, F, \nu)$ then

$$\int f \, d\nu = \int f h \, d\mu.$$

5. ν is a finite measure $\Leftrightarrow$ $h \in L^1(\Omega\, F, \mu)$.

6. If $\tilde{h}$ is another such function with corresponding measure $\tilde{\nu}$, $\tilde{\nu}(E) = \int_E \tilde{h} \, d\mu$, and $h = \tilde{h}$ a.e. then $\nu = \tilde{\nu}$.

It turns out that every measure ν which is absolutely continuous with respect to μ is of this form:

Theorem 1.29. (Radon-Nikodym Theorem). *Let μ and ν be σ-finite measures on a measurable space $(\Omega, \mathcal{F})$. Suppose that $\nu \prec \mu$. Then there exists an $\mathcal{F}$-measurable function $h \geq 0$ such that*

$$\nu(E) = \int_E h \, d\mu \qquad \forall E \in \mathcal{F}.$$

h is essentially unique. That is, if $\tilde{h}$ is another $\mathcal{F}$-measurable function with the property that $\nu(E) = \int_E \tilde{h} \, d\mu \; \forall E \in \mathcal{F}$, then $h = \tilde{h}$ a.e.

Proof. See [2] or [3]. □

Remark 1.23. σ-finiteness cannot be dropped in this theorem. For example, consider the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with the Lebesgue measure λ and the counting measure μ. Obviously, $\lambda \prec \mu$ as the only set with zero counting measure is the empty set. Now suppose, there exists h as in the theorem. Let $\{x\}$ be any singleton. We obtain

$$0 = \lambda(\{x\}) = \int_{\{x\}} h\, d\mu = \int h\chi_{\{x\}}\, d\mu = h(x)\mu(\{x\}) = h(x)$$

so that $h(x) = 0$ for all $x \in \mathbb{R}$. But then for all $E \in \mathcal{B}(\mathbb{R})$,

$$\lambda(E) = \int_E h\, d\mu = \int_E 0\, d\mu = 0$$

which contradicts the fact that $\lambda \neq 0$.

Chapter 2

Fundamental Concepts in Probability

The fundamental object in probability is a finite measure space $(\Omega, \mathcal{F}, \mu)$ where $\mu(\Omega) = 1$. However, one uses a vocabulary different from measure theory:

Measure Theory (Analysis)	*Probability Theory*
measure space$(\Omega, \mathcal{F}, \mu)$	probability space $(\Omega, \mathcal{F}, P)$
finite measure μ, $\mu(\Omega) = 1$	probability measure P
point $w \in \Omega$	outcome ω
measurable set $A \in \mathcal{F}$	event A
measurable function $f : \Omega \to \mathbb{R}$	random variable $X : \Omega \to \mathbb{R}$
measurable function $f : \Omega \to \mathbb{R}^d$	d-dimensional random variable $X : \Omega \to \mathbb{R}^d$
integral of f, $\int f \, d\mu$	expectation (or mean) $E(X)$ of X
$f \in L^p(\Omega)$	X has finite p-th moment, or X is a L^p-random variable
f is integrable	$E(\lvert X \rvert) < \infty$
almost everywhere (a.e.)	almost surely (a.s.)
convergence in measure	convergence in probability
regular Borel measure μ on $\mathbb{R}^d$, $\mu(\mathbb{R}^d) = 1$	distribution
characteristic function χ_A	indicator function $\mathbf{1}_A$
$f^{-1}\big((a, \infty)\big) = \{\omega : f(\omega) > a\}$	$\{X > a\} = \{\omega : X(w) > a\}$
$f^{-1}(A) = \{\omega : f(\omega) \in A\}$	$\{X \in A\} = \{\omega : X(w) \in A\}$
$\mu\big(\{\omega : f(\omega) > a\}\big)$	$P(X > a) = P\big(\{\omega : X(\omega) > a\}\big)$
$\mu\big(\{\omega : f(\omega) \in A\}\big)$	$P(X \in A) = P\big(\{\omega : X(\omega) \in A\}\big)$

We will freely switch between both sets of vocabularies.

2.1 Random Variables

Let us introduce the basic vocabulary of probability by an example:

Example 2.1. Suppose we are throwing an unloaded die. We are thus considering the *sample space*

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

The *outcomes*, i.e. the elements ω of this sample space, are simply the six numbers on the die. As σ-algebra we choose $\mathcal{F} = \mathcal{P}(\Omega)$.

The word "unloaded" means that each number will be thrown with exactly the same probability, so that

$$P(\{\omega\}) = \frac{1}{6} \qquad \forall \omega \in \Omega.$$

Thus, the probability measure P on Ω is the scaled counting measure. Now the (measurable) set $A_1 = \{1, 2, 3\}$ constitutes the event "a number ≤ 3 is thrown" while the set $A_2 = \{2, 4, 6\}$ constitutes the event "an even number is thrown". The probability that the latter happens is obviously

$$P(A_2) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}.$$

Now suppose we play for money, with wins and losses at each throw as outlined in the following table:

Number thrown (outcome ω)	money gained $X(w)$
$\omega_1 = 1$	-300
$\omega_2 = 2$	-100
$\omega_3 = 3$	0
$\omega_4 = 4$	50
$\omega_5 = 5$	100
$\omega_6 = 6$	200

The *random variable* (= function) $X(\omega)$ tells us what the win/loss at each outcome is. For example,

$$P(X > 0) = P(\{\omega : X(\omega) > 0\}) = P(\{4, 5, 6\}) = \frac{1}{2} = 50\%$$

is the probability that we win some money. On the other hand, the probability that we loose some money is

$$P(X < 0) = P(\{\omega : X(\omega) < 0\}) = P(\{1, 2\}) = \frac{1}{3} = 33.\dot{3}\%.$$

In spite of these favourable odds, throwing the dice is not profitable as the *expected win* (=average win) is

$$\mu = E(X) = \int_\Omega X\,dP = \sum_{k=1}^{6} X(\omega_k)P(\{\omega_k\})$$
$$= [-300 - 200 + 0 + 50 + 100 + 200]\cdot\frac{1}{6} = \frac{-50}{6} = -8\frac{1}{3} < 0.$$

This means that if we repeat throwing the die many times, then we will loose $8\frac{1}{3}$ units of money per throw on average. (This is called the "law of large numbers".)

Now suppose, the rules change and at each throw we have different rules for winning/loosing money:

Number thrown (outcome ω)	1st throw $X_1(w)$	2nd throw $X_2(\omega)$	3rd throw $X_3(\Omega)$	...
$\omega_1 = 1$	-300	-500	-1000	...
$\omega_2 = 2$	-100	-200	-800	...
$\omega_3 = 3$	0	0	-500	...
$\omega_4 = 4$	50	100	-400	...
$\omega_5 = 5$	100	200	500	...
$\omega_6 = 6$	200	500	-2000	...

We now have a sequence

$$X_1,\ X_2,\ X_3, \ldots$$

of random variables, called a (discrete) *stochastic process.*

Example 2.2. A factory produces boxes in the shape of a cube whose edges each have length 100cm. Because of old machinery, the actual length of the edges of each box produced differs from 100cm by an error ω; each box produced is still a cube. Experience shows that $-10\text{cm} \le \omega \le 10\text{cm}$.

We thus choose as sample space the set $\Omega = [-10, 10]$. Since the σ-algebra $\mathcal{F}$ should contain all intervals, we choose $\mathcal{F} = \mathcal{B}\big([-10, 10]\big)$. The probability measure P should be chosen so that for all subintervals I of $[-10, 10]$,

$$P(I) = \text{probability that the error lies in } I.$$

For example $I = [-1, 1]$ is the event "the error is $\le 1\,\text{cm}$" and $P\big([-1, 1]\big)$ is the probability that the dimensions of the box fall within 1 cm of the desired value. Similarly, $P\big((0, 10]\big)$ is the probability that the box is larger than desired.

The computed the error in volume will be given by a random variable $X(\omega)$,

$$X(\omega) = (100 + \omega)^3 - 100^3 = 30000\omega + 300\omega^2 + \omega^3. \qquad \text{(units in cm}^3\text{)}$$

Thus $P(|X| \le 2)$ is the probability that the volume of a box produced is within $2\,\text{cm}^3$ of the desired volume of $10^6\,\text{cm}^3$. Similarly, $P(X > 0)$ is the probability that the box produced is larger than desired; hence we must have $P(X > 0) = P\big((0, 10]\big)$. The *expected (average, mean) error* in volume of a box produced is

$$\mu = E(X) = \int_{-10}^{10} X(\omega)\, dP(\omega).$$

We now define the notions of mean, variance and standard derivation. For this, let $(\Omega, \mathcal{F}, P)$ be a probability space. Since $P(\Omega) = 1$, then by exercise 1.10, $L^p(\Omega) \subset L^1(\Omega)$ for all $1 < p < \infty$, and

$$\|X\|_1 \leq \|X\|_p \qquad \forall X \in L^p(\Omega).$$

Definition 2.1. Let $X : \Omega \to \mathbb{R}$ a random variable.

1. Suppose $X \in L^1(\Omega)$. The *mean* μ or also *expected value* of X is

$$\mu = E(X) := \int_\Omega X\, dP.$$

2. Suppose $X \in L^2(\Omega)$ (and hence $X \in L^1(\Omega)$ so that μ is defined). The *variance* of X is

$$\operatorname{var}(X) := E\big[\,(X - \mu)^2\big] = \|X - \mu\|_2^2. \tag{2.1}$$

The *standard deviation* of X is

$$\operatorname{var}(X) := \sqrt{E\big[\,(X - \mu)^2\big]} = \|X - \mu\|_2 \tag{2.2}$$

Remark 2.1. 1. Since P is a finite measure, the constant function μ is square integrable; hence $\operatorname{var}(X)$ is well defined. Variance and standard deviation tell us by how much $X(\omega)$ differs from its mean μ in the square mean.

2. Let $X \in L^2(\Omega)$. Then

$$\begin{aligned}\operatorname{var}(X) &= \int_\Omega (X - \mu)^2\, dP = \int_\Omega (X^2 - 2\mu X + \mu^2)\, dP \\ &= \int_\Omega X^2\, dP - 2\mu \int_\Omega X\, dP + \mu^2 \int 1\, dP = E(X^2) - 2\mu E(X) + \mu^2 \cdot 1.\end{aligned}$$

as $P(\Omega) = 1$. Now since $\mu = E(X)$ we obtain

$$\boxed{\operatorname{var}(X) = E(X^2) - E(X)^2}.$$

Example 2.3. Consider $X(\omega)$ of example 2.1. As $X(\omega) = -8\frac{1}{3} = -\frac{25}{3}$ and

$$E(X^2) = \big[\,(-300)^2 + (-100)^2 + 0^2 + 50^2 + 100^2 + 200^2\,\big] \cdot \frac{1}{6} = \frac{152500}{6} = \frac{228750}{9}$$

we obtain

$$\begin{aligned}\operatorname{var}(X) &= E(X^2) - E(X)^2 = \frac{228750}{9} - \frac{625}{9} = \frac{228125}{9} = 25347\frac{2}{9} \\ \sigma(X) &= \sqrt{\frac{228125}{9}} \approx 159.21.\end{aligned}$$

2.2 Distributions and Density Functions

Many properties of random variables can be discussed independently of the underlying probability space $(\Omega, \mathcal{F}, P)$. In fact, below we will show:

1. Each random variable X naturally defines a measure μ_X on $\mathcal{B}(\mathbb{R})$ (resp. $\mathcal{B}(\mathbb{R}^d)$).

2. If $\mu_X = \mu_{\tilde{X}}$ for two random variables X and $\tilde{X}$ on probability spaces $(\Omega, \mathcal{F}, P)$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, respectively, then loosely speaking, X and $\tilde{X}$ can be identified.

Exercise 2.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $(\tilde{\Omega}, \mathcal{E})$ another measurable space and $f : \Omega \to \tilde{\Omega}$ a fixed measurable function. For each $A \in \mathcal{E}$ define

$$\mu_f(A) := \mu\big(f^{-1}(A)\big).$$

Show:

1. μ_f is a measure on $(\tilde{\Omega}, \mathcal{E})$.

2. If μ is a probability measure, the so is μ_f.

3. If $h : \tilde{\Omega} \to \mathbb{R}$ is any $\mathcal{E}$-measurable function, then $h \circ f$ is $\mathcal{F}$-measurable, and

$$\int_{\tilde{\Omega}} h \, d\mu_f = \int_{\Omega} h \circ f \, d\mu$$

whenever either integral is defined. Furthermore, $h \in L^1(\tilde{\Omega}, \mathcal{E}, \mu_f) \Leftrightarrow h \circ f \in L^1(\Omega, \mathcal{F}, \mu)$.

Definition 2.2. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}^d$ a random variable. Set $(\tilde{\Omega}, \mathcal{E}) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$.

1. The probability measure on $\mathcal{B}(\mathbb{R}^d)$ defined as in assignment 2.1 is called the *distribution of* X. That is, for all $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu_X(A) = P\big(X^{-1}(A)\big) = P\big(\{w \in \Omega : X(w) \in A\}\big).$$

In short, we write

$$\boxed{\mu_X(A) = P(X \in A)}.$$

Observe that by assignment 2.1, for every Borel measurable function $h : \mathbb{R}^d \to \mathbb{R}$ we have

$$\boxed{\int_{\Omega} h \circ X \, dP = \int_{\mathbb{R}^d} h \, d\mu_X} \tag{2.3}$$

provided that either of these integrals is defined.

2. The function $F_X : \mathbb{R}^d \to [0, \infty)$ given by

$$F_X(t) := \mu_X\big((-\infty, t]\big) = P\big(X \in (-\infty, t]\big) = P\big(\{\omega \in \Omega : X(\omega) \leq t\}\big)$$

is called the *distribution function of* X.

Remark 2.2. In the above definition, if $d \neq 1$ and $t = (t_1, t_2, \dots, t_d)$ then we understand $(-\infty, t]$ to denote the d-interval

$$(-\infty, t] := (-\infty, t_1] \times (-\infty, t_2] \times \cdots \times (-\infty, t_d].$$

Thus, if $X = (X_1, X_2, \dots, X_d)$, then

$$F_X(t) := P\big(\{\omega \in \Omega : X_i(\omega) \leq t_i \ \forall i\}\big).$$

In shortened form,

$$\boxed{F_X(t) = P(X \leq t)}.$$

Remark 2.3. From now on we will assume that all random variables are scalar valued, that is $d = 1$. This is for the sake of ease of presentation; most results below have a simple and natural generalization to the vector valued case.

Example 2.4. Consider the random variable $X(\omega)$ of example 2.1. For all $A \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
\mu_X(A) &= P(X \in A) = P\big(\{w \in \Omega : X(w) \in A\}\big) \\
&= \frac{1}{6}\operatorname{card}\big(\{w \in \Omega : X(w) \in A\}\big) && (P \text{ is } \tfrac{1}{6} \text{ counting measure}) \\
&= \frac{1}{6}\operatorname{card}\big(X(\Omega) \cap A\big) && (X \text{ is one-to-one}) \\
&= \frac{1}{6}\operatorname{card}\big(E \cap A\big). && (E = \{-300, -100, 0, 50, 100, 200\})
\end{aligned}$$

and thus the distribution function is given by

$$F_X(t) = \mu_X\big((-\infty, t]\big) = \frac{1}{6}\operatorname{card}\big((-\infty, t] \cap E\big) = \begin{cases} 0 & t < -300 \\ \frac{1}{6} & -300 \leq t < -100 \\ \frac{1}{3} & -100 \leq t < 0 \\ \frac{1}{2} & 0 \leq t < 50 \\ \frac{2}{3} & 50 \leq t < 100 \\ \frac{5}{6} & 100 \leq t < 200 \\ 1 & t \geq 200. \end{cases}$$

Theorem 2.1. *Let $X : \Omega \to \mathbb{R}$ be a random variable. The distribution function $F = F_X$ has the following properties:*

1. *F is increasing.*

2. *F is continuous from the right. That is, for each $a \in \mathbb{R}$,*

$$\lim_{t \to a^+} F(t) = F(a).$$

3. $$\lim_{t \to -\infty} F(t) = 0 \quad \textit{and} \quad \lim_{t \to \infty} F(t) = 1.$$

Proof. Note that by definition, $0 \leq F_X(t) = P\big(X \in (-\infty, t]\big) \leq P(\Omega) = 1$ for all t.

1. Let $t_1 < t_2$. Then $(-\infty, t_1] \subset (-\infty, t_2]$; hence by monotonicity of the measure,

$$F(t_1) = \mu_X\big((-\infty, t_1]\big) \leq \mu_X\big((-\infty, t_2]\big) = F(t_2).$$

This shows that F is increasing.

2. Fix $a \in \mathbb{R}$. Let $\{t_n\}$ be any sequence in $\mathbb{R}$ with $t_n > a$ for all n, and $t_n \to a$. Set

$$s_n := \sup_{k \geq n} t_k.$$

Then $\{s_n\}$ is a decreasing sequence, $a < s_n$ for all n, and $s_n \to a$ as well. Thus considering the intervals $(-\infty, s_n]$, we have

$$\{\,(-\infty, s_n]\,\}_{n=1}^{\infty} \downarrow \qquad \text{and} \qquad \bigcap_{n=1}^{\infty}(-\infty, s_n] = (-\infty, a].$$

Now as μ_X is a finite measure, then by theorem 1.3,

$$\begin{aligned}\lim_{n\to\infty} F(s_n) &= \lim_{n\to\infty} \mu_X\Big(\,(-\infty, s_n]\,\Big) \\ &= \mu_X\Big(\bigcap_{n=1}^{\infty}(-\infty, s_n]\Big) = \mu_X\Big(\,(-\infty, a]\,\Big) = F(a).\end{aligned}$$

However, $a < t_n < s_n$ for all n; thus by monotonicity of F,

$$\lim_{n\to\infty} F(t_n) = F(a)$$

as well. Since $\{t_n\}$ was arbitrary, it now follows that $\lim_{t\to a^+} F(t) = F(a)$.

3. We proceed just as in 2. First let $\{t_n\}$ be an arbitrary sequence in $\mathbb{R}$ with $t_n \to -\infty$. Let s_n be defined as above, so that $s_n \to -\infty$. Then

$$0 \leq \lim_{n\to\infty} F(s_n) = \lim_{n\to\infty} \mu_X\Big(\,(-\infty, s_n]\,\Big) = \mu_X\Big(\bigcap_{n=1}^{\infty}(-\infty, s_n]\Big) = \mu_X(\emptyset) = 0$$

while by monotonicity of F,

$$0 \leq F(t_n) \leq F(s_n)$$

for all n. It follows that $\lim_{n\to\infty} F(t_n) = 0$. As the sequence $\{t_n\}$ was arbitrary, we conclude that $\lim_{t\to-\infty} F(t) = 0$.

Next let $\{t_n\}$ be an arbitrary sequence in $\mathbb{R}$ with $t_n \to \infty$. For each n, set

$$s_n := \inf_{k \geq n} t_k.$$

Then $\{s_n\}$ is an increasing sequence, and $s_n \to \infty$. Applying theorem 1.3 we obtain

$$\lim_{n\to\infty} F(s_n) = \lim_{n\to\infty} \mu_X\Big((-\infty, s_n] \Big) = \mu_X\Big(\bigcup_{n=1}^{\infty}(-\infty, s_n]\Big) = \mu_X(\mathbb{R}) = 1.$$

Using monotonicity of F and the fact that $s_n \leq t_n$ for all n we obtain that

$$F(s_n) \leq F(t_n) \leq 1$$

for all n, so that $\lim_{n\to\infty} F(t_n) = 1$ as well. As the sequence $\{t_n\}$ was arbitrary, we conclude that $\lim_{t\to\infty} F(t) = 1$.

□

Remark 2.4. One can show a converse statement (see [2] or [3]): Given any Borel measurable function F satisfying 1.–3., there exists a unique probability measure μ on $\mathcal{B}(\mathbb{R})$ satisfying

$$\mu\big((-\infty, t] \big) = F(t)$$

for all $t \in \mathbb{R}$. Thus, there is a one-to-one correspondence between distributions and distribution functions.

Example 2.5. Let $(\Omega, \mathcal{F}, P)$ be any probability space.

1. First let X be integrable, that is, $X \in L^1(\Omega, \mathcal{F}, P)$. Choose $h(x) = x$. Then

$$E(X) = \int_\Omega X(\omega)\, dP(\omega) = \int_\Omega (h \circ X)(\omega)\, dP(\omega) \underset{(2.3)}{=} \int_\mathbb{R} h(x)\, d\mu_X(x).$$

By choice of h,

$$\boxed{E(X) = \int_\mathbb{R} x\, d\mu_X(x)}. \tag{2.4}$$

2. Next let X be square integrable, that is, $X \in L^2(\Omega, \mathcal{F}, P)$. Choose $h(x) = x^2$. Then

$$E(X^2) = \int_\Omega X(\omega)^2\, dP(\omega) = \int_\Omega (h \circ X)(\omega)\, dP(\omega) \underset{(2.3)}{=} \int_\mathbb{R} h(x)\, d\mu_X(x).$$

By choice of h,

$$\boxed{E(X^2) = \int_\mathbb{R} x^2\, d\mu_X(x)}. \tag{2.5}$$

Now choose $h(x) = (x - \mu^2)$ where $\mu = E(X)$ denotes the mean of X. We obtain

$$\operatorname{var}(X) = \int_\Omega \big(X(\omega) - \mu\big)^2\, dP(\omega) = \int_\Omega (h \circ X)(\omega)\, dP(\omega) \underset{(2.3)}{=} \int_\mathbb{R} h(x)\, d\mu_X(x).$$

That is,

$$\boxed{\operatorname{var}(X) = \int_\mathbb{R} (x - \mu)^2\, d\mu_X(x)}. \tag{2.6}$$

Definition 2.3. Two random variables X and Y over probability spaces $(\Omega, \mathcal{F}, P)$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ are said to be *identically distributed*, or *equal in distribution*, if

$$\mu_X = \mu_Y.$$

(or equivalently, if $F_X = F_Y$.) We write

$$X \stackrel{d}{=} Y.$$

By the previous example, if $X \stackrel{d}{=} Y$, then $E(X) = E(Y)$ and $\text{var}(X) = \text{var}(Y)$.

Definition 2.4. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and X a random variable with distribution μ_X. Suppose that $\mu_X \prec \lambda$. (that is, μ_X is absolutely continuous with respect to λ.) Then by the Radon-Nikodym theorem, there exists $f_X \in L^1(\mathbb{R})$, $f \geq 0$, such that

$$\boxed{\mu_X(A) = \int_A f_X(x)\, d\lambda(x)} \qquad \forall\, A \in \mathcal{B}(\mathbb{R}). \tag{2.7}$$

Thus, if F_X is the distribution function of X, then

$$\boxed{F_X(t) = \mu_X\big((-\infty, t]\big) = \int_{-\infty}^{t} f_X(x)\, d\lambda(x)}. \tag{2.8}$$

f_X is called the *density function of* X *(or of* F_X *or* μ_X*)*.

Remark 2.5. Let h be a Borel measurable function with either $h \geq 0$ or $h \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$. Then by assignment 1.13,

$$\int_{\mathbb{R}} h\, d\mu_X \int_{\mathbb{R}} h f_X\, d\lambda. \tag{2.9}$$

Thus, (2.4), (2.5) and (2.6) become

$$E(X) = \int_{\mathbb{R}} x f_X(x)\, d\lambda(x) \qquad (\text{if } X \in L^1(\Omega)) \tag{2.10}$$

$$E(X^2) = \int_{\mathbb{R}} x^2 f_X(x)\, d\lambda(x) \qquad (\text{if } X \in L^2(\Omega)) \tag{2.11}$$

$$\text{var}(X) = \int_{\mathbb{R}} (x-\mu)^2 f_X(x)\, d\lambda(x) \qquad (\text{if } X \in L^2(\Omega)). \tag{2.12}$$

Example 2.6. Return to example 2.4 of throwing a die. We claim that the distribution μ_X has *no* density function.

In fact, by example 2.4,

$$\mu_X(A) = \frac{1}{6}\text{card}(A \cap E) \qquad \forall\, A \in \mathcal{B}(\mathbb{R}),$$

where $E = \{-300, -100, 0, 50, 100, 200\}$. In particular,

$$\mu_X(E) = 1 \neq 0 = \lambda(E).$$

This shows that $\mu_X \not\prec \lambda$ (μ_X is not absolutely continuous with respect to λ.) Then by assignment 1.13, part 3, μ_X has no density function.

Example 2.7. Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ and $P = \lambda$. Consider the two random variables

$$X(\omega) = \omega \qquad \text{and} \qquad Y(\omega) = 1 - \omega.$$

Obviously, $X \neq Y$ a.e. We claim that $X \stackrel{d}{=} Y$.

Observe that range(X)=range(Y). Thus

$$F_X(t) = \mu_X\big((-\infty, t]\big) = P\big(\{\omega : X(w) \leq t\}\big) = \begin{cases} \lambda(\emptyset) = 0 & \text{if } t < 0 \\ \lambda\big([0,t]\big) = t & \text{if } 0 \leq t \leq 1 \\ \lambda\big([0,1]\big) = 1 & \text{if } t > 1 \end{cases}$$

while

$$F_Y(t) = \mu_Y\big((-\infty, t]\big) = P\big(\{\omega : Y(w) \leq t\}\big) = \begin{cases} \lambda(\emptyset) = 0 & \text{if } t < 0 \\ \lambda\big([1-t,1]\big) = t & \text{if } 0 \leq t \leq 1 \\ \lambda\big([0,1]\big) = 1 & \text{if } t > 1. \end{cases}$$

So $F_X = F_Y = F$ and hence $\mu_X = \mu_Y = \mu$. This proves the claim.

Observe that for all $t \in \mathbb{R}$,

$$F(t) \doteq \int_{-\infty}^{t} \mathbf{1}_{(0,1)}(x)\, d\lambda(x)$$

Hence the distribution function is $f(t) = \mathbf{1}_{(0,1)}(t)$.

In general, a distribution μ_X is called *uniform* if its density function is of the form $f_X = \frac{1}{b-a}\mathbf{1}_{(a,b)}$. In this case we write "μ_X is $U(a,b)$."

Example 2.8. Let us begin with the density function

$$f(t) = \frac{1}{\sqrt{2\pi}s}\, e^{-(t-\mu)^2/2s^2}$$

where μ and $s > 0$ are constants. Using residue theory or polar coordinates one can show that

$$\int_{\mathbb{R}} e^{-x^2}\, dx = \sqrt{\pi}.$$

Thus for all μ and s we have

$$\int_{\mathbb{R}} f(t)\, dt = \frac{1}{\sqrt{2\pi}s} \int_{\mathbb{R}} e^{-(t-\mu)^2/2s^2}\, dt = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-x^2}\, dx = 1 \tag{2.13}$$

where we have substituted $x = (t-\mu)/(\sqrt{2}s)$.

Now set

$$F(t) := \int_{-\infty}^{t} f(x)\, dx = \frac{1}{\sqrt{2\pi}s} \int_{-\infty}^{t} e^{-(x-\mu)^2/2s^2}\, dx.$$

Using (2.13) it is easy to see that

1. $F(t)$ is increasing,

2. $F(t)$ is differentiable, hence continuous,

3. $F(t) \to 0$ as $t \to -\infty$ and $F(t) \to 1$ as $t \to \infty$.

Hence, F is a distribution function. Observe that $q(t)f(t) \in L^p(\mathbb{R})$ for any polynomial $q(t)$ and any $1 \le p < \infty$.

Now let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}$ a random variable whose distribution function is F. Then

$$\begin{aligned} E(X) &\underset{(2.10)}{=} \int_{\mathbb{R}} x f(x)\, dx = \frac{1}{\sqrt{2\pi}s} \int_{\mathbb{R}} (x - \mu + \mu)\, e^{-(x-\mu)^2/2s^2}\, dx \\ &= \frac{1}{\sqrt{2\pi}s} \int_{\mathbb{R}} (x-\mu)\, e^{-(x-\mu)^2/2s^2}\, dx + \frac{\mu}{\sqrt{2\pi}s} \int_{\mathbb{R}} e^{-(x-\mu)^2/2s^2}\, dx \\ &\underset{x \to x+\mu}{=} \frac{1}{\sqrt{2\pi}s} \int_{\mathbb{R}} x\, e^{-x^2/2s^2}\, dx + \frac{\mu}{\sqrt{2\pi}s} \int_{\mathbb{R}} e^{-x^2/2s^2}\, dx \underset{(2.13)}{=} 0 + \mu = \mu \end{aligned}$$

because the first integrand is an odd function. That is, μ is the mean of X. Similarly,

$$\begin{aligned} \operatorname{var}(X) &\underset{(2.12)}{=} \int_{\mathbb{R}} (x-\mu)^2 f(x)\, dx = \frac{1}{\sqrt{2\pi}s} \int_{\mathbb{R}} (x-\mu)^2\, e^{-(x-\mu)^2/2s^2}\, dx \\ &\underset{x \to x+\mu}{=} \frac{1}{\sqrt{2\pi}s} \int_{\mathbb{R}} x \cdot x\, e^{-x^2/2s^2}\, dx \\ &\underset{\text{by parts}}{=} \frac{1}{\sqrt{2\pi}s} \left\{ \left[x \cdot \frac{1}{-2/2s^2} e^{-x^2/2s^2} \right]_{-\infty}^{\infty} - \int_{\mathbb{R}} \frac{1}{-2/2s^2} e^{-x^2/2s^2}\, dx \right\} \\ &= 0 + \frac{s^2}{\sqrt{2\pi}s} \int_{\mathbb{R}} e^{-x^2/2s^2}\, dx \underset{(2.13)}{=} s^2 \end{aligned}$$

as $\lim\limits_{|x| \to \infty} e^{-x^2} = 0$. In summary,

$$E(X) = \mu, \qquad \operatorname{var}(X) = s^2, \qquad \sigma(X) = s.$$

We say that the distribution μ_F determined by F is a *Gaussian* or *normal distribution* with mean μ and variance s^2 and write "μ_F is $N(\mu, s^2)$".

In case $\mu = 0$ and $s = 1$ (i.e. μ_F is $N(0,1)$) then μ_F is called *standard normal distribution.*

Exercise 2.2. Let X be a $N(0,r)$-random variable. Show:

1. $X \in L^p(\Omega)$ for all $1 \le p < \infty$.

2. $E(X^2) = r$ and $E(X^4) = 3r^2$.

3. αX is $N(0, \alpha^2 r)$ for each scalar $\alpha \neq 0$.

Exercise 2.3. For each $k \in \mathbb{N}_o$, let δ_k denote the one-point Dirac measure at k. That is, for all subsets A of $\mathbb{R}$,

$$\delta_k(A) = \begin{cases} 1 & \text{if } k \in A \\ 0 & \text{if } k \notin A. \end{cases}$$

Fix $a > 0$ and set

$$\mu_a(A) = e^{-a} \sum_{k=1}^{\infty} \frac{a^k}{k!} \delta_k(A) \qquad (A \subset \mathbb{R}).$$

Show:

1. μ_a is a Probability measure on $\mathbb{R}$ (called the *Poisson distribution*).

2. The mean and variance of μ_a equal a. (That is, if X is any random variable with distribution μ_a, then $E(X) = \text{var}(X) = a$.)

3. μ_a has no density function.

The following note is for the sake of completeness only; we will not make use of it.

Remark 2.6. (Convergence in distribution) Let $X_n, X : \Omega \to \mathbb{R}^d$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$. We say that X_n *converges to* X *in distribution* and write $X_n \xrightarrow{d} X$ if

$$\int_{\mathbb{R}^d} f \, d\mu_{X_n} \to \int_{\mathbb{R}^d} f \, d\mu_X$$

for all bounded and continuous functions $f : \mathbb{R}^d \to \mathbb{R}$.

Note 1: One can replace "for all bounded and continuous functions f" by all "for all continuous functions f vanishing at infinity" (i.e. $\forall f \in C_o(\mathbb{R}^d)$.)

Observe that

$$\begin{aligned} X_n \xrightarrow{d} X &\Leftrightarrow \int_{\mathbb{R}^d} f \, d\mu_{X_n} \to \int_{\mathbb{R}^d} f \, d\mu_X \\ &\Leftrightarrow \int_{\Omega} f \circ X_n \, dP \to \int_{\Omega} f \circ X \, dP \\ &\Leftrightarrow E\big(f(X_n)\big) \to E\big(f(X)\big) \end{aligned}$$

for all $f : \mathbb{R}^d \to \mathbb{R}$ bounded and continuous (respectively, for all $f : C_o(R^d)$.)

Note 2: One can show that if $X_n \to X$ a.s., or in probability, or in $\|\cdot\|_p$, then $X_n \xrightarrow{d} X$.

Note 3: In advanced books on measure theory (see [3]) one shows that the set $M(\mathbb{R}^d)$ of all finite Borel measures on $\mathbb{R}^d$ is a Banach space, and that $M(\mathbb{R}^d)$ is isometrically isomorphic to the dual space of $C_o(\mathbb{R}^d)$. Hence,

$$X_n \xrightarrow{d} X \Leftrightarrow < f, \mu_{X_n} > \to < f, \mu_X > \ \forall f \in C_o(\mathbb{R}^d) \Leftrightarrow \mu_{X_n} \overset{\text{weak-*}}{\to} \mu_X.$$

2.3 Independence

Independent Events

Let us explain the concept of independence with an example first.

Example 2.9. Consider again our example of throwing a die:

$$\Omega = \{1,2,3,4,5,6\}, \qquad \mathcal{F} = \mathcal{P}(\Omega), \qquad P(A) = \frac{1}{6}\text{card}(A).$$

Let us set

$$E = \{1,2,3\} \qquad \text{(we throw a number} \leq 3)$$
$$A = \{2,4,6\} \qquad \text{(we throw an even number)}$$

Then $P(A) = P(E) = \frac{1}{3}$.

Suppose now that event E holds, i.e. we have thrown a number ≤ 3. The probability that this number is even is now only $\frac{1}{3}$, and not $\frac{1}{2}$! We say the the events E and A are *dependent.* In fact, set

$$P(A|E) := \text{probability of } A \text{ given event } E \text{ has occurred}$$

Also

$$P(A \cap E) = \text{probability that both } A \text{ and } E \text{ has occur.}$$

Then

$$P(A|E) = \frac{P(A \cap E)}{P(E)}. \tag{2.14}$$

provided that $P(E) \neq 0$. For example, in the above

$$P(A|E) = \frac{P(\{2\})}{P(\{1,2,3\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3} \neq P(A) = \frac{1}{2}.$$

Now let

$$F := \{1,2,3,4\} \qquad \text{(we throw a number} \leq 4)$$

The probability that an element of F is even is $\frac{1}{2}$, in fact

$$P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{P(\{2,4\})}{P(\{1,2,3,4\})} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2} = P(A)!$$

In this case, we say that the events A and F are *independent.*

Thus, events A and E are independent iff $P(A|E) = P(A)$ where $P(A|E)$ is defined as in (2.14). Since in 2.14 the denominator must not be zero, we multiply this equation by $P(E)$ and obtain a definition valid for null sets as well:

Definition 2.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Two events $A, E \in \mathcal{F}$ are called *independent* if

$$\boxed{P(A \cap E) = P(A)P(E)}.$$

If this identity does not hold, then A and E are called *dependent.*

For an infinite number of events we have:

Definition 2.6. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A collection $\{E_\alpha\}_{\alpha \in A}$ of events is called *independent*, if for every finite subcollection $\{E_{\alpha_1}, \ldots, E_{\alpha_k}\}$ of *distinct* elements of $\mathcal{F}$ we have

$$P\big(E_{\alpha_1} \cap E_{\alpha_2} \cap \ldots E_{\alpha_k}\big) = P(E_{\alpha_1})P(E_{\alpha_2}) \ldots P(E_{\alpha_k}). \tag{2.15}$$

Exercise 2.4. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and $P\big(\{\omega_i\}\big) = \frac{1}{4}$, $i = 1 \ldots 4$. Find three events A, B, C which are

1. pairwise independent (i.e. any two sets are independent)
2. but not independent.

Independent Random Variables

Return to the example of throwing a die.

Example 2.10. Let us throw the die several times in a row. Let X_n denote the win/loss at the n-th throw, as in example 2.1. Common sense says that the result of any throw should not depend on the other throws. That is, for any n distinct throws $n_1, \ldots, n_k$ and any k sets $B_1, \ldots, B_k \in \mathcal{B}(\mathbb{R})$ we have

$$P\big(X_{n_1} \in B_1,\, X_{n_2} \in B_2, \ldots, X_{n_k} \in B_k\big) = P(X_{n_1} \in B_1)P(X_{n_2} \in B_2) \ldots P(X_{n_k} \in B_k)$$

or equivalently,

$$P\big(X_{n_1}^{-1}(B_1) \cap X_{n_2}^{-1}(B_2) \cap \cdots \cap X_{n_k}^{-1}(B_k)\big) = P\big(X_{n_1}^{-1}(B_1)\big)P\big(X_{n_2}^{-1}(B_2)\big) \ldots P\big(X_{n_k}^{-1}(B_k)\big). \tag{2.16}$$

That is the sets

$$X_{n_1}^{-1}(B_1),\ X_{n_2}^{-1}(B_2), \ldots, X_{n_k}^{-1}(B_k)$$

or in different notation, the sets

$$\{X_{n_1} \in B_1\},\ \{X_{n_2} \in B_2\}, \ldots, \{X_{n_k} \in B_k\}$$

are independent.

Definition 2.7. A collection $\{X_\alpha\}_{\alpha \in A}$ of random variables on a probability space $(\Omega, \mathcal{F}, P)$ is called *independent* if the collection of events

$$\big\{\, X_{\alpha_k} \in B_k \,\big\}_{k=1}^{N} = \big\{\, X_{\alpha_k}^{-1}(B_k) \,\big\}_{k=1}^{N}$$

is independent for all finite subcollections $\big\{\, X_{\alpha_k} \,\big\}_{k=1}^{N}$ and all Borel sets B_k.

Remark 2.7. Usually we will deal with scalar valued random variables as in the above definition. However, this definition makes perfect sense for d-dimensional random variables $X_\alpha : \Omega \to \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$, or even for measurable maps $X_\alpha : \Omega \to (X_\alpha, \mathcal{E}_\alpha)$ where the $(X_\alpha, \mathcal{E}_\alpha)$ are possibly different measurable spaces.

Exercise 2.5. Let $\Omega = [-1,1]$, $\mathcal{F} = \mathcal{B}\big([-1,1]\big)$ and $P = \frac{\lambda}{2}$.

1. Let $X(\omega) = \omega$ and $Y(\omega) = \omega^2$. Show that X and Y are dependent.
2. Let $Z(\omega) = \omega$ be the Heavyside function, $Z(\omega) = \frac{\omega}{|\omega|}$. Show that Y and Z are independent, but X and Z are dependent.

Now let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}^d$ and $Y : \Omega \to \mathbb{R}^m$ be two random variables. Define

$$Z : \Omega \to \mathbb{R}^{d+m} = R^d \times R^m \qquad \text{by} \qquad Z(w) = \big(X(\omega), Y(\omega)\,\big).$$

Claim: Z is a $\mathcal{B}(\mathbb{R}^{d+m})$-measurable random variable.

In fact, first let

$$A \times B \in \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^m) = \mathcal{B}(\mathbb{R}^{d+m})$$

be a measurable rectangle. Then

$$\begin{aligned} Z^{-1}(A \times B) &= \{\omega : X(\omega) \in A \text{ and } Y(\omega) \in B\} \\ &= \{\omega : X(\omega) \in A\} \cap \{\omega : Y(\omega) \in B\} = \underbrace{X^{-1}(A)}_{\in\mathcal{F}} \cap \underbrace{Y^{-1}(B)}_{\in\mathcal{F}} \in \mathcal{F} \end{aligned}$$

since X and Y are measurable. Now as $\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{B}(\mathbb{R}^m)$ is generated by the collection of measurable rectangles, the claim follows from theorem 1.5.

Now consider the following two Borel measures on $\mathbb{R}^{d+m}$:

1. The distribution $\mu_Z = \mu_{(X,Y)}$ determined by $Z = (X,Y)$ called the *joint distribution of* X *and* Y. That is

   $$\mu_{(X,Y)}(E) = P\big(\,(X,Y) \in E\,\big) = P\big(\,\{\omega : (X,Y)(\omega) \in E\}\,\big)$$

   for $E \in \mathcal{B}(\mathbb{R}^{d+m})$. Observe that in case of measurable rectangles $A \times B$,

   $$\begin{aligned} \mu_{(X,Y)}(A \times B) &= P\big(\,\{\omega : X(\omega) \in A \text{ and } Y(\omega) \in B\}\,\big) \\ &= P\big(\,\{\omega : X(\omega) \in A\} \cap \{\omega : Y(\omega) \in B\}\,\big) \\ &= P(\,X \in A,\ Y \in B\,). \end{aligned} \tag{2.17}$$

2. The product $\mu_X \times \mu_Y$ of the distributions (i.e. product measure) μ_X and μ_Y. It is given on measurable reactangles $A \times B$ by

   $$(\mu_X \times \mu_Y)(A \times B) = \mu_X(A)\mu_Y(B) = P(X \in A)\,P(Y \in B). \tag{2.18}$$

It is natural to ask: Do both measures coincide ?

Theorem 2.2. $\mu_{(X,Y)} = \mu_X \times \mu_Y \quad \Leftrightarrow \quad X$ *and* Y *are independent.*

Proof. Recall:

$$
\begin{aligned}
& X \text{ and } Y \text{ are independent} \\
\underset{\text{def.}}{\Leftrightarrow}\quad & \text{the sets } \{X \in A\} \text{ and } \{Y \in B\} \text{ are independent } \forall\, A \in \mathcal{B}(\mathbb{R}^d),\ B \in \mathcal{B}(\mathbb{R}^m) \\
\underset{\text{def.}}{\Leftrightarrow}\quad & P(X \in A,\ Y \in B) = P(X \in A)\, P(Y \in B) \qquad \forall\, A \in \mathcal{B}(\mathbb{R}^d),\ B \in \mathcal{B}(\mathbb{R}^m) \\
\Leftrightarrow\quad & \mu_{(X,Y)}(A \times B) = (\mu_X \times \mu_Y)(A \times B) \qquad \forall\, A \in \mathcal{B}(\mathbb{R}^d),\ B \in \mathcal{B}(\mathbb{R}^m) \\
\underset{\text{thm 1.27}}{\Leftrightarrow}\quad & \mu_{(X,Y)}(E) = (\mu_X \times \mu_Y)(E) \qquad \forall\, E \in \mathcal{B}(\mathbb{R}^{d+m})
\end{aligned}
$$

where we have used (2.17) and (2.18). □

Theorem 2.3. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, and* $X : \Omega \to \mathbb{R}^d$ *and* $Y : \Omega \to \mathbb{R}^m$ *be two* independent *random variables.*

1. *If* $h : \mathbb{R}^{d+m} \to \mathbb{R}$ *is a Borel function with either* $h \geq 0$ *or* $h \circ (X,Y) \in L^1(\Omega)$ *then*
$$E\big[\, h(X,Y) \,\big] = \int_{\mathbb{R}^m} \int_{\mathbb{R}^d} h(x,y)\, d\mu_X(x)\, d\mu_Y(y).$$

2. *If* $f : \mathbb{R}^d \to \mathbb{R}$ *and* $g : \mathbb{R}^m \to \mathbb{R}$ *are Borel functions with either* $f, g \geq 0$ *or* $f \circ X, g \circ Y \in L^1(\Omega)$ *then*
$$E\big[\, f(X) g(Y) \,\big] = E\big[\, f(X) \,\big] E\big[\, g(Y) \,\big]$$

Proof. Observe that the compositions $h(X,Y)$, $f(X)$ and $g(Y)$ are scalar valued random variables.

1. Let h be as given. Then the integral (=expectation) of $h(X,Y)$ is defined, and
$$
\begin{aligned}
E\big[\, h(X,Y) \,\big] &= \int_\Omega h(X,Y)\, dP \\
&= \int_{\mathbb{R}^{d+m}} h(x,y)\, d\mu_{(X,Y)}(x,y) \qquad \text{(by exercise 2.1)} \\
&= \int_{\mathbb{R}^{d+m}} h(x,y)\, d(\mu_X \times \mu_Y)(x,y) \qquad \text{(by theorem 2.2)} \\
&= \int_{\mathbb{R}^m} \int_{\mathbb{R}^d} h(x,y)\, d\mu_X(x)\, d\mu_Y(y).
\end{aligned}
\tag{2.19}
$$
In the last line we use Tonelli's theorem in case $h \geq 0$ and Fubini's theorem in case $h \circ (X,Y) \in L^1(\Omega)$, which is equivalent to $h \in L^1(\mathbb{R}^{d+m}, \mu_X \times \mu_Y)$ by exercise 2.1.

2. Let f and g be as stated and set

$$h(x,y) := f(x)g(y).$$

Then by theorem 1.26, h is Borel measurable. Note that if $f,g \geq 0$ then $h(X,Y) \geq 0$ so that $E\big[\,h(X,Y)\,\big]$ is defined. On the other hand, if $f \circ X, g \circ Y \in L^1(\Omega)$ it is not obvious why $h \circ (X,Y)$ should be in $L^1(\Omega)$. For this, we implicitely use independence of X and Y which was required in the proof of part 1. In fact,

$$\begin{aligned}
\int_\Omega |h(X,Y)|\,dP \underset{\text{part 1.}}{=} & \int_{\mathbb{R}^m}\int_{\mathbb{R}^d} |h(x,y)|\,d\mu_X(x)\,d\mu_Y(y) \\
= & \int_{\mathbb{R}^m}\int_{\mathbb{R}^d} |f(x)|\,|g(y)|\,d\mu_X(x)\,d\mu_Y(y) \\
= & \int_{\mathbb{R}^m} |g(y)| \Big[\int_{\mathbb{R}^d} |f(x)|\,d\mu_X(x)\Big]\,d\mu_Y(y) \qquad (2.20) \\
= & \Big[\int_{\mathbb{R}^d} |f(x)|\,d\mu_X(x)\Big]\Big[\int_{\mathbb{R}^m} |g(y)|\,d\mu_Y(y)\Big] \\
\underset{(2.3)}{=} & \Big[\int_\Omega |f \circ X(\omega)|\,dP\Big]\Big[\int_\Omega |g \circ Y(\omega)|\,dP\Big] < \infty
\end{aligned}$$

by assumption. This shows that $h \circ (X,Y) \in L^1(\Omega)$. Thus, we can compute as in (2.20) but without absolute values,

$$\begin{aligned}
E\big[\,f(X)g(Y)\,\big] &= E\big[\,h(X,Y)\,\big] \\
&= \Big[\int_\Omega f \circ X(\omega)\,dP\Big]\Big[\int_\Omega g \circ Y(\omega)\,dP\Big] = E\big[\,f \circ X\,\big]E\big[\,g \circ Y\,\big].
\end{aligned}$$

This proves the theorem. □

Choosing $f(x) = x$ and $g(y) = y$ we obtain as a special case:

Corollary 2.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X, Y : \Omega \to \mathbb{R}$ be two* independent *scalar random variables with $X, Y \in L^1(\Omega)$. Then $XY \in L^1(\Omega)$ and $E[XY] = E[X]E[Y]$.*

Exercise 2.6. Show that the converse of theorem 2.3 also holds: *If*

$$E\big[\,f(X)g(Y)\,\big] = E\big[\,f(X)\,\big]E\big[\,g(Y)\,\big]$$

for all bounded Borel functions $f,g \geq 0$, then X and Y are independent. (Hint: Given $A \in \mathcal{B}(\mathbb{R}^d)$ and $B \in \mathcal{B}(\mathbb{R}^m)$, consider $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$.)

Exercise 2.7. Prove: *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}^d$ and $Y : \Omega \to \mathbb{R}^m$ be two independent random variables Then for all Borel-measurable functions*

$$f : \mathbb{R}^d \to \mathbb{R}^k, \qquad g : \mathbb{R}^m \to \mathbb{R}^l$$

the random variables

$$f(X) = f \circ X \qquad \text{and} \qquad g(Y) = g \circ Y$$

are also independent.

Independent σ-algebras

Definition 2.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\{\mathcal{F}_\alpha\}_{\alpha \in A}$ a family of σ-subalgebras of $\mathcal{F}$. (That is, each $\mathcal{F}_\alpha \subseteq \mathcal{F}$ and $\mathcal{F}_\alpha$ is itself a σ-algebra.) We say that the family $\{\mathcal{F}_\alpha\}_{\alpha \in A}$ is *independent* if for any finite number of indices $\alpha_1, \dots, \alpha_N$, every collection of sets $\{E_k\}_{k=1}^N$ with $E_k \in \mathcal{F}_{\alpha_k}$ is independent.

Since every σ-algebra contains the space Ω, it is not difficult to obtain the following simple characterization of independent σ-algebras and independent random variables:

Exercise 2.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Show:

1. A finite family $\mathcal{F}_1, \mathcal{F}_2, \dots \mathcal{F}_N$ of σ-subalgebras of $\mathcal{F}$ is independent iff
$$P(E_1 \cap E_2 \cap \ldots \cap E_N) = P(E_1)P(E_2)\cdots P(E_N) \qquad \forall E_k \in \mathcal{F}_k. \tag{2.21}$$
2. A finite family $X_1, X_2, \dots, X_N$ of random variables $X_k : \Omega \to \mathbb{R}^{d_k}$ is independent iff the σ-algebras $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_N)$ are independent.

Example 2.11. Let $(\Omega, \mathcal{F}, P)$ be as in exercise 2.5. Set
$$\mathcal{F}_1 := \{E \in \mathcal{F} : -E = E\},$$
the collection of symmetric Borel sets. Also, set
$$\mathcal{F}_2 := \big\{\, \emptyset,\ [-1.0)\,[0,1],\ [-1,1]\,\big\}.$$
It is an easy exercise to verify that F_1 and $\mathcal{F}_2$ are independent σ-algebras.

Remark 2.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}^d$ and $Y : \Omega \to \mathbb{R}^m$ be two $\mathcal{F}$-measurable random variables. By theorem 1.9,
$$\mathcal{F}_1 := \sigma(X) = \big\{E = X^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^d)\big\} \qquad (\text{"}\sigma\text{-algebra generated by } X\text{"})$$
$$\mathcal{F}_2 := \sigma(Y) = \big\{F = Y^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^m)\big\} \qquad (\text{"}\sigma\text{-algebra generated by } Y\text{"})$$
are σ-subalgebras of $\mathcal{F}$. Then

X and Y are independent
$$\Leftrightarrow\ P\big(X^{-1}(A) \cap Y^{-1}(B)\big) = P\big(X^{-1}(A)\big)P\big(Y^{-1}(B)\big) \quad \forall A \in \mathcal{B}(\mathbb{R}^d),\ B \in \mathcal{B}(\mathbb{R}^m)$$
$$\Leftrightarrow\ P\big(E \cap F\big) = P\big(E\big)P\big(F\big) \qquad \forall E \in \mathcal{F}_1,\ F \in \mathcal{F}_2$$
$\Leftrightarrow$ $\mathcal{F}_1 = \sigma(X)$ and $\mathcal{F}_2 = \sigma(Y)$ are independent.

This motivates the following definition:

Definition 2.9. Let $X : \Omega \to \mathbb{R}^d$ be a random variable on a probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{F}_o$ a σ-subalgebra of $\mathcal{F}$. We say that X and $\mathcal{F}_o$ are *independent*, if $\sigma(X)$ and $\mathcal{F}_o$ are independent σ-algebras.

Exercise 2.9. Keep the above notation. Show:

X and $\mathcal{F}_o$ are independent
$$\Leftrightarrow \quad X \text{ and } \mathbf{1}_A \text{ are independent random variables} \quad \forall A \in \mathcal{F}_o.$$

Example 2.12. (Continuation of example 2.11.) Let $X(\omega) = \mathbf{1}_{[0,1]}$. Then $\sigma(X) = \mathcal{F}_2$ (verify !) and hence X and $\mathcal{F}_1$ are independent.

2.4 Conditional Expectations

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and $\mathcal{F}_o$ a σ-subalgebra of $\mathcal{F}$. In general, a function $f : \Omega \to \mathbb{R}$ which is $\mathcal{F}$-measurable need not be $\mathcal{F}_o$-measurable

Example 2.13. (example 2.11, continued) Let $f(x) = x^2$.

1. Since f is continuous, it is $\mathcal{F}$-measurable, where $\mathcal{F} = \mathcal{B}([-1,1])$ denotes the Borel σ-algebra on $[-1,1]$.

2. However, if we set $\mathcal{F}_2 = \{\emptyset,\ [-1,0),\ [0,1],\ [-1,1]\}$, then f is *not* $\mathcal{F}_2$-measurable. For example

$$\left\{w \in [-1,1] : f(w) < \frac{1}{4}\right\} = \left(-\frac{1}{2}, \frac{1}{2}\right) \notin \mathcal{F}_2.$$

3. Notice that f is $\mathcal{F}_1$-measurable, where $\mathcal{F}_1$ consists of all symmetric Borel sets,

$$\mathcal{F}_1 = \{E \in \mathcal{B}([-1,1]) : -E = E\}.$$

In fact, for all $a \in \mathbb{R}$ we have

$$\{w \in [-1,1] : f(w) < a\} = \begin{cases} \emptyset \in \mathcal{F}_1, & a \leq 0 \\ (-\sqrt{a}, \sqrt{a}) \in \mathcal{F}_1, & 0 < a \leq 1 \\ [-1,1] \in \mathcal{F}_1, & a > 1. \end{cases}$$

It is easy to see that a Borel function $g : [-1,1] \to \mathbb{R}$ is $\mathcal{F}_1$-measurable if and only if it is even.

The idea of conditional expectation is to find an $\mathcal{F}_o$-measurable function g which is "closest" to f.

Definition 2.10. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\mathcal{F}_o$ a σ-subalgebra of $\mathcal{F}$. Furthermore, let $X : \Omega \to \mathbb{R}$ be a $\mathcal{F}$-measurable random variable, $X \in L^1(\Omega, \mathcal{F}, P)$.

The *conditional expectation of* X *given* $\mathcal{F}_o$ is the random variable $Y : \Omega \to \mathbb{R}$ satisfying:

(C1) Y is $\mathcal{F}_o$-measurable.

(C2) $\int_A Y\,dP = \int_A X\,dP \quad \forall A \in F_o$.

We write

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

For this definition to make sense, we must show that Y exists, and is unique.

Theorem 2.5. *$E[X|\mathcal{F}_o]$ exists and is essentially unique. That is, if Y and $\tilde{Y}$ are random variables satisfying (C1) and (C2) above, then $Y = \tilde{Y}$ a.e. Furthermore, $E[X|\mathcal{F}_o] \in L^1(\Omega, \mathcal{F}_o, P)$.*

Proof. 1) Existence of Y.

a) Suppose first that $X \geq 0$. Set

$$\nu(A) := \int_A X\,dP \qquad \forall A \in \mathcal{F}.$$

By assignment 1.13, ν is a finite (as $X \in L^1(\Omega, \mathcal{F}, P)$) measure on $(\Omega, \mathcal{F})$, hence on $(\Omega, \mathcal{F}_o)$, and $\nu \prec P$. Thus by the Radon-Nikodym Theorem, there exists and $\mathcal{F}_o$-measurable random variable $Y : \Omega \to [0, \infty)$ such that

$$\nu(A) = \int_A Y\,dP \qquad \forall A \in \mathcal{F}_o.$$

That is,

$$\int_A X\,dP = \nu(A) = \int_A Y\,dP \qquad \forall A \in \mathcal{F}_o.$$

b) Given an arbitrary $X \in L^1(\Omega, \mathcal{F}, P)$, write $X = X^+ - X^-$. By part a), there exist $\mathcal{F}_o$-measurable random variables $Y^+ . Y^- : \Omega \to [0, \infty)$ such that for all $A \in \mathcal{F}_o$,

$$\int_A X^+\,dP = \int_A Y^+\,dP < \infty \qquad \text{and} \qquad \int_A X^-\,dP = \int_A Y^-\,dP < \infty. \tag{2.22}$$

Set $Y = Y^+ - Y^-$. Then

(a) Y is $\mathcal{F}_o$-measurable.

(b) For all $A \in \mathcal{F}_o$,

$$\int_A X\,dP = \int_A X^+\,dP - \int_A X^-\,dP \overset{(2.22)}{=} \int_A Y^+\,dP - \int_A Y^-\,dP = \int_A Y\,dP.$$

This proves existence. Note that this proof also shows that

$$E[X^+|\mathcal{F}_o] = Y^+ = E[X|\mathcal{F}_o]^+ \qquad \text{and} \qquad E[X^-|\mathcal{F}_o] = Y^- = E[X|\mathcal{F}_o]^-.$$

In addition,

$$\begin{aligned}\int_\Omega |Y|\,dP &= \int_\Omega Y^+\,dP + \int_\Omega Y^-\,dP \\ &\overset{(2.22)}{=} \int_\Omega X^+\,dP + \int_\Omega X^-\,dP = \int_\Omega |X|\,dP < \infty.\end{aligned} \tag{2.23}$$

That is, $E[X|\mathcal{F}_o] \in L^1(\Omega, \mathcal{F}_o, P)$.

2) Uniqueness. Let Y and $\tilde{Y}$ be random variables satisfying (C1) and (C2). For each $n \in \mathbb{N}$, set

$$A_n := \left\{\omega \in \Omega : (Y - \tilde{Y})(\omega) \geq \frac{1}{n}\right\} \in \mathcal{F}_o.$$

(In short, this set can be written as $A_n = \{Y - \tilde{Y} \geq \frac{1}{n}\}$.) Then

$$\begin{aligned} 0 &= \int_{A_n} X\,dP - \int_{A_n} X\,dP \overset{(C2)}{=} \int_{A_n} Y\,dP - \int_{A_n} \tilde{Y}\,dP \\ &= \int_{A_n} (Y - \tilde{Y})\,dP \geq \int_{A_n} \frac{1}{n}\,dP = \frac{1}{n}P(A_n) \end{aligned}$$

which shows that $P(A_n) = 0$ for all n. So if we set

$$A^+ = \left\{w \in \Omega : (Y - \tilde{Y})(\omega) > 0\right\}$$

then

$$A^+ = \bigcup_{n=1}^{\infty} A_n$$

and thus

$$P(A^+) = \sum_{n=1}^{\infty} P(A_n) = 0.$$

By symmetry, if

$$A^- = \left\{w \in \Omega : (Y - \tilde{Y})(\omega) < 0\right\} = \left\{w \in \Omega : (\tilde{Y} - Y)(\omega) > 0\right\}$$

then $P(A^-) = 0$. Thus,

$$P\left(\{\omega \in \Omega : Y(x) \neq \tilde{Y}(x)\}\right) = P(A^+ \cup A^-) = P(A^+) + P(A^-) = 0.$$

That is, $Y = \tilde{Y}$ a.e. □

Example 2.14. Suppose that X is already *$\mathcal{F}_o$-measurable.* Then $Y := X$ satisfies (C1) and (C2) above. That is,

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

Example 2.15. Suppose, X is *independent* of $\mathcal{F}_o$. That is,

$$P\left(X^{-1}(B) \cap A\right) = P\left(X^{-1}(B)\right)\, P(A)$$

for all $A \in \mathcal{F}_o$ and all Borel subsets B of $\mathbb{R}$.

Claim: $E[X|\mathcal{F}_o]$ is the constant function,

$$E[X|\mathcal{F}_o] = E(X).$$

In fact, as X and $\mathbf{1}_A$ are independent by assignment 2.9, we have

i) The constant function $Y = E(X)$ is trivially $\mathcal{F}_o$-measurable, and

ii) For all $A \in \mathcal{F}_o$,

$$\int_A X\,dP = \int_\Omega X\mathbf{1}_A\,dP = E(X\mathbf{1}_A)$$
$$= E(X)\,E(\mathbf{1}_A) = E(X)\int_A 1\,dP = \int_A E(X)\,dP.$$

The claim now follows from (C1) and (C2).

Example 2.16. (Continuation of examples 2.11 and 2.13.) Recall that

$$\Omega = [-1,1], \qquad \mathcal{F} = \mathcal{B}([-1,1]), \qquad P = \frac{1}{2}\lambda$$
$$\mathcal{F}_1 = \{E \in \mathcal{F} : E = -E\}, \qquad \mathcal{F}_2 = \{\emptyset,\ [-1,0),\ [0,1],\ \Omega\}.$$

Using inversion invariance of the Lebesgue measure together with the definition of the Lebesgue integral, it is an easy exercise to check that if f is a Borel measurable function, then for every Borel set E we have:

i) If f is even, then

$$\int_E f\,d\lambda = \int_{-E} f\,d\lambda.$$

ii) If f is odd, then

$$\int_E f\,d\lambda = -\int_{-E} f\,d\lambda.$$

iii) If $\tilde{f}(x) := f(-x)$ then

$$\int_E \tilde{f}\,d\lambda = \int_{-E} f\,d\lambda.$$

Now let $X : [-1,1] \to \mathbb{R}$ be $\mathcal{F}$-measurable. We want to compute $E[X|\mathcal{F}_1]$ and $E[X|\mathcal{F}_2]$.

1. Assume first that X is *even*. We claim that X is $\mathcal{F}_1$-measurable. In fact, as X is even, then for each Borel subset B of $\mathbb{R}$ and each $\omega \in [-1,1]$,

   $$w \in X^{-1}(B) \Leftrightarrow -w \in X^{-1}(B),$$

   that is, $X^{-1}(B)$ is a symmetric subset of $\mathcal{B}([-1,1])$. But since X is $\mathcal{F}$-measurable, then $X^{-1}(B) \in \mathcal{F}$ as well, and hence $X^{-1}(B) \in \mathcal{F}_1$. This proves the claim.

   It follows from the claim and example 2.14 that

   $$E[X|\mathcal{F}_1] = X. \tag{2.24}$$

   On the other hand, since $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent (see example 2.11) and X is $\mathcal{F}_1$-measurable, then X and $\mathcal{F}_2$ are independent. Hence by example 2.14,

   $$E[X|\mathcal{F}_2] = \text{mean of } X = E(X). \tag{2.25}$$

2. Next assume that X is *odd*. Then for every $A \in \mathcal{F}_1$ we have

$$\int_A X(w)\,dP = -\int_{-A} X(w)\,dP \underset{A=-A}{=} -\int_A X(w)\,dP$$

from which we conclude that

$$\int_A X(w)\,dP = 0 = \int_A 0(w)\,dP. \tag{2.26}$$

Since $A \in \mathcal{F}_1$ was arbitrary, it follows that

$$E[X|\mathcal{F}_1] = 0. \tag{2.27}$$

Now observe that every $\mathcal{F}_2$-measurable function Y is of the form

$$Y = a\mathbf{1}_{[-1,0)} + b\mathbf{1}_{[0,1]}$$

for some real constants a and b. Thus, if $Y = E[X|\mathcal{F}_2]$ then

$$\begin{aligned} 0 \underset{(2.26)}{=} \int\limits_{[-1,1]} X\,dP \underset{(C2)}{=} \int\limits_{[-1,1]} Y\,dP \\ = \int\limits_{[-1,1]} \left[a\mathbf{1}_{[-1,0)} + b\mathbf{1}_{[0,1]}\right] \frac{1}{2}d\lambda = \frac{1}{2}\int\limits_{[-1,0)} a\,d\lambda + \frac{1}{2}\int\limits_{[0,1]} b\,d\lambda = \frac{1}{2}(a+b) \end{aligned}$$

so that $a = -b$. On the other hand,

$$\int\limits_{[0,1]} X\,dP \underset{(C2)}{=} \int\limits_{[0,1]} Y\,dP = b\int\limits_{[0,1]} dP = \frac{1}{2}b.$$

That is,

$$b = 2\int\limits_{[0,1]} X\,dP = 2E\left(X\mathbf{1}_{[0,1]}\right)$$

and hence

$$E[X|\mathcal{F}_2] = 2E\left(X\mathbf{1}_{[0,1]}\right)[\mathbf{1}_{[0,1]} - \mathbf{1}_{[-1,0)}]. \tag{2.28}$$

For example, if $X(w) = w$ then $b = 2\int_0^1 w\left(\frac{1}{2}d\lambda\right) = \frac{1}{2}$ so that

$$E[X|\mathcal{F}_2] = \frac{1}{2}\left[\mathbf{1}_{[0,1]} - \mathbf{1}_{[-1,0)}\right].$$

3. Now we split an arbitrary random variable X into its even and odd parts and write $X = X_e + X_o$ where

$$X_e(w) = \frac{X(w) + X(-w)}{2} = \left(\frac{X + \tilde{X}}{2}\right)(w) \qquad \text{(an even function)}$$

$$X_o(w) = \frac{X(w) - X(-w)}{2} = \Big(\frac{X - \tilde{X}}{2}\Big)(w) \qquad \text{(an odd function).}$$

Then by linearity of the conditional expectation (see the following theorem),

$$E[X|\mathcal{F}_1] = E[X_e + X_o|\mathcal{F}_1] = E[X_e|\mathcal{F}_1] + E[X_o|\mathcal{F}_1] = X_e + 0 = X_e$$

while also

$$E[X|\mathcal{F}_2] = E[X_e + X_o|\mathcal{F}_2] = E[X_e|\mathcal{F}_2] + E[X_o|\mathcal{F}_2]. \tag{2.29}$$

Now by (2.25) above,

$$E[X_e|\mathcal{F}_2] = E(X_e) = \int_{[-1,1]} \frac{X + \tilde{X}}{2}\, dP \overset{\text{iii)}}{=} \int_{[-1,1]} X\, dP = E(X)\mathbf{1}_{[-1,1]}$$

and by (2.28),

$$E[X_o|\mathcal{F}_2] = 2E\left(X_o\mathbf{1}_{[0,1]}\right)[\mathbf{1}_{[0,1]} - \mathbf{1}_{[-1,0)}].$$

Note that

$$\begin{aligned} 2E\left(X_o\mathbf{1}_{[0,1]}\right) &= 2\int_{[0,1]} \frac{X - \tilde{X}}{2}\, dP \overset{\text{iii)}}{=} \int_{[0,1]} X\, dP - \int_{[-1,0)} X\, dP \\ &= E\left(X\mathbf{1}_{[0,1]}\right) - E\left(X\mathbf{1}_{[-1,0)}\right) = E\left(X(\mathbf{1}_{[0,1]} - \mathbf{1}_{[-1,0)})\right). \end{aligned}$$

Thus, (2.29) becomes

$$E[X|\mathcal{F}_2] = E(X)\mathbf{1}_{[-1,1]} + E\left(X(\mathbf{1}_{[0,1]} - \mathbf{1}_{[-1,0)})\right)[\mathbf{1}_{[0,1]} - \mathbf{1}_{[-1,0)}]. \tag{2.30}$$

Exercise 2.10. Let $\mathcal{F}_3$ denote the σ-algebra generated by the collection of sets $\{\, [-1,-\frac{1}{2}),\ [-\frac{1}{2},0),\ [0,\frac{1}{2}),\ [\frac{1}{2},1]\,\}$, and let $X(\omega) = \omega$. Find $E[X|\mathcal{F}_3]$.

Theorem 2.6. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{F}_o$ a σ-subalgebra of $\mathcal{F}$, and $X, \tilde{X} : \Omega \to \mathbb{R}$ $\mathcal{F}$-measurable and integrable. Then*

1. *(Linearity) For all $\alpha, \beta \in \mathbb{R}$,*

$$E[\alpha X + \beta\tilde{X}\,|\,\mathcal{F}_o] = \alpha E[X|\mathcal{F}_o] + \beta E[\tilde{X}|\mathcal{F}_o]$$

2. $$E\left(E[X|\mathcal{F}_o]\right) = E(X)$$

3. *If $Z : \Omega \to \mathbb{R}$ is $\mathcal{F}_o$-measurable and $ZX \in L^1(\Omega, \mathcal{F}, P)$ then*

$$E[ZX|\mathcal{F}_o] = ZE[X|\mathcal{F}_o]$$

Proof. 1. Let $Y = E[X|\mathcal{F}_o]$, $\tilde{Y} = E[Y|\mathcal{F}_o]$ and $\alpha, \beta \in \mathbb{R}$. By linearity of the integral we have for all $A \in \mathcal{F}_o$,

$$\int_A (\alpha X + \beta \tilde{X})\, dP = \alpha \int_A X\, dP + \beta \int_A \tilde{X}\, dP$$
$$\underset{(C2)}{=} \alpha \int_A Y\, dP + \beta \int_A \tilde{Y}\, dP = \int_A (\alpha Y + \beta \tilde{Y})\, dP$$

That is,

$$E[\alpha X + \beta \tilde{X}|\mathcal{F}_o] = \alpha Y + \beta \tilde{Y} = \alpha E[X|\mathcal{F}_o] + \beta E[\tilde{X}|\mathcal{F}_o].$$

2. Let $Y = E[X|\mathcal{F}_o]$. Then

$$E(Y) = \int_\Omega Y\, dP \underset{(C2)}{=} \int_\Omega X\, dP = E(X).$$

3. Throughout, let $Y = E[X|\mathcal{F}_o]$. Note that ZY is $\mathcal{F}_o$-measurable, as Z and Y are. We need to show:

$$E[ZX|\mathcal{F}_o] = ZY,$$

that is,

$$\int_A ZX\, dP = \int_A ZY\, dP \tag{2.31}$$

for all $A \in \mathcal{F}_o$.

1. First let $Z = \mathbf{1}_E$ for some $E \in \mathcal{F}_o$ and $X \geq 0$. By the proof of proposition 2.7, part 1), $Y \geq 0$ as well. Then for all $A \in \mathcal{F}_o$,

$$\int_A ZX\, dP = \int_A \mathbf{1}_E X\, dP = \int_{A\cap E} X\, dP \underset{(C2)}{=} \int_{A\cap E} Y\, dP$$
$$= \int_A \mathbf{1}_E Y\, dP = \int_A ZY\, dP,$$

that is, (2.31) holds.

2. Next let Z be simple and non-negative, say

$$Z = \sum_{k=1}^N c_k \mathbf{1}_{E_k} \qquad (E_k \in \mathcal{F}_o,\ c_k \geq 0).$$

while still $X \geq 0$. Then for all $A \in \mathcal{F}_o$,

$$\int_A ZX\, dP = \int_A \Big[\sum_{k=1}^N c_k \mathbf{1}_{E_k}\Big] X\, dP = \sum_{k=1}^N c_k \int_A \mathbf{1}_{E_k} X\, dP$$
$$\underset{(a)}{=} \sum_{k=1}^N c_k \int_A \mathbf{1}_{E_k} Y\, dP = \int_A \Big[\sum_{k=1}^N c_k \mathbf{1}_{E_k}\Big] Y\, dP = \int_A ZY\, dP$$

by linearity of the integral.

3. Now suppose, $Z \geq 0$ and still, $X \geq 0$. Pick an increasing sequence $\{Z_n\} \uparrow$, $Z_n \geq 0$, of $\mathcal{F}_o$-measurable simple functions such that $Z_n(\omega) \to Z(w)$ at all ω. Then for all $\omega \in \Omega$,

$$\begin{array}{lll} Z_n X \geq 0, & \{Z_n X\} \uparrow, & Z_n(\omega)X(\omega) \to Z(\omega)X(\omega) \\ Z_n Y \geq 0, & \{Z_n Y\} \uparrow, & Z_n(\omega)Y(\omega) \to Z(\omega)Y(\omega) \end{array}$$

Hence by the Monotone Convergence Theorem, for all $A \in \mathcal{F}_o$,

$$\int_A ZX\,dP = \lim_{n\to\infty} \int_A Z_n X\,dP \underset{\text{(b)}}{=} \lim_{n\to\infty} \int_A Z_n Y = \int_A ZY\,dP.$$

4. Finally, let X, Z be arbitrary, with $ZX \in L^1(\Omega, \mathcal{F}, P)$. Observe that

$$ZX = (Z^+ - Z^-)(X^+ - X^-) = (Z^+X^+ + Z^-X^-) - (Z^-X^+ + Z^+X^-) \quad (2.32)$$

and similarly,

$$ZY = (Z^+ - Z^-)(Y^+ - Y^-) = (Z^+Y^+ + Z^-Y^-) - (Z^-Y^+ + Z^+Y^-). \quad (2.33)$$

Now by the proof of proposition 2.7, part 1),

$$E[X^+|\mathcal{F}_o] = Y^+ \qquad \text{and} \qquad E[X^-|\mathcal{F}_o] = Y^-.$$

Furthermore, by assumption

$$Z^\pm X^\pm \leq |Z|\,|X| = |ZX| \in L^1,$$

so that $Z^\pm X^\pm \in L^1$ as well. Thus by part (c), for all $\omega \in \Omega$,

$$\int_A Z^\pm X^\pm\,dP = \int_A Z^\pm Y^\pm\,dP < \infty \quad (2.34)$$

and hence,

$$\int_A ZX\,dP \underset{(2.32)}{=} \int_A Z^+X^+\,dP + \int_A Z^-X^-\,dP - \int_A Z^-X^+\,dP - \int_A Z^+X^-\,dP$$
$$\underset{(2.34)}{=} \int_A Z^+Y^+\,dP + \int_A Z^-Y^-\,dP - \int_A Z^-Y^+\,dP - \int_A Z^+Y^-\,dP \underset{(2.33)}{=} \int_A ZY\,dP.$$

This proves part 3. and the proposition.

□

Theorem 2.7. (Transitivity of the conditional expectation) *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\mathcal{F}_2 \subset \mathcal{F}_1 \subset \mathcal{F}$ σ-subalgebras of $\mathcal{F}$. Then for every random variable $X \in L^1(\Omega, \mathcal{F}, P)$,*

$$E[X|\mathcal{F}_2] = E\big[E[X|\mathcal{F}_1] \bigm| \mathcal{F}_2\big].$$

Proof. Let $Y_1 = E[X|\mathcal{F}_1]$, that is,

$$\int_A X\,dP = \int_A Y_1\,dP \tag{2.35}$$

for all $A \in \mathcal{F}_1$. Similarly let $Y_2 = E[X|\mathcal{F}_2]$, that is,

$$\int_A X\,dP = \int_A Y_2\,dP \tag{2.36}$$

for all $A \in \mathcal{F}_2$. We must show that $Y_2 = E[Y_1|\mathcal{F}_2]$, that is,

$$\int_A Y_1\,dP = \int_A Y_2\,dP$$

for all $A \in \mathcal{F}_2$. However this identity follows directly from (2.35) and (2.36). □

The following technical result will be required for the discussion in chapter 3.

Theorem 2.8. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_o$ a σ-subalgebra of $\mathcal{F}$. Furthermore, let $X_1 : \Omega \longrightarrow \mathbb{R}^d$ be $\mathcal{F}_o$-measurable, and $X_2 : \Omega \longrightarrow \mathbb{R}^m$ be $\mathcal{F}$-measurable and independent of X_1. Given a bounded Borel measurable function $H : \mathbb{R}^d \times \mathbb{R}^m \longrightarrow \mathbb{R}$, set*

$$h(x_1) := E\big[\,H(x_1, X_2)\,\big] = \int_\Omega H\big(x_1, X_2(\omega)\,\big)\,dP(\omega), \qquad (x_1 \in \mathbb{R}^d). \tag{2.37}$$

Then

1. *h is Borel measurable, and*
2. *$h(X_1) = E[\,H(X_1, X_2)\,|\,\mathcal{F}_o\,]$.*

Proof. Let us first show that the integral defining h exists. By theorem 1.26, for fixed x_1, the function

$$x_2 \longmapsto H(x_1, x_2)$$

is $\mathcal{B}(\mathbb{R}^m)$-measurable. Hence the composition

$$w \in \Omega \longmapsto H\big(\,x_1, X_2(\omega)\,\big) \tag{2.38}$$

is $\mathcal{F}$-measurable, for all $x_1 \in \mathbb{R}^d$. Now as H is bounded and $P(\Omega) < \infty$, each function (2.38) is integrable. That is, the integral in (2.37) exists, for each $x_1 \in \mathbb{R}^d$.

Next we show that the function h is $\mathcal{B}(\mathbb{R}^d)$-measurable. Let μ_{X_1} and μ_{X_2} denote the distributions of X_1 and X_2, respectively. As H is bounded, and $\mu_{X_1} \times \mu_{X_2}$ a probability measure ($(\mu_{X_1} \times \mu_{X_2})(\mathbb{R}^{d+m}) = \mu_{X_1}(\mathbb{R}^d)\,\mu_{X_2}(\mathbb{R}^m) = 1 \cdot 1 = 1$) it follows that $H(x_1, x_2)$ is $\mu_{X_1} \times \mu_{X_2}$-integrable. Now

$$h(x_1) = \int_\Omega H\big(x_1, X_2(\omega)\,\big)\,dP(\omega) = \int_{\mathbb{R}^m} H(x_1, x_2)\,d\mu_{X_2}(x_2). \tag{2.39}$$

As H is $\mu_{X_1} \times \mu_{X_2}$-integrable, Fubini's theorem says that the right-hand integral exists a.e. x_1 (because H is bounded and μ_{X_2} a finite measure, this integral even exists at every x_1) and defines a Borel-measurable function $h(x_1)$ which is an element $L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu_{X_2})$.

Now we compute $h(X_1)$. Let $Y : \Omega \to \mathbb{R}$ be an arbitrary bounded, $\mathcal{F}_o$-measurable function and consider the random variable

$$Z = (Y, X_1) : \Omega \to \mathbb{R} \times \mathbb{R}^d.$$

By the claim in the discussion before theorem 2.2, Z is $\mathcal{F}_o$-measurable. Note that the function $\varphi : \mathbb{R}^{1+d} \times \mathbb{R}^m \to \mathbb{R}$ given by

$$\varphi\big((y, x_1), x_2 \big) = yH(x_1, x_2) \qquad (y \in \mathbb{R},\ x_1 \in \mathbb{R}^d,\ x_2 \in \mathbb{R}^m)$$

is Borel measurable by example 1.15. Now by assumption, X_2 is independent of $\mathcal{F}_o$, and hence independent of Z. It follows that

$$\mu_{(Z,X_2)} = \mu_Z \times \mu_{X_2}.$$

Since Y and H are bounded, then $YH(X_1, X_2)$ is integrable, and we have

$$\begin{aligned} \int_\Omega YH(X_1, X_2)\, dP &= \int_\Omega \varphi\big((Y, X_1), X_2 \big)\, dP = \int_\Omega \varphi(Z, X_2)\, dP \\ &= \int_{\mathbb{R}^{1+d} \times \mathbb{R}^m} \varphi(z, x_2)\, d\mu_{(Z,X_2)}(z, x_2) \\ &\underset{\text{Fubini}}{=} \int_{\mathbb{R}^{1+d}} \int_{R^m} yH(x_1, x_2)\, d\mu_{X_2}(x_2)\, d\mu_Z(z) \\ &\underset{(2.39)}{=} \int_{\mathbb{R}^{1+d}} yh(x_1)\, d\mu_Z(z) = \int_\Omega Yh(X_1)\, dP. \end{aligned} \tag{2.40}$$

Now given $A \in \mathcal{F}_o$, choose $Y = \mathbf{1}_A$. Then the above becomes

$$\int_A H(X_1, X_2)\, dP = \int_\Omega \mathbf{1}_A H(X_1, X_2)\, dP = \int_\Omega \mathbf{1}_A h(X_1)\, dP = \int_A h(X_1)\, dP,$$

which shows since $h(X_1)$ is $\mathcal{F}_o$-measurable that

$$E\big[H(X_1, X_2) \,\big|\, \mathcal{F}_o \big] = h(X_1).$$

This proves the proposition. □

Chapter 3

Stochastic Processes and Brownian Motion

We are now ready to introduce the concept of stochastic processes and discuss some of their properties. A process of particular interest is Brownian motion, and we will discuss some of its peculiar features. Throughout this chapter, $(\Omega, \mathcal{F}, P)$ will denote a fixed probability space.

3.1 Stochastic Processes

Definition 3.1. Let $I = [a, b]$ or $I = [a, \infty)$. A family $\{X_t\}_{t \in I}$ of $\mathcal{F}$-measurable random variables

$$X_t : \Omega \to \mathbb{R}^d$$

is called a (d-dimensional) *stochastic process.*

Remark 3.1. Given $t \in I$ and $\omega \in \Omega$, we also set

$$X(t, w) = X(t)(w) := X_t(w).$$

We thus can look at a stochastic process in several ways:

1. For each fixed $t \in I$, the map

$$\omega \in \Omega \mapsto X_t(\omega)$$

is a random variable, as in the definition.

2. For each fixed $\omega \in \Omega$, the map

$$t \in I \mapsto X_t(\omega)$$

is a function defined on I, called a *sample path.*

3. The map

$$(t, w) \mapsto X_t(\omega) = X(t, \omega)$$

is a function defined on $I \times \Omega$.

4. The map
$$t \in I \mapsto X(t) := X_t$$
is a vector valued function defined on I, with values in the linear space of random variables on $(\Omega, \mathcal{F}, P)$.

Throughout, we will use various notations to denote a stochastic process, such as
$$\{X(t) : t \in I\}, \quad \{X(t) : t \geq 0\}, \quad \{X_t\}_{a \leq t \leq b}, \quad \{X_t\}_{t \in I}$$
for example. For simplicity, we will mostly choose $I = [0, \infty)$ and write $\{X(t)\}_{t \geq 0}$.

In practice, we identify "similar" processes:

Definition 3.2. Two stochastic processes $\{X_t\}_{t\in I}$ and $\{Y_t\}_{t\in I}$ are said to be *modifications* (or *versions*) of each other, if for every $t \in I$,
$$X_t(\omega) = Y_t(\omega) \quad \text{a.s.}$$

Remark 3.2. Observe the subtle difference between the following two concepts:

1. "$\{X_t\}_{t\in I}$ and $\{Y_t\}_{t\in I}$ are versions of each other" means that for each $t \in I$,
$$A_t := \{\omega : X_t(\omega) \neq Y_t(\omega)\}$$
is a null set.

2. "$\{X_t\}_{t\in I}$ and $\{Y_t\}_{t\in I}$ are equal a.s." means that
$$A := \{\omega : \exists t \in I, \ X_t(\omega) \neq Y_t(\omega)\}$$
is a null set.

Obviously, 2. implies 1.

Remark 3.3. Suppose, $\{X_t\}_{t\in I}$ and $\{Y_t\}_{t\in I}$ are modifications of each other. For each $t \in I$, let μ_t and ν_t denote the distributions of the random variables X_t, respectively Y_t. We claim that $\mu_t = \nu_t$.

In fact, for all $B \in \mathcal{B}(\mathbb{R}^d)$,
$$\mu_t(B) = P(X_t \in B) = P\big(X_t^{-1}(B)\big) \tag{3.1}$$
while also
$$\nu_t(B) = P(Y_t \in B) = P\big(Y_t^{-1}(B)\big). \tag{3.2}$$
Let $A := \{\omega : X_t(\omega) \neq Y_t(\omega)\}$. Then $A \in \mathcal{F}$ and $P(A) = 0$. Now
$$X_t^{-1}(B) = \underbrace{\{\omega \in A : X_t(\omega) \in B\}}_{\text{a null set}} \cup \{\omega \in A^c : X_t(\omega) \in B\} \quad \text{and}$$
$$Y_t^{-1}(B) = \underbrace{\{\omega \in A : Y_t(\omega) \in B\}}_{\text{a null set}} \cup \{\omega \in A^c : Y_t(\omega) \in B\}$$

Now by choice of A, the two sets on the right coincide. Hence, $X_t^{-1}(B)$ and $Y_t^{-1}(B)$ differ only by a null set. It thus follows from (3.1) and (3.2) that
$$\mu_t(B) = P\big(X_t^{-1}(B)\big) = P\big(Y_t^{-1}(B)\big) = \nu_t(B).$$

Example 3.1. Let φ be a $N(0,1)$ random variable. Let us first show the following:

Claim: For each $k \in \mathbb{R}$, $A_k := \{\omega \in \Omega : \varphi(\omega) = k\}$ is a null set.

In fact, observe that

$$P(A_k) = P(\varphi \in \{k\}) = \mu_\varphi(\{k\}).$$

Now if F is the distribution function of φ,

$$F(t) = \int_{-\infty}^{t} f(x)\,dx \qquad \text{with} \qquad f(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$$

then

$$\begin{aligned} \mu_\varphi(\{k\}) &= \mu_\varphi\Big(\bigcap_{m=1}^{\infty}(k-\frac{1}{n},k]\Big) \underset{\text{thm1.3}}{=} \lim_{n\to\infty}\mu_\varphi\Big((k-\frac{1}{n},k]\Big) \\ &= \lim_{n\to\infty}\Big(F(k)-F(k-\frac{1}{n})\Big) = \lim_{n\to\infty}\int_{k-\frac{1}{n}}^{k} f(x)\,dx = 0 \end{aligned}$$

by continuity of the integral. This proves the claim. (Note that this argument applies to *any* random variable φ having a density function, because then $\mu_\varphi \prec \lambda$.)

Now set

$$X_t(\omega) = t - \varphi(\omega)^2 \qquad (t \geq 0)$$

and

$$Y_t(\omega) = \begin{cases} t - \varphi(\omega)^2 & \text{if } t \neq \varphi(\omega)^2 \\ t & \text{if } t = \varphi(\omega)^2. \end{cases}$$

Claim: X_t and Y_t are $\mathcal{F}$-measurable random variables $\forall\, t \geq 0$.

In fact, since constant functions and φ^2 are measurable, this is obvious for X_t. Now compare Y_t with X_t. Obviously, $Y_0 = X_0$. On the other hand, if $t > 0$ we set

$$\begin{aligned} B_t &= \{\omega \in \Omega : X_t(\omega) \neq Y_t(\omega)\} = \{\omega \in \Omega : \varphi(\omega)^2 = t\} \\ &= \{\omega \in \Omega : \varphi(\omega) = -\sqrt{t}\} \cup \{\omega \in \Omega : \varphi(\omega) = \sqrt{t}\} = A_{-\sqrt{t}} \cup A_{\sqrt{t}}. \end{aligned}$$

(Written in short form, $B_t = \{X_t \neq Y_t\} = \{\varphi^2 \neq t\} = A_{-\sqrt{t}} \cup A_{\sqrt{t}}$.) Then B_t is a null set by the above claim. It follows that

1. $X_t(\omega) = Y_t(\omega)$ a.s.

2. For all $a \in \mathbb{R}$,

$$\{w : Y_t(\omega) \leq a\} = \begin{cases} (B_t)^c \cap \{w : X_t(\omega) \leq a\} \in \mathcal{F} & \text{if } a < t \\ \Big((B_t)^c \cap \{w : X_t(\omega) \leq a\}\Big) \cup A_{-\sqrt{t}} \cup A_{\sqrt{t}} \in \mathcal{F} & \text{if } a \geq t \end{cases}$$

which shows that Y_t is $\mathcal{F}$-measurable.

This proves the claim and furthermore, that $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ are modifications of each other.

For ease of notation, we will from now on assume that $I = [0, \infty)$, although the discussion can easily be applied to arbitrary intervals I. There are several choices to make a stochastic processes compatible with the topological structure on I:

Definition 3.3. A stochastic process $\{X_t\}_{t\geq 0}$ is called

1. *left continuous*, if

$$\lim_{s\to t^+} X_s(\omega) = X_t(\omega) \qquad \forall\, t > 0,\ \forall \omega \in \Omega$$

2. *right continuous*, if

$$\lim_{s\to t^-} X_s(\omega) = X_t(\omega) \qquad \forall\, t \geq 0,\ \forall \omega \in \Omega$$

3. *continuous*, if it is both left and right continuous.

An alternative and clearer definition would be to say that "sample paths are (left/right) continuous".

Example 3.2. (Example 3.1, continued)

1. Let $\omega \in \Omega$ be fixed. The sample path $t \to X_t(\omega) = t - \varphi(\omega)^2$ is a linear function, hence is continuous on $[0, \infty)$. That is, $\{X_t\}_{t\geq 0}$ is continuous.

2. Next pick ω such that $\varphi(\omega) \neq 0$ and let $t = \varphi(\omega)^2$, so that $Y_t(\omega) = t$. Then

$$\lim_{s\to t^+} Y_s(\omega) = \lim_{s\to t^+} s - \varphi(\omega)^2 = t - \varphi(\omega)^2 \neq t = Y_t(\omega)$$

which shows that $\{Y_t\}_{t\geq 0}$ is not (left) continuous at t. Observe, however, that $\{X_t\}_{t\geq 0}$ is a *continuous version* of $\{Y_t\}_{t\geq 0}$.

3. As mentioned above, $\varphi(\omega) \neq 0$ for almost all ω. The argument in 2. shows that for such ω there exists t such that $s \mapsto Y_s(\omega)$ is not continuous at t. Hence there cannot exist a continuous process $\{Z_t\}_{t\geq 0}$ such that $Y_t = Z_t$ a.s.

Let $\{X_t\}_{t\geq 0}$ be a stochastic process. If $s \neq t$, then the σ-algebras generated by X_s and X_t may be different. However, we want the events at time $s < t$ to be also available at time t. We therefore introduce the concept of filtration:

Definition 3.4. A family $\{\mathcal{F}_t\}_{t\geq 0}$ of σ-algebras is called a *filtration* if

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \qquad \text{for all } 0 \leq s < t.$$

Definition 3.5. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration. A stochastic process $\{X_t\}_{t\geq 0}$ is called $(\mathcal{F}_t)$-*adapted* if for each $t \geq 0$, the random variable X_t is $\mathcal{F}_t$-measurable.

Remark 3.4. Let $\{X_t\}_{t\geq 0}$ be $(\mathcal{F}_t)$-adapted. Since X_t is $\mathcal{F}_t$-measurable and $\mathcal{F}_t \subseteq \mathcal{F}_s$ for $s > t$, it follows that X_t is $\mathcal{F}_s$-measurable for all $s > t$ as well. However, X_t need not be $\mathcal{F}_s$-measurable for $s < t$.

Example 3.3. Let $\{X_t\}_{t\geq 0}$ be a stochastic process. Recall that for each t,

$$\sigma(X_t) \underset{\text{thm 1.6}}{=} \{X_t^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d)\},$$

is the smallest σ-algebra in which X_t is measurable. Now for each t, set

$$\mathcal{F}_t^X := \sigma\big(\{X_s : 0 \leq s \leq t\}\big) = \sigma\big(\{X_s^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d),\ 0 \leq s \leq t\}\big) \subseteq \mathcal{F},$$

the σ-algebra generated by the collection of random variables $\{X_s\}$ with $0 \leq s \leq t$. By this definition,

1. $\mathcal{F}_r^X \subseteq \mathcal{F}_t^X$ for $r < t$, i.e. $\{F_t^X\}_{t\geq 0}$ is a filtration, and
2. $\{X_t\}_{t\geq 0}$ is $(\mathcal{F}_t^X)$-adapted, since for each t, $\sigma(X_t) \subseteq \mathcal{F}_t^X$.

We call $\{F_t^X\}_{t\geq 0}$ the *natural filtration* of $\{X_t\}_{t\geq 0}$.

Now suppose $\{X_t\}_{t\geq 0}$ is adapted to another filtration $\{F_t\}_{t\geq 0}$.

1. Since X_t is $\mathcal{F}_t$-measurable, then

   $$X_t^{-1}(B) \in \mathcal{F}_t \qquad \forall\, B \in \mathcal{B}(\mathbb{R}^d)$$

   and thus $\sigma(X_t) = \{X_t^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d)\} \subseteq \mathcal{F}_t$ for each t.

2. As $\{\mathcal{F}_t\}$ is a filtration, then by 1.,

   $$\sigma(X_s) \subseteq \mathcal{F}_s \subseteq F_t \qquad \forall\, 0 \leq s \leq t$$

   and hence

   $$\mathcal{F}_t^X = \sigma\big(\{X_s : 0 \leq s \leq t\}\big) \subseteq \mathcal{F}_t$$

   for all $t \geq 0$.

Thus, the natural filtration is the *smallest* filtration to which $\{X_t\}_{t\geq 0}$ is adapted.

Often it is not enough to consider X_t to be $\mathcal{F}_t$ measurable for each fixed t. Instead, we want the function $X(t,\omega)$ of two variables to be measurable:

Definition 3.6. Let $\{F_t\}_{t\geq 0}$ be a filtration. A stochastic process $\{X_t\}_{t\geq 0}$ is called (F_t)-*progressive*, if for every $T \geq 0$, the function

$$X(t,\omega) : [0,T] \times \Omega \to \mathbb{R}^d$$

is $\mathcal{B}\big([0,T]\big) \otimes \mathcal{F}_T$-measurable.

For simplicity of presentation, above we chose a process defined on $[0,\infty)$. It is obvious how to define the notions of "$\mathcal{F}_t$-adapted" and "progressive" to processes $\{X_t\}_{a\leq t\leq b}$ defined on an arbitrary closed and bounded interval.

Remark 3.5. If $\{X_t\}_{t\geq 0}$ is (F_t)-progressive, then its is also (F_t)-adapted.

To see this, fix an arbitrary $T\geq 0$. Since $X(t,\omega)$ is $\mathcal{B}([0,T])\otimes\mathcal{F}_T$-measurable, then in particular, by theorem 1.26, the restriction to one variable, $X_T:\omega\mapsto X(T,\omega)$ is $\mathcal{F}_T$-measurable. Changing T to t it follows that $\{X_t\}_{t\geq 0}$ is (F_t)-adapted.

For the converse we have:

Theorem 3.1. *Let $\{X_t\}_{t\geq 0}$ be left-continuous and adapted to the filtration $\{F_t\}_{t\geq 0}$. Then $\{X_t\}_{t\geq 0}$ is progressive.*

Proof. Let $T\geq 0$ be arbitrary. If $T=0$, then $[0,T]\times\Omega$ can be identified with Ω and $\mathcal{B}([0,T])\otimes\mathcal{F}_T$ with $\mathcal{F}_0$ as

$$\mathcal{B}([0,T])\otimes\mathcal{F}_T=\mathcal{B}(\{0\})\otimes\mathcal{F}_0=\{\,\{0\}\times A : A\in\mathcal{F}_0\,\},$$

and $X(t,\omega):\{0\}\times\Omega\to\mathbb{R}^d$ with $X_0(\omega)$. Since $X_0(\omega)$ is $\mathcal{F}_0$-measurable, then $X(t,\omega)$ is $\mathcal{B}(\{0\})\otimes\mathcal{F}_0$-measurable.

We thus may assume that $T>0$. We now construct a sequence $X^n(t,w)$ of $\mathcal{B}([0,T])\otimes\mathcal{F}_T$-measurable functions such that

$$X^n(t,w)\to X(t,w) \tag{3.3}$$

pointwise. It then will follow from theorem 1.11 that X is also $\mathcal{B}([0,T])\otimes\mathcal{F}_T$-measurable.

For each $n\in\mathbb{N}$, partition $[0,T]$ into subintervals of length $\frac{T}{n}$ by choosing partitions

$$P_n=\Big\{0,\frac{T}{n},\frac{T}{n},\ldots,\frac{jT}{n},\ldots,\frac{nT}{n}\Big\}.$$

Now let $t\in(0,T]$ be given. For each n there exists a unique $j=j(t,n)$, $0\leq j<n$, such that $t\in\big(\frac{jT}{n},\frac{(j+1)T}{n}\big]$. Let us set

$$p(t,n):=\frac{j(t,n)T}{n}.$$

Then

$$t-\frac{T}{n}\leq p(t,n)<t$$

for all n, so that

$$p(t,n)\to t^- \qquad \text{as } n\to\infty.$$

Now set

$$X^n(t,\omega):=X_0(\omega)\mathbf{1}_{\{0\}}(t)+\sum_{j=0}^{n-1}X_{\frac{jT}{n}}(\omega)\mathbf{1}_{\left(\frac{jT}{n},\frac{(j+1)T}{n}\right]}(t). \tag{3.4}$$

We note:

i) For given $t > 0$, $X^n(t, w) = X_{p(t,n)}(\omega)$.

ii) Since $\{X_t\}$ is $(\mathcal{F}_t)$-adapted, each term is $\mathcal{B}[0,T] \otimes \mathcal{F}_T$-measurable by example 1.15, hence so is the finite sum $X^n(t, \omega)$.

Let $n \to \infty$. If $t = 0$, then $X^n(t, \omega) = X^n(0, \omega) = X_0(\omega) = X(0, \omega)$ so trivially, $X^n(0, \omega) \to X(0, \Omega)$. If $0 < t \leq T$, then

$$X^n(t, \omega) = X_{p(t,n)}(\omega) \to X_t(\omega) \qquad \forall \omega$$

since $p(t, n) \to t^-$ and X_t is *left* continuous. This proves (3.3) and the theorem. □

3.2 Brownian Motion

Throughout this section, $\{\mathcal{F}_t\}_{t \geq 0}$ will denote a filtration on $(\Omega, \mathcal{F}, P)$.

Definition 3.7. An $(\mathcal{F}_t)$-adapted process $\{B_t\}_{t \geq 0}$, $B_t : \Omega \to \mathbb{R}$, is called $(\mathcal{F}_t)$-*Brownian motion* if

(B1) sample paths $t \to B_t(\omega)$ are continuous $\forall \omega \in \Omega$,

(B2) for every $0 \leq r < t$, the random variable

$$B_t - B_r$$

is independent of $\mathcal{F}_r$,

(B3) for every $0 \leq r < t$, the random variable

$$B_t - B_r$$

is Gaussian with mean zero and variance $t - r$.

If in addition,

(B4) $B_0 = 0$, then $\{B_t\}_{t \geq 0}$ is called *standard* Brownian motion.

Remark 3.6. 1. Let $0 \leq r < t$. Then

$$\Delta B := B_t - B_r$$

is called an *increment*. (B3) says that increments have normal distribution $N(0, t - r)$. In particular,

(a) $\Delta B = B_t - B_r \in L^p(\Omega)$ for all $1 \leq p < \infty$ by exercise 2.2.

(b) $E[B_t - B_r] = E[\Delta B] = 0$ and hence

$$E\big[(B_t - B_r)^2\big] = E\big((\Delta B)^2\big] = E\big[(\Delta B)^2\big] - E[\Delta B]^2 = \text{var}[\Delta B] = t - r$$

or in short,

$$E\big[(\Delta B)^2\big] = \Delta t$$

where $\Delta t = t - r$.

(c) By (B2), example 2.14 and (b) we have that

$$E[\Delta B|\mathcal{F}_r] = E[B_t - B_r|\mathcal{F}_r] = E(B_t - B_r) = 0.$$

In applications, this has the following meaning: If B_t denotes the amount of money in a game, then $E[B_t - B_r|\mathcal{F}_r]$, represents the winnings over the time interval $[r, t]$ as predicted at time r. The above equation says that the predicted winnings are zero. This is called a "fair game".)

2. If $\{B_t\}_{t\geq 0}$ is *standard* Brownian motion, then choosing $r = 0$ we see that B_t is $N(0, t)$ for all $t > 0$.

3. Recall that the collection of events $A \in \mathcal{F}_s$ is interpreted as containing all information at time s. So if $0 \leq s \leq r < t$, then (B2) implies that $\Delta B = B_t - B_r$ is independent of the information at time s which is interpreted as "increments are independent of the past".

4. One can show that there exists a great variety of Brownian motions.

5. By theorem 3.1, Brownian motion is progressive.

Exercise 3.1. Let $\{B_t\}_{t\geq 0}$ be $(\mathcal{F}_t)$-adapted Brownian motion

1. For each $t \geq 0$ set $\tilde{B}_t := B_t - B_0$. Show that $\{\tilde{B}_t\}_{t\geq 0}$ is *standard* Brownian motion.

2. (Left shift of Brownian motion.) Fix $t_o > 0$ and for each $t \geq 0$, set $\hat{\mathcal{F}}_t := \mathcal{F}_{t+t_o}$ and set $\hat{B}_t := B_{t_o+t}$. Show that $\{\hat{\mathcal{F}}\}_{t\geq 0}$ is a filtration, and $\{\hat{B}_t\}_{t\geq 0}$ is $(\hat{\mathcal{F}}_t)$-adapted Brownian motion.

Theorem 3.2. *Let $\{B_t\}_{t\geq 0}$ be $(\mathcal{F}_t)$-adapted Brownian motion, and $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$. Then the random variables*

$$B_{t_0}, \quad B_{t_1} - B_{t_0}, \quad B_{t_2} - B_{t_1}, \ldots, \quad B_{t_n} - B_{t_{n-1}}$$

are independent. (We say that "Brownian motion has independent increments".)

Proof. We use exercise 2.8 and induction on n.

Induction start ($n = 1$). Since $\{B_t\}_{t\geq 0}$ is $(\mathcal{F}_t)$-adapted, then $\sigma(B_{t_0}) \subseteq \mathcal{F}_{t_0}$. Now by property (B2), $\sigma(B_{t_1} - B_{t_0})$ is independent of $\mathcal{F}_{t_0}$. It follows that $\sigma(B_{t_0})$ and $\sigma(B_{t_1} - B_{t_0})$ are independent σ-algebras, that is, $B_{t_1} - B_{t_0}$ and B_{t_0} are independent random variables by exercise 2.8,

Induction step. Suppose the assertion holds for some $k \geq 1$. That is (using exercise 2.8),

$$\sigma(B_{t_0}), \quad \sigma(B_{t_1} - B_{t_0}), \quad \sigma(B_{t_2} - B_{t_1}), \ldots, \sigma(B_{t_k} - B_{t_{k-1}}) \tag{3.5}$$

are independent σ-algebras. We need to show that

$$\sigma(B_{t_0}), \quad \sigma(B_{t_1} - B_{t_0}), \quad \sigma(B_{t_2} - B_{t_1}), \ldots, \sigma(B_{t_k} - B_{t_{k-1}}), \quad \sigma(B_{t_{k+1}} - B_{t_k}) \tag{3.6}$$

are independent σ-algebras. For this, let

$$E_0 \in \sigma(B_{t_0}), \qquad E_i \in \sigma(B_{t_i} - B_{t_{i-1}}) \qquad\qquad (1 \le i \le k)$$

which are independent by induction assumption (3.5), and let $E_{k+1} \in \sigma(B_{t_{k+1}} - B_{t_k})$. Now as $\{B_t\}_{t\ge 0}$ is $(\mathcal{F}_t)$-adapted and $\{\mathcal{F}_t\}_{t\ge 0}$ is a filtration, we have that

$$A := E_0 \cap E_1 \cap \cdots \cap E_k \in \mathcal{F}_{t_k}.$$

On the other hand, $\mathcal{F}_{t_k}$ and $\sigma(B_{t_{k+1}} - B_{t_k})$ are independent by (B2). Hence,

$$\begin{aligned} P(E_0 \cap E_1 \cap \cdots \cap E_k \cap E_{k+1}) &= P(A \cap E_{k+1}) = P(A)P(E_{k+1}) \\ &\underset{(3.6)}{=} P(E_0)P(E_1)P(E_2)\dots P(E_k)P(E_{k+1}). \end{aligned}$$

Since $E_0, E_1, \dots, E_{k+1}$ were arbitrary it follows that the σ-algebras in (3.6) are independent. The theorem now follows by induction. $\square$

Notation: In many of the proofs which follow we will make use of the following notation:

Let $t > 0$ be fixed, and let $\{P_n\}_{n=1}^{\infty}$ be a sequence of partitions of $[0, t]$ such that

$$\|P_n\| \to 0 \qquad \text{as} \qquad t \to \infty.$$

Recall here that if we list the partition points of each partition as

$$P_n = \{0 = t_0^{(n)} < t_1^{(n)} < t_2^{(n)} < \cdots < t_{m_n}^{(n)} = t\}$$

(so the n-th partition has $m_n + 1$ partition points) then

$$\|P_n\| = \max_{0 \le j \le m_n - 1} \left[t_{j+1}^{(n)} - t_j^{(n)} \right].$$

When working with a given partition we will often drop the symbol (n) for ease of notation; for example,

$$\Delta_j := t_{j+1} - t_j \quad \text{will be used instead of} \quad \Delta_j^{(n)} := t_{j+1}^{(n)} - t_j^{(n)}.$$

Similarly,

$$\Delta B_j := B_{t_{j+1}} - B_{t_j} \quad \text{will be used instead of} \quad \Delta B_j^{(n)} := B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}.$$

Theorem 3.3. *Let $\{B_t\}_{t\ge 0}$ be $(\mathcal{F}_t)$-Brownian motion. Given $t > 0$, let $\{P_n\}_{n=1}^{\infty}$ be a sequence of partitions of $[0, t]$ with $\|P_n\| \to 0$ as $n \to \infty$. For each n, set*

$$S_n(\omega) := \sum_{j=1}^{m_n - 1} \left[B_{t_{j+1}^{(n)}}(\omega) - B_{t_j^{(n)}}(\omega) \right]^2.$$

Then $S_n(\omega) \xrightarrow{\|\cdot\|_2} t$ *in* $L^2(\Omega)$ *as* $n \to \infty$.

Proof. After dropping the index (n) we need to show that

$$S_n = \sum_j \left(\Delta B_j\right)^2 \to t \qquad \text{in } L^2(\Omega).$$

Now by (B3), ΔB_j has distribution $N(0, \Delta t_j)$, hence by exercise 2.2,

$$E\left[(\Delta B_j)^2\right] = \Delta t_j \qquad \text{and} \qquad E\left[(\Delta B_j)^4\right] = 3(\Delta t_j)^2.$$

Also, by theorem 3.2, ΔB_i and ΔB_j are independent for $i \neq j$. Hence by exercise 2.7 $(\Delta B_i)^2$ and $(\Delta B_j)^2$ are also independent for $i \neq j$. Thus,

$$\begin{aligned}
\|S_n - t\|_2^2 &= \int_\Omega \left(S_n(\omega) - t\right)^2 dP = E\left[(S_n - t)^2\right] = E\left[S_n^2 - 2tS_n + t^2\right] \\
&= E\left[\left(\sum_i (\Delta B_i)^2\right)\left(\sum_j (\Delta B_j)^2\right) - 2t\sum_j (\Delta B_j)^2 + t^2\right] \\
&= E\left[\sum_i (\Delta B_i)^4 + \sum_{i\neq j} (\Delta B_i)^2 (\Delta B_j)^2 - 2t\sum_j (\Delta B_j)^2 + t^2\right] \\
&= \sum_i E[(\Delta B_i)^4] + \sum_{i\neq j} E[(\Delta B_i)^2(\Delta B_j)^2] - 2t\sum_j E[(\Delta B_j)^2] + E[t^2] \\
&= \sum_i 3(\Delta t_i)^2 + \sum_{i\neq j} E[(\Delta B_i)^2(\Delta B_j)^2] - 2t\sum_j \Delta t_j + t^2
\end{aligned}$$

where we have used linearity of the integral. Now applying corollary 2.4,

$$\begin{aligned}
\|S_n - t\|_2^2 &= 3\sum_i (\Delta t_i)^2 + \sum_{i\neq j} E[(\Delta B_i)^2]\, E[(\Delta B_j)^2] - 2t\cdot t + t^2 \\
&= 3\sum_i (\Delta t_i)^2 + \sum_{i\neq j} \Delta t_i\, \Delta t_j - 2t^2 + t^2 \\
&= 2\sum_i (\Delta t_i)^2 + \left(\sum_i \Delta t_i\right)\left(\sum_j \Delta t_j\right) - t^2 \\
&= 2\sum_i (\Delta t_i)^2 + t\cdot t - t^2 \\
&\leq 2\|P_n\| \sum_i \Delta t_i = 2\|P_n\|\, t \to 0 \quad \text{as} \quad n \to \infty.
\end{aligned}$$

This proves the theorem. $\square$

Remark 3.7. A function F is said to be of *bounded variation* on $[a, b]$, if there exists $M > 0$ such that

$$\sum_{j=0}^{m-1} \left| F(t_{j+1}) - F(t_j) \right| < M \tag{3.7}$$

for *all* partitions $P = \{a = t_0 < t_1 < t_2 < \cdots < t_m = b\}$ of $[a, b]$. (Here m is arbitrary ! Intuitively, the graph of F has finite length.)

Claim: Every $F \in C^1[a,b]$ is of bounded variation. For let

$$K := \max_{a \le t \le b} \left| F'(t) \right|.$$

Then for any partition P of $[a,b]$,

$$\sum_{j=0}^{m-1} \left| F(t_{j+1}) - F(t_j) \right| \le \sum_{j=0}^{m-1} \left| K\,(t_{j+1} - t_j) \right| = K(b-a) < \infty$$

by the Mean Value Theorem.

Claim: If F is of bounded variation on $[a,b]$, then it is bounded.

In fact, let M be as in (3.7). For each $t \in [a,b]$ we have

$$\left|F(t)\right| \le \left| F(t) - F(a) \right| + \left|F(a)\right| \underset{(3.7)}{\le} M + \left|F(a)\right|$$

Claim: Let F be continuous and of bounded variation on $[a,b]$. If $\{P_n\}_{n=1}^{\infty}$ is a sequence of partitions of $[0,t]$ with $\|P_n\| \to 0$ as $n \to \infty$ and if

$$S_n := \sum_{j=1}^{m_n-1} \left[F(t_{j+1}^{(n)}) - F(t_j^{(n)}) \right]^2,$$

then $S_n \to 0$ as $n \to \infty$.

In fact, let $\epsilon > 0$ be arbitrary, but given. Since F is uniformly continuous, there exists $\delta > 0$ such that

$$\left|F(t) - F(\tilde{t})\right| < \frac{\epsilon}{M} \qquad \text{whenever} \qquad |t - \tilde{t}| < \delta, \qquad t, \tilde{t} \in [a,b]$$

where M is the constant in (3.7). Thus if n is sufficiently large, we have

$$\left|F(t_{j+1}^{(n)}) - F(t_j^{(n)})\right| < \frac{\epsilon}{M}$$

for all partition points $t_j^{(n)}$ since$\|P_n\| \to 0$. It follows that

$$\begin{aligned} |S_n| &= \sum_{j=1}^{m_n-1} \left|F(t_{j+1}^{(n)}) - F(t_j^{(n)})\right| \cdot \left|F(t_{j+1}^{(n)}) - F(t_j^{(n)})\right| \\ &\le \frac{\epsilon}{M} \sum_{j=1}^{m_n-1} \left|F(t_{j+1}^{(n)}) - F(t_j^{(n)})\right| \le \frac{\epsilon}{M} M = \epsilon \end{aligned}$$

for sufficiently large n. This proves the claim.

Example 3.4. Let $\{B_t\}_{t \ge 0}$ be Brownian motion and $T > 0$. Then sample paths

$$t \in [0,T] \to B_t(\omega)$$

are *not* of bounded variation a.e. ω.

To see this, choose a sequence $\{P_n\}$ of partitions of $[0,T]$ with $\|P_n\| \to 0$. Let S_n be as in theorem 3.3, so that

$$S_n = \sum_{j=1}^{m_n-1} \left[B_{t_{j+1}^{(n)}}(\omega) - B_{t_j^{(n)}}(\omega)\right]^2 \xrightarrow{\|\cdot\|_2} T.$$

Now by theorems 1.25 and 1.24 there exists a subsequence S_{n_k} such that

$$S_{n_k}(\omega) \to T \qquad \text{a.e. } \omega$$

Relabeling the sequence $\{P_{n_k}\}$ to $\{P_n\}$, there exists a set $A \in \mathcal{F}$, $P(A^c) = 0$ such that

$$S_n(\omega) = \sum_{j=1}^{m_n-1} \left[B_{t_{j+1}^{(n)}}(\omega) - B_{t_j^{(n)}}(\omega)\right]^2 \to T \neq 0 \qquad \forall \omega \in A. \tag{3.8}$$

Choosing $F(t) = B_t(\omega)$ in the previous claim, we see that

$$t \in [0,T] \to B_t(\omega)$$

cannot be of bounded variation $\forall \omega \in A$, that is, a.e.

Definition 3.8. An $(\mathcal{F}_t)$-adapted process $\{X_t\}_{t\geq 0}$ is said to have the *Markov property* relative to $\{\mathcal{F}_t\}_{t\geq 0}$, if

$$E[\, f(X_t) \,|\, \mathcal{F}_r \,] = E[\, f(X_t) \,|\, \sigma(X_r) \,] \tag{3.9}$$

for every bounded Borel function $f : \mathbb{R} \to \mathbb{R}$ and every $0 \leq r < t$.

Remark 3.8. 1. For ease of notation, one usually writes $E[\, f(X_t) \,|\, X_r \,]$ instead of $E[\, f(X_t) \,|\, \sigma(X_r) \,]$. Then (3.9) becomes:

$$E[\, f(X_t) \,|\mathcal{F}_r \,] = E[\, f(X_t) \,|\, X_r \,]$$

for every bounded Borel function f and every $0 \leq r < t$.

2. (Interpretation) Recall that $[\, f(X_t) \,|\mathcal{F}_r \,]$ is the $\mathcal{F}_r$-measurable random variable "closest" to $f(X_t)$, i.e. which best predicts $f(X_t)$, given time r. Thus, (3.9) means that besides the information which we know through X_r already, no other information available at time r (=events in $\mathcal{F}_r$) can help predict the future value $f(X_t)$.

3. If $\{X_t\}_{t\geq 0}$ has the Markov property relative to $\{\mathcal{F}_t\}_{t\geq 0}$, then it also has this property relative to the natural filtration $\{\mathcal{F}_t^X\}_{t\geq 0}$. In fact, since

$$\sigma(X_r) \subseteq \mathcal{F}_r^X \subseteq \mathcal{F}_r$$

for all r, then by theorem 2.7,

$$\begin{aligned} E[\, f(X_t) \,|\mathcal{F}_r^X \,] &= E\Big[\, E[\, f(X_t) \,|\, \mathcal{F}_r \,] \,\Big|\, \mathcal{F}_r^X \,\Big] \\ &= E\Big[\, \underbrace{E[\, f(X_t) \,|\, X_r \,]}_{\sigma(X_r)\text{-meas.} \,\Rightarrow\, \mathcal{F}_r^X\text{-meas.}} \,\Big|\, \mathcal{F}_r^X \,\Big] = E[\, f(X_t) \,|\, X_r \,] \end{aligned}$$

for $0 \leq r < t$, which was to be shown.

Theorem 3.4. *If $\{B_t\}_{t\geq 0}$ is $(\mathcal{F}_t)$-Brownian motion, then it has the Markov property. That is,*

$$E\big[\,f(B_t)\,\big|\,\mathcal{F}_r\,\big] = E\big[\,f(B_t)\,\big|\,B_r\,\big]$$

or every bounded Borel function f and every $0 \leq r < t$.

Proof. By $(B2)$, B_{t-r} and $\mathcal{F}_r$ are independent. Hence we can apply theorem 2.8 to

$$\mathcal{F}_o = \mathcal{F}_r, \quad X_1 = B_r, \quad X_2 = B_t - B_r, \quad H(x_1, x_2) = f(x_1 + x_2)$$

to obtain that

$$h(B_r) = E\big[\,f\big(B_r + (B_t - B_r)\big)\,\big|\,\mathcal{F}_r\,\big]$$

that is,

$$h(B_r) = E\big[\,f(B_t)\,\big|\,\mathcal{F}_r\,\big]. \tag{3.10}$$

On the other hand, since $h(x_1)$ is Borel, then the composition $h(B_r)$ is $\sigma(B_r)$-measurable so that

$$\begin{aligned} E\big[\,f(B_t)\,\big|\,B_r\,\big] &\underset{\text{thm 2.7}}{=} E\Big[\,E\big[\,f(B_t)\,\big|\,\mathcal{F}_r\,\big]\,\Big|\,B_r\Big] \underset{(3.10)}{=} E\big[\,h(B_r)\,\big|\,F_r\,\big] \\ &\underset{\text{expl 2.14}}{=} h(B_r) \underset{(3.10)}{=} E\big[\,f(B_t)\,\big|\,\mathcal{F}_r\,\big]. \end{aligned}$$

This proves the theorem. □

Remark 3.9. We can even give a formula for $E\big[\,f(B_t)\,\big|\,\mathcal{F}_r\,\big]$. Keeping the notation of the theorem, we have

$$\begin{aligned} h(x_1) = E\big[\,H(x_1, B_t - b_r)\,\big] &= \int_\Omega H\big(x_1, \underbrace{(B_t - B_r)}_{N(0,t-r)}(\omega)\,\big)\,dP(\omega) \\ &= \int_{\mathbb{R}} H(x_1, x_2)\frac{1}{\sqrt{2\pi}(t-r)}e^{-x_2^2/[2(t-r)]}\,dx_2 \\ &= \int_{\mathbb{R}} f(x_1 + x_2)\frac{1}{\sqrt{2\pi}(t-r)}e^{-x_2^2/[2(t-r)]}\,dx_2 \qquad x_2 \to x_2 - x_1 \\ &= \frac{1}{\sqrt{2\pi}(t-r)}\int_{\mathbb{R}} f(x_2)e^{-(x_2-x_1)^2/[2(t-r)]}\,dx_2 \end{aligned}$$

and thus

$$\begin{aligned} E\big[\,f(B_t)\,\big|\,\mathcal{F}_r\,\big](\omega) = h(B_r)(\omega) &= h\big(B_r(\omega)\big) \\ &= \frac{1}{\sqrt{2\pi}(t-r)}\int_{\mathbb{R}} f(y)e^{-(y-B_r(\omega))^2/[2(t-r)]}\,dy. \end{aligned}$$

3.3 Martingales

Throughout this section, $(\Omega, \mathcal{F}, P)$ will denote a fixed probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. A martingale is an integrable stochastic process whose value at time r is the best predictor for its value at $t > r$:

Definition 3.9. An $(\mathcal{F}_t)$-adapted process $\{M_t\}_{t\geq 0}$ is called an $(\mathcal{F}_t)$-*martinagle*, if

(Ma1) $\quad M_t \in L^1(\Omega)$ for all $t \geq 0$,

(Ma2) $\quad E[\, M_t \,|\, \mathcal{F}_r \,] = M_r$ for all $0 \leq r < t$.

If in addition,

(Ma1') $\quad M_t \in L^2(\Omega)$ for all $t \geq 0$,

then it is called a *square integrable martingale.*

Theorem 3.5. *Let $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$ be $(\mathcal{F}_t)$-martingales.*

1. *For all real numbers α, β, the process $\{\alpha X_t + \beta Y_t\}_{t\geq 0}$ is also an $(\mathcal{F}_t)$-martingale (so the set of $(\mathcal{F}_t)$-martingales is a real vector space),*

2. *The expected value is constant, that is,*

$$E(X_t) = E(X_r) \qquad \forall\, 0 \leq r < t.$$

Proof. Lets first prove 1. Since for each $t \geq 0$, $X_t, Y_t \in L^1(\Omega, \mathcal{F}_t, P)$ and the latter is a vector space, it follows that $\alpha X_t + \beta Y_t \in L^1(\Omega, \mathcal{F}_t, P)$. That is, $\{\alpha X_t + \beta Y_t\}_{t\geq 0}$ is an $(\mathcal{F}_t)$ adapted, integrable process. Furthermore, by linearity of the conditional expectation, for each $0 \leq r < t$,

$$E\big[\, \alpha X_t + \beta Y_t \,\big|\, \mathcal{F}_r \,\big] = \alpha E[\, X_t \,|\, \mathcal{F}_r \,] + \beta E[\, Y_t \,|\, \mathcal{F}_r \,] \underset{\text{(Ma2)}}{=} \alpha X_r + \beta Y_r.$$

This shows that $\{\alpha X_t + \beta Y_t\}_{t\geq 0}$ is an $(\mathcal{F}_t)$-martinagle.

To prove 2., observe that for all $0 \leq r < t$,

$$E(X_t) \underset{\text{thm 2.6}}{=} E\big(\, E[\, X_t \,|\, \mathcal{F}_r \,]\big) \underset{\text{(Ma2)}}{=} E(X_r).$$

This proves the theorem. ∎

The limits of martingales are also martingales:

Theorem 3.6. *Let $1 \leq p < \infty$, and let*

$$\Big\{ M_n(t) : t \geq 0 \Big\}_{n=1}^{\infty}$$

be a sequence of $(\mathcal{F}_t)$-martingales with $M_n(t) \in L^p(\Omega) \quad \forall\, t \geq 0, \ \forall n.$

If $\{M_n(t)\}_{n=1}^{\infty}$ converges in $L^p(\Omega)$ for all t, say

$$M_n(t) \xrightarrow{\|\cdot\|_p} M(t) \qquad \forall t \geq 0$$

then $\{M(t) : t \geq 0\}$ is also an $(\mathcal{F}_t)$-martingale.

Proof. For each t, since $M_n(t) \in L^p(\Omega, \mathcal{F}_t, P)$, since $M_n(t) \to M(t)$ in $L^p(\Omega, \mathcal{F}, P)$ and since $L^p(\Omega, \mathcal{F}_t, P)$ is complete, it follows that $M(t) \in L^p(\Omega, \mathcal{F}_t, P)$. Thus by exercise 1.10, $M(t) \in L^1(\Omega, \mathcal{F}_t, P)$ for all t, and hence $\{M(t) : t \geq 0\}$ is $(\mathcal{F}_t)$-adapted.

We thus need only verify the martingale property (M2). That is, we need to show that for all $0 \leq r < t$,

$$\int_A M(t)\,dP = \int_A M(r)\,dP \qquad (\forall\, A \in \mathcal{F}_r). \tag{3.11}$$

In fact, for all s and all $A \in \mathcal{F}_s$,

$$\Big|\int_A M_n(s)\,dP - \int_A M(s)\,dP\Big| \leq \int_A |M_n(s) - M(s)|\,dP \leq \int_\Omega |M_n(s) - M(s)|\,dP$$
$$= \|M_n - M\|_1 \leq \|M_n - M\|_p \to 0 \qquad \text{as} \quad n \to \infty$$

by exercise 1.10. That is,

$$\int_A M_n(s)\,dP \to \int_A M(s)\,dP. \tag{3.12}$$

for all s and all $A \in \mathcal{F}_s$. Now as $\{M_n : t \geq 0\}$ is a martingale, identity (3.11) holds for each M_n and hence applying (3.12) to $s = r$ and $s = t$ and $A \in \mathcal{F}_r \subseteq \mathcal{F}_t$ we have

$$\int_A M(t)\,dP \underset{(3.12)}{=} \lim_{n\to\infty} \int_A M_n(t)\,dP \underset{(3.11)}{=} \lim_{n\to\infty} \int_A M_n(r)\,dP \underset{(3.12)}{=} \int_A M(r)\,dP.$$

This proves the theorem. □

Remark 3.10. In the above proof, we have to be careful what limit process we choose. Recall that functions in $L^p(\Omega, \mathcal{F}_t, P)$ are defined up to a null set only. By concluding in the proof that "$M(t) \in L^p(\Omega, \mathcal{F}_t, P)$" we mean: "There exists $M(t) \in L^p(\Omega, \mathcal{F}_t, P)$ such that $M_n(t) \xrightarrow{\|\cdot\|_p} M(t)$."

On the other hand the statement "If $M(t) \in L^p(\Omega, \mathcal{F}, P)$ and $M_n(t) \xrightarrow{\|\cdot\|_p} M(t)$ then $M(t) \in L^p(\Omega, \mathcal{F}_t, P)$" is not correct in general; $M(t)$ need not be $\mathcal{F}_t$-measurable. However, if $\mathcal{F}_0$ (and hence $\mathcal{F}_t$) contains all null sets of $\mathcal{F}$ then by applying theorem 1.12 to the σ-algebra $\mathcal{F}_t$ we can be sure that this latter statement is correct.

Example 3.5. *Every standard $(\mathcal{F}_t)$-Brownian motion $\{B_t\}_{t\geq 0}$ is a square integrable $(\mathcal{F}_t)$-martingale.*

In fact, since $\{B_t\}_{t\geq 0}$ is standard, then

$$B_t = B_t - 0 = B_t - B_0$$

is $N(0, t)$. Hence by exercise 2.2, $B_t \in L^p(\Omega)$ for all $1 \leq p < \infty$. In particular, B_t is square integrable.

Next we need to verify the martingale property. By (B2) we have for all $0 \leq r < t$,

$$\begin{aligned} E[\,B_t \,|\, \mathcal{F}_r\,] &= E[\,(B_t - B_r) + B_r \,|\, \mathcal{F}_r\,] \\ &= E[\,B_t - B_r \,|\, \mathcal{F}_r\,] + E[\,B_r \,|\, \mathcal{F}_r\,] \\ &= E[\,B_t - B_r\,] + B_r && \text{(by examples 2.14 and 2.15)} \\ &= 0 + B_r = B_r. && \big(B_t - B_r \text{ is } N(0, t - r)\big) \end{aligned}$$

Thus proves the assertion.

Exercise 3.2. Let $\{B_t\}_{t\geq 0}$ be standard $(\mathcal{F}_t)$-Brownian motion. Show:

1. Set $Y_t = B_t^2 - t$. Then $\{Y_t\}_{t\geq 0}$ is an $\{F_t\}$-martingale.
   (Hint: write $B_t^2 = B_r^2 + 2B_r(B_t - B_r) + (B_t - B_r)^2$ for $0 \leq r < t$.)

2. If $X(\omega)$ is an $N(0,r)$ random variable, then $e^{X(\omega)} \in L^p(\Omega)$ for all $1 \leq p < \infty$.

3. Set $Z_t := e^{(\theta B_t - \theta^2 t/2)}$ ($\theta \neq 0$ is constant). Then $\{Z_t\}_{t\geq 0}$ is an $\{F_t\}$-martingale.
   (Hint: write $B_t = (B_t - B_r) + B_r$ for $0 \leq r < t$.)

Chapter 4

Stochastic Integrals

Throughout this chapter, $(\Omega, \mathcal{F}, P)$ will denote a probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and $\{B_t\}_{t\geq 0}$ (one-dimensional) $(\mathcal{F}_t)$-Brownian motion.

Given an $(\mathcal{F}_t)$-progressive stochastic process $\{f_t\}_{t\geq 0}$, we want to work with integrals

$$\text{(I)} \qquad F(\omega) = \int_a^b f(s,\omega)\, d\lambda(s) \qquad (\omega \in \Omega)$$

and

$$\text{(II)} \qquad I(\omega) = \int_a^b f(s,\omega)\, dB_s(\omega) \qquad (\omega \in \Omega).$$

In short,

$$F = \int_a^b f(s)\, ds \qquad \text{and} \qquad I = \int_a^b f(s)\, dB_s.$$

If $s \to f(s,\omega)$ is integrable for all ω then obviously, the integral (I) exists. However we still need to show that $F(\omega)$ is $\mathcal{F}_b$-measurable.

Consider the integral (II). Recall that for a.e. ω, the map $s \to B_s(\omega)$ is *not* of bounded variation. Hence, given a partition P of $[a,b]$, the values of $\Delta B_j(\omega) = B_{t_{j+1}}(\omega) - B_{t_j}(\omega)$ will vary extremely as $\Delta t_j = t_{j+1} - t_j \to 0$. We thus can not expect that the sums

$$\sum_j f(t_j,\omega)\Delta B_j(\omega)$$

converge for each ω, as $\|P\| \to 0$. In fact, we will define this integral as a limit in the *mean square* instead of a pointwise limit.

4.1 The Classes $\mathcal{V}[a,b]$ and $\mathcal{W}[a,b]$

Return for a while to the integral (I). If $f \in L^1\big([a,b]\times\Omega, \mathcal{B}([a,b])\otimes\mathcal{F}_b, \lambda\times P\big)$ then by Fubini's theorem, this integral is defined a.e. and there exists $F \in L^1(\Omega, \mathcal{F}_b, P)$ such that

$$F(\omega) = \int_a^b f(s,\omega)\, dP \qquad \text{a.e. } \omega \in \Omega. \tag{4.1}$$

In fact, inspection of the proof of Fubini's theorem (see [2], [3]) shows that (4.1) holds at every ω where $s \to f(s,w)$ is integrable. Thus in particular, if $f : [a,b] \times \Omega$ is $\mathcal{F}_b$-measurable and bounded, then (4.1) holds at every ω.

If f is *not* integrable on $[a,b] \times \Omega$ we still have:

Theorem 4.1. *Let $f : [a,b] \times \Omega$ be $\mathcal{B}([a,b]) \otimes \mathcal{F}_b$-measurable. Suppose that*

$$s \to f(s,w) \in L^1[a,b] \qquad \forall \omega \in \Omega. \tag{4.2}$$

Then

$$F(\omega) = \int_a^b f(s,\omega)\, ds$$

is $\mathcal{F}_b$-measurable.

Proof. The idea is to approximate f by bounded processes.

For each $n \in \mathbb{N}$, let

$$A_n := \{(s,\omega) \in [a,b] \times \Omega : |f(s,\omega)| \leq n\}.$$

Then A_n is a $\mathcal{B}([a,b]) \otimes \mathcal{F}_b$-measurable by assumption on f; hence $f_n := f\mathbf{1}_{A_n}$ is a $\mathcal{B}([a,b]) \otimes \mathcal{F}_b$-measurable function. Observe that for all $(s,\omega) \in [a,b] \times \Omega$,

(a) $|f_n(s,\omega)| \leq n$,

(b) $|f_n(s,\omega)| \leq |f(s,\omega)|$,

(c) $f_n(s,\omega) \to f(s,\omega)$.

Now as each f_n is bounded, by the above remarks,

$$F_n(\omega) = \int_a^b f_n(s,\omega)\, ds$$

is defined for all ω, and is $\mathcal{F}_b$-measurable. Since $\{f_n\}$ is dominated by $f \in L^1[a,b]$, we can apply the Lebesgue dominated convergence theorem to obtain that

$$F(\omega) = \int_a^b f_n(s,\omega)\, ds = \lim_{n\to\infty} \int_a^b f_n(s,\omega)\, ds = \lim_{n\to\infty} F_n(,\omega)$$

for each $\omega \in \Omega$. Thus by theorem 1.11, F is $\mathcal{F}_b$-measurable. □

For the integral (I) we will be working with progressive processes in the following class:

Definition 4.1. Given a closed interval $[a,b]$ with $a \geq 0$, set

$$\mathcal{W}[a,b] := \{g : [a,b] \times \Omega \to \mathbb{R} \ : \{g_t\}_{a \leq t \leq b} \text{ is } (\mathcal{F}_t)\text{-progressive and } t \mapsto g(t,\omega) \in L^1[a,b] \ \forall \omega \}.$$

We also set

$$\mathcal{W}[0,\infty) := \{g : [0,\infty) \times \Omega \to \mathbb{R} \ : \ g \in \mathcal{W}[0,T] \quad \forall\, T \geq 0\}.$$

Thus, the elements of $\mathcal{W}[0,\infty)$ are $(\mathcal{F}_t)$-progressive processes with $t \mapsto g(t,\omega)$ integrable on each $[0,T]$ for fixed ω. Note however that $t \mapsto g(t,\omega)$ need not be integrable on $[0,\infty)$.

Corollary 4.2. *Let $g \in \mathcal{W}[a,b]$. Set*

$$G_t(\omega) = \int_a^t g(s,\omega)\,ds. \qquad (a \le t \le b)$$

Then $\{G_t\}_{a\le t\le b}$ is $(\mathcal{F}_t)$-adapted and continuous (hence progressive).

Proof. Pick t, $a < t \le b$. By assumption, $g(s,\omega)$ is $\mathcal{B}([a,t]) \otimes \mathcal{F}_t$ measurable. Applying the previous theorem to the interval $[a,t]$ we thus obtain that G_t is $(\mathcal{F}_t)$-measurable. Since $G_a = 0$ it follows that $\{G_t\}_{a\le t\le b}$ is $(\mathcal{F}_t)$-adapted.

Next we show that $\{G_t\}$ is left-continuous on $[a,b]$. In fact, let $t_o \in (a,b]$ and fix $\omega \in \Omega$. Let $\epsilon > 0$ be given. By theorem 1.21 there exists $\delta > 0$ such that

$$\int_E |g(s,\omega)|\,ds < \epsilon$$

whenever $E \subset [a,t_o]$ and $\lambda(E) < \delta$. In particular, if $\max(a, t_o - \delta) < t < t_o$ we have

$$|G_t(\omega) - G_{t_o}(\omega)| = \left|\int_t^{t_o} g(s,\omega)\,ds\right| \le \int_t^{t_o} |g(s,\omega)|\,ds < \epsilon.$$

As ϵ was arbitrary, it follows that $\lim\limits_{t\to t_o^-} G_t(\omega) = G_{t_o}(\omega)$. Right continuity on $[a,b]$ is proved similarly. Hence continuity follows. □

Corollary 4.3. *Let $g \in \mathcal{W}[0,\infty)$. Set*

$$G_t(\omega) = \int_0^t g(s,\omega)\,ds. \qquad (t \ge 0)$$

Then $\{G_t\}_{t\ge 0}$ is $(\mathcal{F}_t)$-adapted and continuous (hence progressive).

Proof. Apply the previous corollary to each interval $[0,T]$, $T > 0$. □

For the integral (II) we work with a smaller class of stochastic processes:

Definition 4.2. Given a closed interval $[a,b]$ with $a \ge 0$, set

$$\mathcal{V}[a,b] := \big\{ f \in L^2([a,b]\times\Omega) \ : \ \{f_t\}_{a\le t\le b} \text{ is } (\mathcal{F}_t)\text{-progressive } \big\}.$$

We also set

$$\mathcal{V}[0,\infty) := \big\{ f(t,\omega) : [0,\infty)\times\Omega \to \mathbb{R} \ : \ f \in \mathcal{V}[0,T] \quad \forall T \ge 0 \big\}.$$

Thus, the elements of $\mathcal{V}[0,\infty)$ are $(\mathcal{F}_t)$-progressive processes which are square integrable on each $[0,T]\times\Omega$. Note that f need not be square integrable on $[0,\infty)\times\Omega$.

Remark 4.1. 1. It is straightforward to verify that $\mathcal{V}[a,b]$ is a closed subspace of $L^2([a,b]\times\Omega)$ in the L^2-norm,

$$\|f\|_2^2 = \int_\Omega \int_a^b f(t,\omega)^2\,dt\,dP = E\left[\int_a^b f(t,\omega)^2\,dt\right].$$

2. If $[a,b]\subseteq[c,d]$, then each $f\in\mathcal{V}[a,b]$ can be considered an element of $\mathcal{V}[c,d]$ by setting $f(s,\omega)=0$ for $s\notin[a,b]$. This gives rise to an isometric embedding of of $\mathcal{V}[a,b]$ in of $\mathcal{V}[c,d]$.

   Conversely, each $f\in\mathcal{V}[c,d]$ when restricted to $[a,b]$ becomes an element of $\mathcal{V}[a,b]$.

Just as every integrable function on a measure space can be arbitrarily approximated by an integrable simple function, we want to show that every element of $\mathcal{V}[a,b]$ can be arbitrarily approximated by a square-integrable simple function in the following sense.

Definition 4.3. A function $\phi:[a,b]\times\Omega\to\mathbb{R}$ is called *simple*, if it is of the form

$$\phi(t,\omega) = \phi_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{j=0}^{m-1}\phi_j(\omega)\mathbf{1}_{(t_j,t_{j+1}]}(t) \tag{4.3}$$

for some partition

$$P = \{a=t_0<t_1<t_2<\cdots<t_m=b\}$$

of $[a,b]$ where each ϕ_j is an $\mathcal{F}_{t_j}$-measurable random variable.

Remark 4.2. 1. Since $\{a\}$ is a null subset of $[a,b]$, we may drop the first term $\phi_0(\omega)\mathbf{1}_{\{0\}}$ in practice.

2. Let $t\in(a,b]$ be given. Then there exists a *unique* j such that $t\in(t_j,t_{j+1}]$, and

$$\phi(t)=\phi_j. \tag{4.4}$$

   It follows that the process $\{\phi(t): a\le t\le b\}$ is

   (a) $(\mathcal{F}_t)$-*adapted*: In fact, since ϕ_j is $\mathcal{F}_{t_j}$-measurable and $t_j<t$, then $\phi(t)$ is $\mathcal{F}_t$-measurable.

   (b) *left-continuous*: In fact, given $t\in(a,b]$ as above, we have for all s with $t_j<s<t$ that

$$\phi(s)=\phi_j=\phi(t)$$

   and hence $\lim\limits_{s\to t^-}\phi(s,\omega)=\lim\limits_{s\to t^-}\phi(t,\omega)=\phi(t,\omega)$ for all ω.

   (c) $(\mathcal{F}_t)$-*progressive.* This follows from (a) and (b).

(d) an element of $\mathcal{W}[a,b]$. This follows from (c) and the fact that for each ω,

$$\int_a^b |\phi(s,\omega)|\, ds = \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \phi(s,\omega)\, ds = \sum_{j=0}^{m-1} \phi_j(\omega)\Delta t_j < \infty.$$

3. Let's compute the L^p norm of ϕ. We have for $1 \le p < \infty$,

$$\begin{aligned} \int_\Omega \int_a^b |\phi(t,\omega)|^p\, dt\, dP &= \int_\Omega \left[\sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} |\phi(t,\omega)|^p\, dt \right] dP \\ &\underset{(4.3)}{=} \int_\Omega \left[\sum_{j=0}^{m-1} |\phi_j(\omega)|^p\, \Delta t_j \right] dP \qquad (\Delta t_j = t_{j+1} - t_j) \\ &= \sum_{j=0}^{m-1} \Delta t_j \int_\Omega |\phi_j(\omega)|^p\, dP \end{aligned} \tag{4.5}$$

and hence

$$\Delta t_{\min} \sum_{j=0}^{m-1} \int_\Omega |\phi_j(\omega)|^p\, dP \le \int_\Omega \int_a^b |\phi(t,\omega)|^p\, dt\, dP \le \Delta t_{\max} \sum_{j=0}^{m-1} \int_\Omega |\phi_j(\omega)|^p\, dP$$

where $\Delta t_{\min} = \min_j \Delta t_j$ and $\Delta t_{\max} = \max_j \Delta t_j = \|P\|$, or equivalently,

$$\Delta t_{\min} \sum_{j=0}^{m-1} E\big(|\phi_j|^p\big) \le E\Big[\int_a^b |\phi(t,\omega)|^p\, dt\Big] \le \Delta t_{\max} \sum_{j=0}^{m-1} E\big(|\phi_j|^p\big).$$

It follows that

$$\phi \in L^p\big([a,b] \times \Omega\big) \quad \Leftrightarrow \quad \phi_j \in L^p(\Omega) \quad \forall\, 1 \le j \le m$$

In case $p = 2$ then

$$\phi \in \mathcal{V}[a,b] \quad \Leftrightarrow \quad \phi_j \in L^2(\Omega) \quad \forall\, 1 \le j \le m \tag{4.6}$$

in which case its norm in $\mathcal{V}[a,b]$ is given by

$$\big\|\phi\big\|_{L^2([a,b]\times\Omega)} = \left[\sum_{j=0}^{m-1} \Delta t_j\, \|\phi_j\|_2^2 \right]^{1/2} = \left[\sum_{j=0}^{m-1} \Delta t_j\, E\big(\phi_j^2\big) \right]^{1/2}$$

by (4.5).

Definition 4.4. Let us set

$$\mathcal{S}[a,b] := \{\, \phi \in \mathcal{V}[a,b] : \phi \text{ is simple} \,\}.$$

It is easily seen that $\mathcal{S}[a,b]$ is a linear subspace of $\mathcal{V}[a,b]$. In fact, it turns out that its closure is $\mathcal{V}[a,b]$:

Theorem 4.4. *Let $f \in \mathcal{V}[a,b]$. There exists a sequence $\{\phi_n\}_{n=1}^{\infty}$ in $\mathcal{S}[a,b]$ such that*

$$\phi_n \xrightarrow{\|\cdot\|_2} f$$

in the norm of $L^2([a,b]\times\Omega)$.

Proof. To simplify notation, we will assume that $a = 0$ and set $T = b$, that is consider the interval $[0,T]$. The proof for general a is done in exactly the same way.

We prove the theorem in stages:

Lemma 4.5. *Let $f \in \mathcal{V}[0,T]$ satisfy*

1. *f is bounded,*

2. *The process $\{f(t) : 0 \le t \le T\}$ is left-continuous.*

Then the assertion of the theorem hold.

Proof. Let M be such that $|f(t,\omega)| \le M$ for all $(t,\omega) \in [0,T]\times\Omega$.

For each n, partition the interval $[0,T]$ into 2^n subintervals of length $\frac{T}{2^n}$ each and set

$$\phi_n := f(0)\mathbf{1}_{\{0\}} + \sum_{j=0}^{2^n-1} f\Big(\frac{jT}{2^n}\Big)\mathbf{1}_{(\frac{jT}{2^n},\frac{(j+1)T}{2^n}]} \tag{4.7}$$

just as in the proof of theorem 3.1 (with a slight change of notation). Since $\{f_t\}$ is $\mathcal{F}_t$-adapted, then obviously, ϕ_n is simple. Furthermore, since f is bounded, then $f(\frac{jT}{2^n})(\omega) \in L^2(\Omega)$ for each $0 \le j < 2^n$. Thus by (4.6), $\phi_n \in \mathcal{V}[0,T]$.

As shown in the proof of theorem 3.1,

$$\phi_n(t,\omega) \to f(t,\omega) \qquad \text{pointwise on } [0,T]\times\Omega.$$

Now as $|\phi_n(t,\omega)| \le M$ for all (t,ω), we can apply the dominated convergence theorem for L^p-spaces (theorem 1.20) to obtain that

$$\phi_n(t,\omega) \to f(t,\omega) \qquad \text{in } L^2([0,T]\times\Omega).$$

This proves the lemma. □

Lemma 4.6. *Let $h \in \mathcal{V}[0,T]$ be bounded. Then there exists a sequence $\{g_n\}_{n=1}^{\infty}$ in $\mathcal{V}[0,T]$ and satisfying*

1. *g_n is bounded for each n,*

2. *the process $\{g_n(t) : 0 \le t \le T\}$ is left-continuous*

such that $g_n \xrightarrow{\|\cdot\|_2} h$ in the norm of $L^2([0,T]\times\Omega)$.

Proof. The idea is to represent each g_n as an integral with the variable t appearing in the limit of integration. Then boundedness and continuity of g_n are easily derived.

1. *Construct g_n:* Since $\{h_t\}_{0\le t\le T}$ is $(\mathcal{F}_t)$-progressive and bounded, then by corollary 4.2, the process $\{H_t\}_{0\le t\le T}$ given by

$$H_t(\omega) := \int_0^t h(s,\omega)\,ds$$

is continuous and also $(\mathcal{F}_t)$-progressive. For each n, set

$$g_n(t,\omega) := n\Big[H_t(\omega) - H_{t_n}(\omega)\Big] = n\int_{t_n}^t h(s,\omega)\,ds$$

where $t_n = \max\{t-\frac{1}{n},0\}$.

2. *Show that $g_n \in \mathcal{V}[0,T]$:* Observe that

(a) each $g_n(t,w)$ is $(\mathcal{F}_t)$-adapted, as H_t is $(\mathcal{F}_t)$-adapted and $t_n \le t$.

(b) for each $\omega\in\Omega$, the map $t\to g_n(t,\omega)$ is left continuous, as $t\to H_t(\omega)$ is continuous, and $t\to H_{t_n}(\omega)$ is the composition of the continuous maps $t\to \max(t-\frac{1}{n},0)$ and $t\to H_t(\omega)$.

Hence by theorem 3.1, $\{g_n(t,\omega): t\ge 0\}$ is $(\mathcal{F}_t)$-progressive on $[0,T]$.

Now as h is bounded, say $h(t,\omega)\le M$, we have for all $0\le t\le T$ and $\omega\in\Omega$,

$$|g_n(t,\omega)| \le n\int_{t_n}^t |h(s,\omega)|\,ds \le n\cdot M\cdot(t-t_n) \le n\cdot M\cdot\frac{1}{n} = M.$$

That is g_n is bounded and in particular, square-integrable. Hence, $g_n\in\mathcal{V}[0,T]$.

3.. *Show that $g_n \xrightarrow{\|\cdot\|_2} h$:* By Lebesgue's theorem (theorem 1.17) we have for each ω that

$$\lim_{n\to\infty} g_n(t,\omega) = \lim_{n\to\infty} n\int_{t_n}^t h(s,\omega)\,ds = \lim_{n\to\infty}\frac{1}{\lambda\big((t_n,t]\big)}\int_{t_n}^t h(s,\omega)\,ds = h(t,\omega)$$

a.e. $t\in(0,T]$. As $|g_n(t,\omega)|, |h(\omega)| \le M$ for all (t,ω) it follows from the Lebesgue Dominated Convergence Theorem for L^p spaces (theorem 1.20) that for each fixed ω,

$$g_n(t,\omega) \quad \overset{\|\cdot\|_{L^2[0,T]}}{\longrightarrow} \quad h(t,\omega)$$

that is,

$$K_n(\omega) := \|g_n(\cdot,\omega) - h(\cdot,\omega)\|^2_{L^2[0,T]} = \int_0^T |g_n(s,\omega)-h(s,\omega)|^2\,ds \to 0$$

as $n\to\infty$. Observe that for all ω,

$$K_n(\omega) \le \|g_n(,\omega)\|^2_{L^2[0,T]} + \|h(\cdot,\omega)\|^2_{L^2[0,T]} \le \int_0^T M^2\,ds + \int_0^T M^2\,ds = 2M^2T.$$

Applying the Lebesgue Dominated Convergence Theorem we obtain by Tonelli theorem that

$$\|g_n - h\|^2_{L^2([0,T]\times\Omega)} = \int_\Omega \int_0^T \left|g_n(s,\omega) - h(s,\omega)\right|^2 ds\, dP = \int_\Omega K_n(\omega)\, dP \to 0$$

as $n \to \infty$ which proves the lemma. □

Lemma 4.7. *Let $f \in \mathcal{V}[0,T]$ be bounded. Then there exists a sequence $\{h_n\}_{n=1}^\infty$ in $\mathcal{V}[0,T]$, with each h_n bounded, such that $h_n \xrightarrow{\|\cdot\|_2} f$ in the norm of $L^2([0,T]\times\Omega)$.*

Proof. We clip f to make it bounded: For each $n \in \mathbb{N}$, let

$$h_n(t,\omega) := \begin{cases} -n & \text{if } f(t,\omega) < -n \\ f(t,\omega) & \text{if } -n \le f(t,\omega) \le n \\ n & \text{if } f(t,\omega) > n. \end{cases}$$

Then

1. $h_n(t,\omega) \to f(t,\omega)$ for all $(t,\omega) \in [0,T]\times\Omega$,
2. $|h(t,\omega)| \le n$, i.e. h_n is bounded,
3. each h_n is $(\mathcal{F}_t)$-progressive. To see this, let $t \in [0,T]$ be given. We must show that h_n is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable. Set

$$\begin{aligned} A_+ &:= \{(s,\omega) \in [0,t]\times\Omega : f(t,\omega) > n\} \\ A_- &:= \{(s,\omega) \in [0,t]\times\Omega : f(t,\omega) < -n\} \\ A &:= A_+ \cup A_-. \end{aligned}$$

   Then $A_+, A_-, A \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t$ since f is $(\mathcal{F}_t)$-progressive and hence is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable. Now as

$$h_n = f\mathbf{1}_{A^c} + n\mathbf{1}_{A+} - n\mathbf{1}_{A_-}$$

   is follows that h_n is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable.

Now as $|h_n| \le f$ and $f \in L^2([0,T]\times\Omega)$ we conclude that $h_n \in \mathcal{V}[a,b]$. Furthermore, using (a) we can apply the Lebesgue Dominated Convergence Theorem for L^p spaces to obtain that

$$h_n \xrightarrow{\|\cdot\|_2} f$$

as well. Thus, the lemma is proved. □

We are now ready to complete the proof of theorem 4.4. Fix $f \in \mathcal{V}[0,T]$ and let $\epsilon > 0$ be given. By lemma 4.7 there exists a bounded process $h \in \mathcal{V}[0,T]$ such that

$$\|f - h\|_{L^2([0,T]\times\Omega)} < \frac{\epsilon}{3}.$$

Now by lemma 4.6 there exists a bounded and left-continuous $g \in \mathcal{V}[0, T]$ such that

$$\|h - g\|_{L^2([0,T]\times\Omega)} < \frac{\epsilon}{3}.$$

Then by lemma 4.5 there exists $\phi \in \mathcal{S}[0, T]$ such that

$$\|g - \phi\|_{L^2([0,T]\times\Omega)} < \frac{\epsilon}{3}.$$

Combining the above three inequalities we obtain by the triangle inequality that

$$\|f - \phi\|_{L^2([0,T]\times\Omega)} < \epsilon.$$

Now choosing $\epsilon = \frac{1}{n}$ we obtain a sequence $\{\phi_n\}$ in $\mathcal{S}[0, T]$ such that

$$\|f - \phi_n\|_{L^2([0,T]\times\Omega)} < \frac{1}{n} \qquad \forall n.$$

This proves the theorem. □

Definition 4.5. A function $\phi : [0, \infty) \times \Omega \to \mathbb{R}$ is called *simple*, if it is of the form

$$\phi(t, \omega) = \phi_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{j=0}^{\infty} \phi_j(\omega)\mathbf{1}_{(t_j,t_{j+1}]}(t) \tag{4.8}$$

where $0 = t_0 < t_1 < t_2 < \cdots < t_m < \ldots$ with $\lim_{j\to\infty} t_j = \infty$ and each ϕ_j is an $\mathcal{F}_{t_j}$-measurable random variable.

We set

$$\mathcal{S}[0, \infty) := \{\phi \in \mathcal{V}[0, \infty) : \phi \text{ is simple }\}.$$

Remark 4.3.

1. For each fixed t, exactly one term in (4.8) is nonzero. Thus we need not worry about convergence of this infinite series.

2. If $\phi \in \mathcal{S}[0, b]$ for some b, then obviously, $\phi \in \mathcal{S}[0, \infty)$ as well. Conversely, the restriction of an element $\phi \in \mathcal{S}[0, \infty)$ to an interval $[a, b]$ defines an element of $\mathcal{S}[a, b]$.

3. Let $\phi : [0, \infty) \times \Omega \to \mathbb{R}$ be *simple*. Then ϕ is left-continuous. Furthermore, just as in remark 4.2 one verifies that

$$\phi \in \mathcal{S}[0, \infty) \quad \Leftrightarrow \quad \phi_j \in L^2(\Omega) \quad \forall j = 1, 2, \ldots$$

4. Let $f \in \mathcal{V}[0, \infty)$. Then $f \in \mathcal{V}[0, n]$ for all $n \in \mathbb{N}$. By theorem 4.4 there exists $\phi_n \in \mathcal{S}[0, n]$ (and hence $\phi_n \in \mathcal{S}[0, \infty)$) such that

$$\|\phi_n - f\|_{L^2([0,n]\times\Omega)} < \frac{1}{n}$$

for all n.

Consider now the sequence $\{\phi_n\}$. Let $T > 0$ be arbitrary, but given. By construction, for all $n \geq \frac{1}{T}$ we have

$$\|\phi_n - f\|_{L^2([0,T]\times\Omega)} \leq \|\phi_n - f\|_{L^2([0,n]\times\Omega)} < \frac{1}{n} \to 0$$

as $n \to \infty$.

We want the equivalent of theorem 4.4 for the class $\mathcal{W}[a,b]$. Inspection of the proof shows that theorem 4.4 is valid for the space $L^1([a,b]\times\Omega)$ (and in fact for each L^p space). That is, we obtain:

Let $\{g_t\}_{a\le t\le b}$ be $(\mathcal{F}_t)$-progressive with $g\in L^1([a,b]\times\Omega)$. There exists a sequence $\{\phi_n\}$ of simple, $(\mathcal{F}_t)$-adapted and integrable functions such that

$$\phi_n \xrightarrow{\|\cdot\|_1} g. \tag{4.9}$$

in $L^1([a,b]\times\Omega)$

Since the class $\mathcal{W}[a,b]$ is larger than $\mathcal{V}[a,b]$, some additional work is now required.

Theorem 4.8. *Let $g\in\mathcal{W}[a,b]$. There exists a sequence $\{\phi_k\}$ of simple and integrable functions in $\mathcal{W}[a,b]$ such that*

$$\phi_k(\cdot,\omega)\xrightarrow{\|\cdot\|_1} g(\cdot,\omega) \qquad a.e.\ \omega$$

in $L^1[a,b]$.

Proof. For each $n\in\mathbb{N}$, set

$$A_n := \{\,(s,\omega)\in[a,b]\times\Omega\ :\ |g(s,\omega)|\le n\,\}$$

and set $g_n := g\mathbf{1}_{A_n}$. Then g_n is bounded and $(\mathcal{F}_t)$-progressive with $|g_n|\le|g|$ for all n and $g_n(s,\omega)\to g(s,\omega)$ on $[a,b]\times\Omega$. Hence by the LDCT for L^p-spaces (theorem 1.20), $g_n(\cdot,\omega)\to g(\cdot,\omega)$ in $L^1[a,b]$ for each ω, that is,

$$\big\|\,g_n(\cdot,\omega)-g(\cdot,\omega)\,\big\|_{L^1[a,b]}\ \to\ 0 \qquad \forall\omega. \tag{4.10}$$

and hence by theorem 1.23,

$$\big\|\,g_n(\cdot,\omega)-g(\cdot,\omega)\,\big\|_{L^1[a,b]}\ \xrightarrow{\text{prob.}}\ 0.$$

On the other hand, by (4.9), for each n there exists a sequence $\{\phi_n^m\}_{m=1}^\infty$ of simple, integrable functions in $\mathcal{W}[a,b]$ such that

$$\begin{aligned}\int_\Omega \big\|\phi_n^m(\cdot,\omega)-g_n(\cdot,\omega)\big\|_{L^1[a,b]}\,dP &= \int_\Omega\int_a^b |\phi_n^m(s,\omega)-g_n(s,\omega)|\,ds\,dP\\ &= \big\|\phi_n^m-g_n\big\|_{L^1([a,b]\times\Omega)}\to 0\end{aligned}$$

as $m\to\infty$. As L^1-convergence implies convergence in probability (see theorem 1.25), then

$$\big\|\phi_n^m(\cdot,\omega)-g_n(\cdot,\omega)\big\|_{L^1[a,b]}\ \xrightarrow{\text{prob.}}\ 0$$

as $m\to\infty$. Now we do a diagonalization process as in the proof of theorem 1.24. For each k (choosing $\epsilon=\frac{1}{2k}$) we pick n_k such that

$$\mu\Big(\Big\{\omega\ :\ \big\|\,g_{n_k}(\cdot,\omega)-g(\cdot,\omega)\,\big\|_{L^1[a,b]}>\frac{1}{2k}\Big\}\Big)<\frac{1}{2^{k+1}}.$$

Then for each k we pick $\phi_{n_k}^m$ (call it ϕ_k for simplicity) such that

$$\mu\Big(\Big\{\omega : \|\,\phi_k(\cdot,\omega) - g_{n_k}(\cdot,\omega)\,\|_{L^1[a,b]} > \frac{1}{2k}\Big\}\Big) < \frac{1}{2^{k+1}}.$$

It follows that for each k,

$$\begin{aligned}
\mu\Big(\Big\{\omega : \|\,\phi_k(\cdot,\omega) - g(\cdot,\omega)\,\|_{L^1[a,b]} > \frac{1}{k}\Big\}\Big) \\
&\le \mu\Big(\Big\{\omega : \|\,\phi_k(\cdot,\omega) - g_{n_k}(\cdot,\omega)\,\|_{L^1[a,b]} > \frac{1}{2k}\Big\}\Big) \\
&\qquad + \mu\Big(\Big\{\omega : \|\,g_{n_k}(\cdot,\omega) - g(\cdot,\omega)\,\|_{L^1[a,b]} > \frac{1}{2k}\Big\}\Big) \\
&< \frac{1}{2^{k+1}} + \frac{1}{2^{k+1}} = \frac{1}{2^k}.
\end{aligned}$$

Arguing now as in the proof of theorem 1.24 we conclude that

$$\|\,\phi_k(\cdot,\omega) - g(\cdot,\omega)\,\|_{L^1[a,b]} \to 0 \qquad \text{a.e. } \omega.$$

Thus, the theorem is proved. □

4.2 The Itô Integral

Just as for the Lebesgue integral, we first introduce the Itô integral for simple functions. Then we define the integral of an arbitrary $f \in \mathcal{V}[a,b]$ as the limit of a sequence of integrals of simple functions.

Definition 4.6. Let $\phi \in \mathcal{S}[a,b]$. That is,

$$\phi = \phi_0 \mathbf{1}_{\{0\}} + \sum_{j=0}^{m-1} \phi_j \mathbf{1}_{(t_j, t_{j+1}]} \tag{4.11}$$

for some partition

$$P = \{a = t_0 < t_1 < t_2 < \cdots < t_m = b\}$$

of $[a,b]$ where each ϕ_j is a square-integrable $\mathcal{F}_{t_j}$-measurable random variable. We define its *Itô integral* by

$$I(\phi) = \int_a^b \phi(s,\omega)\, dB_s(\omega) \;:=\; \sum_{j=0}^{m-1} \phi_j \left[B_{t_{j+1}}(\omega) - B_{t_j}(\omega)\right] \tag{4.12}$$

or in short,

$$\int_a^b \phi(s)\, dB_s \;=\; \sum_{j=0}^{m-1} \phi_j\, \Delta B_j.$$

Remark 4.4. The verification of the following properties are left as an exercise. Most make use of the easily verified fact that when replacing the partition P by a refinement in (4.11), the integral (4.12) does not change.

1. $I(\phi)$ is independent of the representation (4.11) of ϕ chosen.
2. If $a = b$ then $I(\phi) = 0$.
3. Since each ϕ_j and B_{t_j} are $\mathcal{F}_{t_j}$-measurable, and $t_j \leq b$ for all j, it follows that the Itô integral (4.12) is $\mathcal{F}_b$-measurable.
4. The map $\phi \in \mathcal{S}[a, b] \mapsto I(\phi)$ is linear.
5. If we split $[a, b]$ into subintervals $[a, c]$ and $[c, b]$ then

$$\int_a^b \phi(s)\, dB_s = \int_a^c \phi(s)\, dB_s + \int_c^b \phi(s)\, dB_s.$$

Exercise 4.1. Prove properties 1., 4. and 5. in the above remark.

Lemma 4.9. *The map*

$$\phi \mapsto I(\phi) = \int_a^b \phi(s)\, dB_s$$

is a linear isometry of $\mathcal{S}[a, b]$ *into* $L^2(\Omega, \mathcal{F}_b, P)$. *In particular,*

$$\| I(\phi) \|_{L^2(\Omega)} = \| \phi \|_{L^2([a,b]\times\Omega)}$$

or equivalently ("Itô isometry"),

$$E\left[\left[\int_a^b \phi(s)\, dB_s \right]^2 \right] = E\left[\int_a^b \phi(s, \omega)^2\, ds \right]. \tag{4.13}$$

for all $\phi \in \mathcal{S}[a, b]$.

Proof. By remark 4.4 it is only left to show that $I(\phi) \in L^2(\Omega)$ and that the Itô isometry (4.13) holds. Not that for ϕ as in (4.11) we have

$$\begin{aligned} I(\phi)^2 &= \left[\int_a^b \phi(s)\, dB_s \right]^2 = \left[\sum_{j=0}^{m-1} \phi_j\, \Delta B_j \right]^2 \\ &= \sum_{j=0}^{m-1} \phi_j^2\, (\Delta B_j)^2 + \sum_{\substack{i,j=0 \\ i \neq j}}^{m-1} \phi_i\, \Delta B_i\, \phi_j\, \Delta B_j \\ &= \sum_{j=0}^{m-1} \phi_j^2\, (\Delta B_j)^2 + 2 \sum_{\substack{i,j=0 \\ i < j}}^{m-1} \phi_i\, \Delta B_i\, \phi_j\, \Delta B_j. \end{aligned} \tag{4.14}$$

We need to show that each term is integrable. By (B2), $\Delta B_j = B_{t_{j+1}} - B_{t_j}$ is independent of ϕ_j. Hence,

$$\phi_j^2 \in L^1(\Omega) \qquad \text{and} \qquad (\Delta B_j)^2 \in L^1(\Omega)$$

are also independent. It follows from corollary 2.4 that

$$\underbrace{\phi_j^2}_{\in L^1} \underbrace{(\Delta B_j)^2}_{\in L^1} \in L^1(\Omega)$$

as well, so that $\phi_j \Delta B_j \in L^2(\Omega)$ for all j.

Now applying Hölder's inequality we conclude that

$$\underbrace{\phi_i \Delta B_i}_{\in L^2} \underbrace{\phi_j}_{\in L^2} \in L^1(\Omega)$$

Now using (B2) again, we have for $i < j$ as $\phi_i \Delta B_i \phi_j$ is $\mathcal{F}_{t_j}$-measurable that $\phi_i \Delta B_i \Phi_j \in L^1(\Omega)$ and $\Delta B_j \in L^1(\Omega)$ are independent, hence by corollary 2.4 again,

$$\underbrace{\phi_i \Delta B_i \phi_j}_{\in L^1} \underbrace{\Delta B_j}_{\in L^1} \in L^1(\Omega).$$

Thus, each term in (4.14) is integrable, which shows that $I(\phi) \in L^2(\Omega)$.

To prove the Itô isometry, observe that by (4.14) and linearity of the integral,

$$\begin{aligned} E\big[\, I(\phi)^2 \,\big] &= \sum_{j=0}^{m-1} E\big[\, \phi_j^2 \, (\Delta B_j)^2 \,\big] \; + \; 2 \sum_{\substack{i,j=0 \\ i<j}}^{m-1} E\big[\, \phi_i \, \Delta B_i \, \phi_j \, \Delta B_j \,\big] \\ &\underset{\text{cor 2.4}}{=} \sum_{j=0}^{m-1} E\big[\, \phi_j^2 \,\big] \, E\big[\, (\Delta B_j)^2 \,\big] \; + \; 2 \sum_{\substack{i,j=0 \\ i<j}}^{m-1} E\big[\, \phi_i \, \Delta B_i \, \phi_j \,\big] E\big[\, \Delta B_j \,\big] \\ &= \sum_{j=0}^{m-1} E\big[\, \phi_j^2 \,\big] \, \Delta t_j \qquad \qquad \big(\, \Delta B_j \text{ is } N(0, \Delta t_j)\big) \\ &= E\Bigg[\, \sum_{j=0}^{m-1} \phi_j^2 \, (t_{j+1} - t_j) \Bigg] \qquad \qquad \big(\, \Delta t_j = t_{j+1} - t_j \,\big) \\ &= E\bigg[\, \int_a^b \phi(s, \omega)^2 \, ds \bigg]. \qquad \qquad \Big(\, \phi^2 = \phi_0^2 \, \mathbf{1}_{\{0\}} + \sum_{j=0}^{m-1} \phi_j^2 \, \mathbf{1}_{(t_j, t_{j+1}]} \Big) \end{aligned}$$

This proves the lemma. □

Remark 4.5. Let

$$\phi = \phi_0 \mathbf{1}_{\{0\}} + \sum_{j=0}^{m-1} \phi_j \mathbf{1}_{(t_j, t_{j+1}]} \in \mathcal{S}[0, T] \tag{4.15}$$

for some $T > 0$, where

$$P = \{0 = t_0 < t_1 < t_2 < \cdots < t_m = T\}$$

is a partition of $[0, T]$. As mentioned in remark 4.3, $\phi \in \mathcal{S}[0, t]$ for all $t \geq 0$. (By restriction if $t < T$, or rewriting (4.15) as

$$\phi = \phi_0 \mathbf{1}_{\{0\}} + \sum_{j=0}^{m-1} \phi_j \mathbf{1}_{(t_j, t_{j+1}]} + 0 \cdot \mathbf{1}_{(T,t]}$$

if $t > T$). Thus we can set

$$\boxed{I_t(\phi) := \int_0^t \phi(s)\, dB_s}$$

for all $t \geq 0$. Obviously, $I_t(\phi)$ is constant on $[T, \infty)$.

Lemma 4.10. *Let $\phi \in \mathcal{S}[0, T]$ for some $T > 0$. Then $\{I_t(\phi)\}_{t \geq 0}$ is a continuous, square-integrable $(\mathcal{F}_t)$-martingale.*

Proof. Lemma 4.9 shows that $I_t(\phi) \in L^2(\Omega, \mathcal{F}_t, P)$ for each $t \geq 0$. That is, $\{I_t(\phi)\}_{t \geq 0}$ is $(\mathcal{F}_t)$-adapted and square-integrable.

Continuity: This essentially follows from continuity of Brownian motion. In fact, let $t \in (0, \infty)$ be given. Going to a refinement of the partition P we may assume that $t \in P$, say $t = t_k$ for some $1 \leq k \leq m$. Then

$$\phi = \phi_0 \mathbf{1}_{\{0\}} + \sum_{j=0}^{k-1} \phi_j \mathbf{1}_{(t_j, t_{j+1}]} + \sum_{j=k}^{m-1} \phi_j \mathbf{1}_{(t_j, t_{j+1}]} \tag{4.16}$$

and

$$I_t(\phi) = \int_0^{t_k} \phi(s)\, ds = \sum_{j=0}^{k-1} \phi_j\, \Delta B_j = \sum_{j=0}^{k-2} \phi_j\, \Delta B_j + \phi_{k-1}\big(B_{t_k} - B_{t_{k-1}}\big)$$

Increasing T if necessary by adding a zero term in (4.16), we may assume that $t_k = t < T$.

First let $s \to t^- = t_k^-$. If s is sufficiently close to t_k, then $s \in (t_{k-1}, t_k]$ and hence

$$I_s(\phi) = \sum_{j=0}^{k-2} \phi_j\, \Delta B_j + \phi_{k-1}\big(B_s - B_{t_{k-1}}\big)$$

Thus for all $\omega \in \Omega$,

$$I_t(\phi)(\omega) - I_s(\phi)(\omega) = \phi_{k-1}(\omega)\big(B_{t_k}(\omega) - B_s(\omega)\big) \to 0$$

as $s \to t_k^- = t$ by continuity of $B_t(\omega)$.

Next let $s \to t^+ = t_k^+$. (In this case, we may also allow $t = t_0 = 0$.) If s is sufficiently close to t_k, then $s \in (t_k, t_{k-1}]$ and hence

$$I_s(\phi) = \sum_{j=0}^{k-1} \phi_j \, \Delta B_j + \phi_k \big(B_s - B_{t_k}\big)$$

Thus for all $\omega \in \Omega$,

$$I_s(\phi)(\omega) - I_t(\phi)(\omega) = \phi_k(\omega)\big(B_s(\omega) - B_{t_k}(\omega)\big) \to 0$$

as $s \to t_k^+ = t$ by continuity of $B_t(\omega)$. Thus, continuity follows.

Martingale Property: Let $0 \le r < t$. On $[0, t]$ we write ϕ in the form

$$\phi = \phi_0 \mathbf{1}_{\{0\}} + \sum_{j=0}^{k-1} \phi_j \mathbf{1}_{(t_j, t_{j+1}]} + \sum_{j=k}^{m-1} \phi_j \mathbf{1}_{(t_j, t_{j+1}]}$$

where P is a partition of $[0, t]$ chosen so that $r \in P$, say $r = t_k$. Then

$$P = \{0 = t_0 < t_1 < t_2 < \cdots < t_m = t\}$$

and

$$I_t(\phi) = \sum_{j=0}^{m-1} \phi_j \, \Delta B_j = \sum_{j=0}^{k-1} \phi_j \, \Delta B_j + \sum_{j=k}^{m-1} \phi_j \, \Delta B_j = I_r(\phi) + \sum_{j=k}^{m-1} \phi_j \, \Delta B_j. \tag{4.17}$$

Now for all $k \le j \le m$, as ϕ_j is $\mathcal{F}_{t_j}$-measurable, ΔB_j is independent of $\mathcal{F}_{t_j}$ hence of $\mathcal{F}_{t_k} = \mathcal{F}_r$ and is $N(0, \Delta t_j)$,

$$E\big[\,\phi_j \Delta B_j \,\big|\, \mathcal{F}_r\,\big] \underset{\text{thm 2.6}}{=} \phi_j E\big[\,\Delta B_j \,\big|\, \mathcal{F}_r\,\big] \underset{\text{ex. 2.15}}{=} \phi_j E\big[\,\Delta B_j\,\big] = \phi_j \cdot 0 = 0.$$

Hence in (4.17) we obtain by linearity of the conditional expectation,

$$E\big[\,I_t(\phi)\big|\,\mathcal{F}_r\,\big] = E\big[\,I_r(\phi)\big|\,\mathcal{F}_r\,\big] + \sum_{j=k}^{m-1} E\big[\,\phi_j \, \Delta B_j \big|\,\mathcal{F}_r\,\big] \underset{\text{ex. 2.14}}{=} I_r(\phi) + 0 = I_r(\phi).$$

This proves the lemma. □

Definition 4.7. Lemma 4.9 says that the Itô integral

$$I : \phi \in \mathcal{S}[a, b] \mapsto I(\phi) \in L^2(\Omega, \mathcal{F}_b, P)$$

is a linear isometry between normed linear spaces. Since $\mathcal{S}[a, b]$ is dense in $\mathcal{V}[a, b]$ and $L^2(\Omega, \mathcal{F}_b, P)$ is complete, it follows immediately from a basic theorem on bounded linear operators that I extends uniquely to a *linear isometry*

$$I : \mathcal{V}[a, b] \to L^2(\Omega, \mathcal{F}_b, P).$$

This extension is determined as follows: Given $f \in \mathcal{V}[a, b]$, pick an arbitrary sequence $\{\phi_n\} \in \mathcal{S}[a, b]$ converging to f. Then

$$I(f) = \lim_{n \to \infty} I(\phi_n).$$

$I(f)$ is called the *Itô integral* of f.

Remark 4.6. We easily conclude:

1. If $a = b$ then $I(f) = 0$ as $I(\phi_n) = 0$ for all n.

2. Being a limit in $L^2(\Omega, \mathcal{F}_b, P)$, the integral $I(f)$ is $\mathcal{F}_b$-measurable.

3. If we split $[a, b]$ into subintervals $[a, c]$ and $[c, b]$ then

$$\int_a^b f(s)\, dB_s = \int_a^c f(s)\, dB_s + \int_c^b f(s)\, dB_s.$$

Exercise 4.2. Prove property 3. in the above remark.

Now let $f \in \mathcal{V}[0, \infty)$. Then $f \in \mathcal{V}[0, t]$ for each $t \geq 0$, so $I_t(f)$ is defined and $I_t(f) \in L^2(\Omega, \mathcal{F}_t, P)$. We thus have an $(\mathcal{F}_t)$-adapted process $\{ I_t(f) \}_{t \geq 0}$. Obviously, the map

$$\{f_t\}_{t\geq 0} \mapsto \{I_t(f)\}_{t\geq 0} = \left\{ \int_0^t f(s)\, dB_s : t \geq 0 \right\}$$

is linear.

Theorem 4.11. *Suppose, $\mathcal{F}_0$ contains all P-null sets. Then for each $f \in \mathcal{V}[0, \infty)$, $\{ I_t(f) \}_{t\geq 0}$ is a square-integrable $(\mathcal{F}_t)$-martingale.*

Proof. It is only left to prove the martingale property:

$$E\left[I_t(f) \,\middle|\, \mathcal{F}_r \right] = I_r(f) \qquad \forall\, 0 \leq r < t.$$

Let $\{\phi_n\}$ be a sequence of simple functions in $\mathcal{S}[0, \infty)$ that such for all $t \geq 0$, $\phi_n \to f$ in $L^2([0, t] \times \Omega)$, as shown in remark 4.3. By lemma 4.10, each $\{ I_t(\phi_n)\}_{t\geq 0}$ is an $(\mathcal{F}_t)$-martingale. By definition of the Itô integral, $I_t(\phi_n) \xrightarrow{\|\cdot\|_2} I_t(f)$ for each $t \geq 0$; thus it follows from theorem 3.6 and the remark following it that $\{ I_t(f) \}_{t\geq 0}$ is also an $(\mathcal{F}_t)$-martingale. □

While each $\{ I_t(\phi_n) \}_{t\geq 0}$ is continuous by lemma 4.10, this need no longer be true for $\{ I_t(f) \}_{t\geq 0}$. However, one can prove the following (see [5]):

Theorem 4.12. *Keep the above notation. Suppose that $(\Omega, \mathcal{F}, P)$ is complete and $\mathcal{F}_0$ contains all P-null sets. Then $\{ I_t(\phi_n) \}_{t\geq 0}$ has a continuous version $\{ J_t(\phi_n) \}_{t\geq 0}$.*

From now on, we will always choose this continuous version as the Itô integral. Thus we will throughout assume that $(\Omega, \mathcal{F}, P)$ is complete and $\mathcal{F}_0$ contains all P-null sets.

4.3 Some Special Integrals

In this section, we will discuss two simple Itô integrals which will be needed later.

First we integrate Brownian motion $\{B_t\}_{t\geq 0}$ itself. For this we must verify that $\{B_t\}_{t\geq 0} \in \mathcal{V}[0,\infty)$. We already know that Brownian motion is progressive. Now fix $t > 0$ and observe that since $B_s - B_0$ is $N(0, s-0)$ then for all $t \geq 0$,

$$\int_0^t E\big[(B_s - B_0)^2\big]\,ds = \int_0^t s\,ds = \frac{t^2}{2} < \infty$$

which shows that $\{B_s - B_0\}_{0\leq s\leq t} \in L^2([0,t]\times\Omega)$. So if we assume in addition that $B_0 \in L^2(\Omega)$ then we will be assured that

$$\{B_s\}_{0\leq s\leq t} = \{B_s - B_0\}_{0\leq s\leq t} + \{B_0\}_{0\leq s\leq t} \in L^2([0,t]\times\Omega)$$

for all $t \geq 0$ as required.

Recall that in the usual Riemann integral, if $f(s) = s$ we have

$$2\int_0^t s\,ds = s^2\Big]_0^t = t^2.$$

In case of the Itô integral we have an extra term:

Theorem 4.13. *Let $\{B_t\}_{t\geq 0}$ be $(\mathcal{F}_t)$-Brownian motion with $B_0 \in L^2(\Omega)$. Then for all $t \geq 0$,*

$$B_t^2 - B_0^2 = 2\int_0^t B_s\,dB_s + t$$

Proof. The assertion is obvious for $t = 0$. Thus let $t > 0$ be arbitrary. First we approximate $B(s,\omega)$ by simple functions on $[0,t]\times\Omega$,

$$\phi_n(s,\omega) = \sum_{j=0}^{n-1} B_{\frac{jt}{n}}(\omega)\mathbf{1}_{\left(\frac{jt}{n},\frac{(j+1)t}{n}\right]}(s).$$

Claim: $\phi_n(s,\omega) \xrightarrow{\|\cdot\|_2} B_s(\omega)$ in $L^2([0,t]\times\Omega)$ for each $t \geq 0$.
In fact,

$$\begin{aligned}
\big\|\phi_n(s) - B_s\big\|^2_{L^2([0,t]\times\Omega)} &= \int_0^t\int_\Omega \big[\phi_n(s,\omega) - B_s(\omega)\big]^2 dP\,ds \qquad \text{(Tonelli)}\\
&= \sum_{j=0}^{n-1}\int_{\frac{jt}{n}}^{\frac{(j+1)t}{n}}\int_\Omega \big[\phi_n(s,\omega) - B_s(\omega)\big]^2\,dP\,ds\\
&= \sum_{j=0}^{n-1}\int_{\frac{jt}{n}}^{\frac{(j+1)t}{n}}\int_\Omega \big[B_{\frac{jt}{n}}(\omega) - B_s(\omega)\big]^2\,dP\,ds\\
&= \sum_{j=0}^{n-1}\int_{\frac{jt}{n}}^{\frac{(j+1)t}{n}} E\Big[\big[B_s - B_{\frac{jt}{n}}\big]^2\Big]\,ds.
\end{aligned}$$

As ΔB_t is $N(0,t)$ this becomes

$$
\begin{aligned}
\left\|\phi_n(s) - B_s\right\|^2_{L^2([0,t]\times\Omega)} &= \sum_{j=0}^{n-1} \int_{\frac{jt}{n}}^{\frac{(j+1)t}{n}} \left[s - \frac{jt}{n}\right] ds \\
&= \sum_{j=0}^{n-1} \int_0^{\frac{t}{n}} s\, ds \qquad\qquad (s \to s + \frac{jt}{n}) \\
&= n \cdot \frac{t^2}{2n^2} = \frac{t^2}{2n} \to 0 \qquad \text{as } n \to \infty.
\end{aligned}
$$

This proves the claim.

Now by definition of the Itô integral,

$$
\int_0^t B_s\, dB_s = \lim_{n\to\infty} \int_0^t \phi_n(s)\, dB_s = \lim_{n\to\infty} \sum_{j=0}^{n-1} B_{\frac{jt}{n}}\left[B_{\frac{(j+1)t}{n}} - B_{\frac{jt}{n}}\right] \tag{4.18}
$$

as a limit in $L^2(\Omega)$, while

$$
\begin{aligned}
B_t^2 - B_0^2 &= \sum_{j=0}^{n-1} \left[B^2_{\frac{(j+1)t}{n}} - B^2_{\frac{jt}{n}}\right] \qquad\qquad \text{(telescoping sum)} \\
&= \sum_{j=0}^{n-1} \left[B_{\frac{(j+1)t}{n}} - B_{\frac{jt}{n}}\right]^2 + 2\sum_{j=0}^{n-1} B_{\frac{jt}{n}}\left[B_{\frac{(j+1)t}{n}} - B_{\frac{jt}{n}}\right] \\
&\to t + 2\int_0^t B_s\, dB_s \qquad \text{as } n \to \infty
\end{aligned}
$$

in $L^2(\Omega)$ by theorem 3.3 and (4.18). This proves the theorem. □

Remark 4.7. It follows immediately that for $0 \le r < t$,

$$
B_t^2 - B_r^2 = \left[2\int_0^t B_s\, dB_s + t\right] - \left[2\int_0^r B_s\, dB_s + r\right] = 2\int_r^t B_s\, dB_s + (t - r).
$$

Next we consider an integral of the form

$$
\int_0^t h(s)\, dB_s
$$

where $h(s)$ is a *deterministic function*, that is, independent of ω. We will need the following result:

Theorem 4.14. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\{X_n\}_{n=1}^\infty$ a sequence of $N(0, r_n)$ random variables (where $r_n > 0$ for all n).*

1. *If $X_n \to X$ in probability, and $r_n \to r$, then X is $N(0,r)$.*
2. *If $X_1, \ldots, X_n$ are independent, then $X_1 + \cdots + X_n$ is $N(0, r_1 + \cdots + r_n)$.*

Theorem 4.15. *Let $t > 0$ and $h : [0,t] \to \mathbb{R}$ be continuous, $h \neq 0$. Set*

$$Z := \int_0^t h(s)\, dB_s.$$

Then Z is a $N(0,r)$ random variable, where

$$r = \int_0^t h(s)^2\, ds = \|h\|^2_{L^2[0,t]}.$$

Proof. We approximate h by deterministic simple functions,

$$\phi_n(s) := h(0)\mathbf{1}_{\{0\}} + \sum_{j=0}^{n-1} h\Big(\frac{jt}{n}\Big)\mathbf{1}_{\left(\frac{jt}{n}, \frac{(j+1)t}{n}\right]}(s).$$

Since h is uniformly continuous on $[0,t]$ then $\phi_n \to h$ uniformly on $[0,t]$. Now all these functions are constant with respect to ω, hence $\phi_n \to h$ uniformly on $[0,t]\times\Omega$ as well. Thus, $\phi_n \xrightarrow{\|\cdot\|_2} h$ in $L^2([0,t]\times\Omega)$.

Now by definition of the Itô integral.

$$Z(\omega) = \lim_{n\to\infty}\int_0^t \phi_n(\omega)\, dB_s = \lim_{n\to\infty}\sum_{j=0}^{n-1} \underbrace{h\Big(\frac{jt}{n}\Big)}_{:=\alpha_{n,j}} \underbrace{\Big[B_{\frac{(j+1)t}{n}} - B_{\frac{jt}{n}}\Big](\omega)}_{:=\Delta B_{n,j}(\omega)}$$

For convenience, set

$$Z_n(\omega) := \sum_{j=0}^{n-1} Z_{n,j}(\omega) \qquad \text{where} \qquad Z_{n,j}(\omega) = \alpha_{n,j}\Delta B_{n,j}(\omega)$$

so that $Z_n \to Z$ in $L^2(\Omega)$.

1. Since Brownian motion has independent increments, for each n, the random variables
$$\{Z_{n,j} = \alpha_{n,j}\Delta B_{n,j}\}_{j=0}^{n-1}$$
are independent, and by (B3) and exercise 2.2 they are $N(0,(\alpha_{n,j})^2\Delta t)$ where $\Delta t = \frac{t}{n}$.

2. Thus by theorem 4.14, each Z_n is $N(0,\alpha_n)$ where
$$\alpha_n = \sum_{j=0}^{n-1}(\alpha_{n,j})^2\Delta t = \sum_{j=0}^{n-1}\Big[h\Big(\frac{jt}{n}\Big)\Big]^2\Delta t > 0$$
provided that $h(\frac{jt}{n}) \neq 0$ for at least one j. (This is the case if n is sufficiently large as $h \neq 0$ and h is continuous; in all series above we remove the terms corresponding to $h(\frac{jt}{n}) = 0$.)

Now let $n \to \infty$. Then by theorem 4.14 again, we obtain that $Z(\omega)$ is $N(0,r)$ where

$$r = \lim_{n\to\infty}\alpha_n = \lim_{n\to\infty}\sum_{j=1}^{n-1}\Big[h\Big(\frac{jt}{n}\Big)\Big]^2\Delta t = \int_0^t h(s)^2\, ds$$

as a Riemann integral. This proves the theorem. □

Chapter 5

Itô's Formula

The Riemann and Lebesgue integrals are difficult to compute directly from their definitions. Instead, we usually make use of the Fundamental Theorem of Calculus for their computation. The situation for the Itô integral is similar; the main tool here is called Itô's formula.

Recall the chain rule of differentiation: If $F(x)$ and $g(s)$ are differentiable functions, then

$$\frac{d}{ds} F\big(g(s)\big) = F'\big(g(s)\big)g'(s). \tag{5.1}$$

Now if $F, g \in C^1(\mathbb{R})$ then by the Fundamental Theorem of Calculus we obtain the integral version of the chain rule,

$$F\big(g(b)\big) - F\big(g(a)\big) \;=\; F\big(g(s)\big)\Big]_a^b \;=\; \int_a^b F'\big(g(s)\big)\, dg(s) \tag{5.2}$$

where $dg(s) = g'(s)\, ds$.

Similarly, let $F(y, x) \in C^1(\mathbb{R}^2)$ and $g(s) \in C^1(\mathbb{R})$. Then by the chain rule,

$$\frac{d}{ds} F\big(s, g(s)\big) = F_y\big(s, g(s)\big) + F_x\big(s, g(s)\big)g'(s). \tag{5.3}$$

So setting $dg(s) = g'(s)\, ds$ we obtain the integrated version,

$$F\big(b, g(b)\big) - F\big(a, g(a)\big) \;=\; F\big(s, g(s)\big)\Big]_a^b = \int_a^b F_y\big(s, g(s)\big)\, ds + \int_a^b F_x\big(s, g(s)\big)\, dg(s). \tag{5.4}$$

Our goal is to obtain similar formulas for the Itô integral. Note that if in (5.2) we choose $F(x) = x^2$, $g(s) = s$ and $a = 0$ and $b = t$ we obtain

$$g(t)^2 - g(0)^2 = 2\int_0^t g(s)\, dg(s)$$

Now if formula (5.2) generalized directly to Brownian motion, setting $g(s) = B_s$ we would obtain

$$B_t^2 - B_0^2 = 2\int_0^t B_s, dB_s$$

which is incorrect: Theorem 4.13 shows that there is an additional term t. In general, the generalizations of the above formulas will contain additional terms.

Throughout this chapter, $(\Omega, \mathcal{F}, P)$ will denote a *complete* probability space, $\mathcal{F}_0$ containing all null sets. Hence the Itô integral

$$I_t(f)(\omega) = \int_0^t f(s,\omega)\, dB_s(\omega) \qquad (t \geq 0)$$

will have a continuous version. We will furthermore assume that $\{B_t\}_{t\geq 0}$ is either Brownian motion with B_0 normally distributed, say $N(0, s_o)$ for some $s_o > 0$, or standard Brownian motion (i.e. $B_0 = 0$ in which case we set $s_o = 0$). Then as pointed out in the previous section, $\{B_t\}_{t\geq 0} \in \mathcal{V}[0,\infty)$. In addition, by theorem 4.14, $B_t = (B_t - B_0) + B_0$ is $N(0, t + s_o)$ for all t, and hence $B_t \in L^2(\Omega)$ for all t.

5.1 Itô's Formula

We begin with a particularly simple case of Itô 's formula.

Theorem 5.1. (Itô's Formula for Brownian Motion.) *Let $F \in C^2(\mathbb{R})$ and suppose that*

$$\text{(A)} \qquad F'\big(B_s(\omega)\big) \in L^2\big([0,t] \times \Omega\big) \qquad \forall\, t \geq 0.$$

Then for all $0 \leq a \leq b$,

$$\boxed{F(B_b) - F(B_a) = \int_a^b F'(B_s)\, dB_s + \frac{1}{2}\int_a^b F''(B_s)\, ds}. \tag{5.5}$$

Remark 5.1. 1. Observe the extra term involving the second derivative of F when compared to (5.2).

2. Since $s \to B_s(\omega)$ and F' are continuous, so is the composition $s \to F'\big(B_s(\omega)\big)$, for each ω. Obviously, the process $\{F'(B_s)\}_{s\geq 0}$ is $(\mathcal{F}_t)$-adapted by continuity of F'. Hence by theorem 3.1, this process is also $(\mathcal{F}_t)$-progressive. Thus, condition (A) is equivalent to:

$$\text{(A')} \qquad \big\{F'(B_s)\big\}_{s\geq 0} \in \mathcal{V}[0,\infty)$$

3. If we integrate over the fixed finite interval $[a,b]$ only, then it is enough to require that $\{F'(B_s)\}_{a\leq s\leq b} \in \mathcal{V}[a,b]$, or equivalently, that $F'\big(B_s(\omega)\big) \in L^2\big([a,b] \times \Omega\big)$.

4. One can define the Itô integral for processes $\{f_t\}_{t\geq 0}$ not in $\mathcal{V}[0,\infty)$, but only satisfying

$$s \mapsto f(s,\omega) \in L^2[0,t] \qquad \forall\, t \geq 0, \quad \text{a.e. } \omega.$$

Since F' is continuous and Brownian motion has continuous sample paths, the process $\{F'(B_s)\}_{s\geq 0}$ satisfies this assumption. Hence working with this more general case, condition (A) is not required.

Example 5.1. Let $F(x) = e^x$. Obviously, $F \in C^2(\mathbb{R})$; in fact, $F = F' = F''$.

First consider the process $\{e^{B_s}\}_{s\geq 0}$. By our standing assumption on Brownian motion, $\{B_s\}$ is $N(0, s + s_o)$. Thus by (2.9),

$$E\Big[\big(e^{B_s}\big)^2\Big] = E\Big[e^{2B_s}\Big] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(s+s_o)}}\, e^{2x}\, e^{-x^2/2(s+s_o)}\, dx$$
$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(s+s_o)}}\, e^{-[x^2-4x(s+s_o)]/2(s+s_o)}\, dx$$
$$= e^{2(s+s_o)} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(s+s_o)}}\, e^{-[x-2(s+s_o)]^2/2(s+s_o)}\, dx \underset{(2.13)}{=} e^{2(s+s_o)}$$

and hence

$$\int_0^t E\Big[\big(e^{B_s}\big)^2\Big]\, ds = \int_0^t e^{2(s+s_o)}\, ds < \infty.$$

It follows that $\{F'(B_s)\}_{s\geq 0} = \{e^{B_s}\}_{s\geq 0} \in \mathcal{V}[0,\infty)$, i.e. condition (A) is satisfied.

Now we apply Itô's formula to obtain for all $t \geq 0$,

$$e^{B_t} - e^{B_0} = \int_0^t e^{B_s}\, dB_s + \frac{1}{2}\int_0^t e^{B_s}\, ds.$$

So $X(t) := e^{B_t}$ is the solution of what is called a *stochastic integral equation*

$$X_t = X_0 + \int_0^t X_s\, dB_s + \frac{1}{2}\int_0^t X_s\, ds. \tag{5.6}$$

with initial condition $X(0) = e^{B_0}$. In order to shorten this equation we write it in *differential form*,

$$dX_t = X_t\, dB_t + \frac{1}{2} X_t\, dt. \tag{5.7}$$

This last equation is called a *stochastic differential equation (SDE)*. We have shown that $X(t) = e^{B_t}$ is the solution of the SDE (5.7) with initial condition $X(0) = e^{B_0}$.

Definition 5.1. A stochastic process $\{X_t\}_{t\geq 0}$ is called an *Itô process*, if there exist $f \in \mathcal{V}[0,\infty)$ and $g \in \mathcal{W}[0,\infty)$ such that for all $t \geq 0$,

$$\text{(B)} \qquad \boxed{X_t = X_0 + \int_0^t f(s)\, dB_s + \int_0^t g(s)\, ds \qquad \text{a.e. } \omega}\,.$$

Remark 5.2. 1. Writing (B) in *differential form* we obtain

$$\text{(C)} \qquad dX_t = f(t)\, dB_t + g(t)\, dt.$$

We call this equation the *differential* of the Itô process $\{X_t\}$.

2. By theorem 4.2 we know that the Itô integral has a continuous version. Furthermore, by corollary 4.2 the integral of g on the right is continuous. Hence, every Itô process has a continuous version which we may choose as the process itself. Since the integrals in (B) are $(\mathcal{F}_t)$-adapted, then by theorem 3.1, the continuous version is $(\mathcal{F}_t)$-progressive.

Example 5.2. 1. Brownian motion $X(t) = B_t$ is itself an Itô process. In fact, setting $X_0 = B_0$, $f(s) = 1$ and $g(s) = 0$, the right-hand side of (B) becomes

$$B_0 + \int_0^t 1\, dB_s = B_0 + (1)\big[B_t - B_0\big] = B_t.$$

2. By example 5.2,

$$e^{B_t} = e^{B_0} + \int_0^t e^{B_s}\, dB_s + \frac{1}{2}\int_0^t e^{B_s}\, ds.$$

So $X(t) = e^{B_t}$ is an Itô process with $f(s) = e^{B_s}$ and $g(s) = \frac{1}{2}e^{B_s}$.

We are now ready to state the general Itô formula for Itô processes. Recall here that by definition, an Itô process is a solution of the SDE

$$dX_t = f(t)\, dB_t + g(t)\, dt$$

where f and g satisfy the assumptions of definition 5.1. Thus by an integral of the form $\int_a^b G(s)\, dX_s$ we will mean

$$\int_a^b G(s)\, dX_s := \int_a^b G(s)f(s)dB_s + \int_a^b G(s)g(s)\, ds. \tag{5.8}$$

Theorem 5.2. (General Itô formula). *let $\{X_t\}_{t\geq 0}$ be an Itô process, and let $F(y,x)$ be continuously differentiable in the first variable y, and twice continuously differentiable in the second variable x. Suppose that*

$$(A) \qquad F_x\big(s, X(s)\big)f(s) \in L^2\big([0,t]\times\Omega\big) \qquad \forall\, t \geq 0.$$

Then for all $0 \leq a \leq b$,

$$F\big(b, X(b)\big) - F\big(a, X(a)\big) = \int_a^b F_y(s, X_s)\, ds + \int_a^b F_x(s, X_s)\, dX_s + \frac{1}{2}\int_a^b F_{xx}(s, X_s) f(s)^2\, ds \qquad a.s. \tag{5.9}$$

Remark 5.3. 1. Observe the extra term involving the second derivative of F when compared to (5.4).

2. By (5.9) and (5.8) we have for all $t \geq 0$,

$$F\big(t, X(t)\big) = F\big(0, X(0)\big) + \int_0^t F_x(s, X_s) f(s)\, dB_s + \int_0^t \Big[F_y(s, X_s) + F_x(s, X_s)g(s) + \frac{1}{2}F_{xx}(s, X_s)f(s)^2\Big] ds \qquad \text{a.s.} \tag{5.10}$$

3. Because we always choose the continuous version of an Itô process, the map $s \longrightarrow X_s(\omega)$ is continuous. Furthermore, F_x is continuous by assumption; hence the composition $s \longrightarrow F_x\big(s, X_s(\omega)\big)$ is continuous for each ω.

   Similarly, as each X_s is $\mathcal{F}_s$-measurable and $F_x(s, \cdot)$ is continuous, the composition $\omega \longrightarrow F_x\big(s, X_s(\omega)\big)$ is $\mathcal{F}_s$ measurable for each s. That is,
   $$\big\{F_x(s, X_s)\big\}_{s\geq 0}$$
   is $(\mathcal{F}_s)$-adapted. Hence by theorem 3.1, it is $(\mathcal{F}_s)$-progressive. On the other hand, by assumption in definition 5.1, $\{f(s)\}_{s\geq 0}$ is $(\mathcal{F}_s)$-progressive. Hence the product process
   $$\big\{F_x(s, X_s)f(s)\big\}_{s\geq 0}$$
   is also $(\mathcal{F}_s)$-progressive.

   It follows that condition (A) is equivalent to:
   $$\text{(A')} \qquad \big\{F_x(s, X_s)f(s)\big\}_{s\geq 0} \in \mathcal{V}[0, \infty)$$
   which is required for the Itô integral in (5.10) to make sense.

4. Since F_y, F_x, F_{xx} are continuous, the same argument shows that the processes
   $$\big\{F_y(s, X_s)\big\}_{s\geq 0}, \quad \big\{F_x(s, X_s)g(s)\big\}_{s\geq 0} \quad \text{and} \quad \big\{F_{xx}(s, X_s)f(s)^2\big\}_{s\geq 0}$$
   are $(\mathcal{F}_s)$-progressive. Furthermore, the maps
   $$s \longrightarrow F_y(s, X_s(\omega)), \quad s \longrightarrow F_x(s, X_s(\omega)) \quad \text{and} \quad s \longrightarrow F_{xx}(s, X_s(\omega))$$
   are continuous ; hence bounded on all intervals $[0, t]$ for each ω. By assumption in definition 5.1, $g(s)$ is integrable over all intervals $[0, t]$ for each ω. Also, $f \in L^2\big([0, t] \times \Omega\big)$ for each $t \geq 0$, hence $f^2 \in L^1\big([0, t] \times \Omega\big)$ which by Fubini's theorem implies that $s \longrightarrow f(s, \omega)^2 \in L^1[0, t]$ for each t, a.e. ω. It follows that
   $$F_y(s, X_s) + F_x(s, X_s)g(s) + \frac{1}{2}F_{xx}(s, X_s)f(s)^2 \in L^1[0, t] \qquad \forall\, t \geq 0, \quad \text{a.s.},$$
   so that the Lebesgue integral in (5.10) is defined a.s. (and is defined at every ω in case f is deterministic).

5. Observe that by definition 5.1, the process $\big\{F(t, X_t)\big\}_{t\geq 0}$ given by (5.10) is again an Itô process.

6. If we only integrate over $[a, b]$, then it is enough to require in (A) that
   $$F_x\big(s, X(s)\big)f(s) \in L^2\big([a, b] \times \Omega\big).$$

7. If $f = 1$ and $g = 0$, then $dX_s = dB_s$ and the general Itô formula (5.9) becomes
   $$F\big(b, B_b\big) - F\big(a, B_a\big) = \int_a^b F_y(s, B_s)\, ds + \int_a^b F_x(s, B_s)\, dB_s + \frac{1}{2}\int_a^b F_{xx}(s, B_s)\, ds \qquad \forall\omega. \tag{5.11}$$
   If in addition, $F = F(x)$ is a function of the variable x only, then we recover the Itô formula for Brownian motion (5.5).

5.2 Proof of Itô's Formula

This section is devoted to the proof of the general Itô formula. First we need to discuss some tools required for the proof. Let us begin with a generalization of theorem 3.3.

Theorem 5.3. *Let $\{g(t)\}_{a\le t\le b}$ be a continuous and bounded $(\mathcal{F}_t)$-adapted process. Let $\{P_n\}_{n=1}^{\infty}$ be a sequence of partitions of $[a,b]$,*

$$P_n = \{a = t_0^{(n)} < t_1^{(n)} < t_2^{(n)} < \cdots < t_{m_n}^{(n)} = b\}$$

satisfying $\|P_n\| \to 0$ as $n \to \infty$. For each n, let

$$S_n = \sum_{j=0}^{m_n-1} g\big(t_j^{(n)}\big)(\omega)\Big[B_{t_{j+1}^{(n)}}(\omega) - B_{t_j^{(n)}}(\omega)\Big]^2$$

be a random variable determined by g and P_n. Then,

$$S_n \xrightarrow{\|\cdot\|_2} \int_a^b g(s)\,ds \qquad \text{as } n \to \infty$$

in $L^2(\Omega, \mathcal{F}_b, P)$.

Proof. Observe that the assumption implies that $\{g(t)\}_{a\le t\le b}$ is $(\mathcal{F}_t)$-progressive. For ease of notation, we drop the index "(n)" and often drop ω. As usual, we also set $\Delta t_j = t_{j+1} - t_j$ and $\Delta B_j = B_{j+1} - B_j$. Then

$$S_n = \sum_{j=0}^{m-1} g(t_j)\big(\Delta B_j\big)^2 \in L^2(\Omega, \mathcal{F}_b, P).$$

Since $s \to g(s,\omega)$ is continuous for all ω, the above integral is a Riemann integral for each fixed ω, and hence

$$\underbrace{\int_a^b g(s,\omega)\,ds}_{\substack{=:Z(\omega),\\ \text{Riemann integral}}} = \lim_{n\to\infty} \underbrace{\sum_{j=0}^{m-1} g(t_j,\omega)\Delta t_j}_{=:Z_n(\omega),\ \text{Riemann sum}} \quad . \tag{5.12}$$

for each $\omega \in \Omega$. As g is bounded, say $|g(t,\omega)| \le M$ on $[a,b]\times\Omega$, we have

(i) $|Z_n(\omega)| \le \sum_{j=0}^{m-1} M\Delta t_j = M(b-a) \in L^2(\Omega)$, and

(ii) by (5.12), $Z_n(\omega) \to Z(\omega)$ for all ω.

It follows from the LDCT for L^p-spaces (theorem 1.20) that

$$Z_n \xrightarrow{\|\cdot\|_2} Z = \int_a^b g(s)\,ds$$

that is, the limit in (5.12) is a limits in $L^2(\Omega, \mathcal{F}_b, P)$ as well. So it is enough show that

$$\|S_n - Z_n\|_2 \to 0 \tag{5.13}$$

because (5.13) will imply, using the triangle inequality, that

$$\|S_n - Z\|_2 \leq \|S_n - Z_n\|_2 + \|Z_n - Z\|_2 \to 0.$$

To prove (5.13), we compute.

$$\begin{aligned}
\|S_n - Z_n\|_2^2 &= E\big[\,(S_n - Z_n)^2\,\big] \\
&= E\Bigg[\Bigg(\sum_{j=0}^{m-1} g(t_j)\Big[\big(\Delta B_j\big)^2 - \Delta t_j\Big]\Bigg)^2\Bigg] \\
&= \sum_{j=0}^{m-1} E\Bigg[\underbrace{g(t_j)^2\Big[\big(\Delta B_j\big)^2 - \Delta t_j\Big]^2}_{(I)}\Bigg] \\
&\qquad + 2\sum_{\substack{i,j=0\\ i<j}}^{m-1} E\Bigg[\underbrace{g(t_i)g(t_j)\Big[\big(\Delta B_i\big)^2 - \Delta t_i\Big]\Big[\big(\Delta B_j\big)^2 - \Delta t_j\Big]}_{(II)}\Bigg].
\end{aligned} \tag{5.14}$$

Here we have made use of the fact that by linearity of the integral,

$$\begin{aligned}
E\Bigg[\Big(\sum_{j=0}^{m-1}\alpha_j\Big)^2\Bigg] &= E\Bigg[\sum_{i,j=0}^{m-1}\alpha_i\alpha_j\Bigg] \\
&= \sum_{i,j=0}^{m-1} E\big[\alpha_i\alpha_j\big] \\
&= \sum_{\substack{i,j=0\\ i=j}}^{m-1} E\big[\alpha_i\alpha_j\big] + \sum_{\substack{i,j=0\\ i<j}}^{m-1} E\big[\alpha_i\alpha_j\big] + \sum_{\substack{i,j=0\\ j<i}}^{m-1} E\big[\alpha_i\alpha_j\big] \\
&= \sum_{j=0}^{m-1} E\big[\,(\alpha_j)^2\,\big] + 2\sum_{\substack{i,j=0\\ i<j}}^{m-1} E\big[\alpha_i\alpha_j\big]
\end{aligned}$$

provided that $E(\alpha_i\alpha_j)$ is defined and finite for all i, j. (In (5.14) this is certainly satisfied with $\alpha_j = g(t_j)\Big[\big(\Delta B_j\big)^2 - \Delta t_j\Big]$ as will be shown now.)

First consider the terms of form (I). Since g is bounded, $g(t_j)^2 \in L^1(\Omega, \mathcal{F}_{t_j}, P)$. As $\Delta B_j \in L^4(\Omega)$, $\Delta t_j \in L^2(\Omega)$, then $\big[(\Delta B_j)^2 - \Delta t_j\big]^2 \in L^1(\Omega)$ by Hölder's inequality. Now by (B2), ΔB_j is independent of $\mathcal{F}_{t_j}$, and hence of $g(t_j)^2$. Hence by exercise (2.7) (applied to $f(x) = x^2 - \Delta t_j$ and $g(y) = y$ in the notation of the exercise), $\big[(\Delta B_j)^2 - \Delta t_j\big]^2$ and $g(t_j)^2$ are independent. We can thus apply corollary 2.4 to obtain that $(I) \in L^1(\Omega)$ and

$$\begin{aligned}
E[\,(I)\,] &= E\big[g(t_j)^2\big]\, E\Big[\,\big[\,(\Delta B_j)^2 - \Delta t_j\,\big]^2\Big] \\
&= E\big[g(t_j)^2\big]\, E\Big[\,(\Delta B_j)^4 - 2\Delta t_j(\Delta B_j)^2 + (\Delta t_j)^2\,\Big] \\
&= E\big[g(t_j)^2\big]\Big\{\,E[\,(\Delta B_j)^4\,] - 2\Delta t_j\, E[\,(\Delta B_j)^2\,] + E[\,(\Delta t_j)^2\,]\,\Big\} \\
&= E\big[g(t_j)^2\big]\Big\{\,3(\Delta t_j)^2 - 2\Delta t_j \cdot \Delta t_j + (\Delta t_j)^2\Big\} \quad \text{(by (B3) and exercise 2.2)} \\
&= E\big[g(t_j)^2\big] \cdot 2(\Delta t_j)^2.
\end{aligned}$$

Next consider the terms of form (II). As $\Delta B_j \in L^2(\Omega)$, $\Delta t_j \in L^1(\Omega)$ and g is bounded, then $\big[(\Delta B_j)^2 - \Delta t_j\big] \in L^1(\Omega)$ and $g(t_i)g(t_j)\big[(\Delta B_i)^2 - \Delta t_i\big] \in L^1(\Omega, \mathcal{F}_{t_j}, P)$. (note that $i < j$ so $g(t_i)$, $g(t_j)$ and $\big[(\Delta B_i)^2 - \Delta t_i\big]$ are all $\mathcal{F}_{t_j}$-measurable !) Using again independence of ΔB_j and $\mathcal{F}_{t_j}$ as above, we see that $\big[(\Delta B_j)^2 - \Delta t_j\big]$ and $g(t_i)g(t_j)\big[(\Delta B_i)^2 - \Delta t_i\big]$ are independent. Applying corollary 2.4 once more, we see that $(II) \in L^1(\Omega)$ and

$$\begin{aligned}
E[\,(II)\,] &= E\Big[\,g(t_i)g(t_j)\,\big[\,(\Delta B_i)^2 - \Delta t_i\big]\,\Big]\, E\Big[\,\big[\,(\Delta B_j)^2 - \Delta t_j\,\big]\,\Big] \\
&= E\Big[\,g(t_i)g(t_j)\,\big[\,(\Delta B_i)^2 - \Delta t_i\big]\,\Big]\Big\{\,E[\,(\Delta B_j)^2\,] - E[\,\Delta t_j\,]\,\Big\} \\
&= E\Big[\,g(t_i)g(t_j)\,\big[\,(\Delta B_i)^2 - \Delta t_i\big]\,\Big]\Big\{\,\Delta t_j - \Delta t_j\,\Big\} \quad (\text{as } \Delta B_j \text{ is } N(0, \Delta t_j)\,) \\
&= 0.
\end{aligned}$$

Hence (5.14) becomes

$$\begin{aligned}
\|S_n - Z_n\|_2^2 &= \sum_{j=0}^{m-1} E\big[\,g(t_j)^2\big] \cdot 2(\Delta t_j)^2 \\
&\le 2M^2\,\|P_n\| \sum_{j=0}^{m-1} \Delta t_j \qquad \text{(monotonicity of the integral)} \\
&= 2M^2\,\|P_n\|(b-a) \to 0 \qquad \text{as } n \to \infty.
\end{aligned}$$

This proves the theorem. □

Remark 5.4. 1. As a special case, choosing $g(t) = 1$, $a = 0$ and $b = t$, we obtain theorem 3.3.

2. (a) Using the definition of the Itô integral one can show that

$$\sum_{j=0}^{m-1} g(t_j)\Delta B_j \xrightarrow{\|\cdot\|_2} \int_a^b g(t)\, dB_t \tag{5.15}$$

as $n \to \infty$, similar to the proof of theorem 4.15.

(b) The above theorem says that

$$\sum_{j=0}^{m-1} g(t_j)(\Delta B_j)^2 \xrightarrow{\|\cdot\|_2} \int_a^b g(t)\,dt = \lim_{n\to\infty} \sum_{j=0}^{m-1} g(t_j)\Delta t_j$$

as $n \to \infty$ (where the right-hand limit is a point-wise limit of Riemann sums, but also an L^2-limit as shown in the proof of the theorem).
For this reason, one often writes "$(\Delta B_t)^2 = \Delta t$" or "$(dB_t)^2 = dt$".

Exercise 5.1. Prove that (5.15) holds for any $g \in \mathcal{V}[a,b]$.

The proof of the next theorem can be found in [1]. It is an important ingredient in the proof that the Itô integral has a continuous version. Compare this theorem with Chebychev's inequality !

Theorem 5.4. *Let $\{M_t\}_{a\le t\le b}$ be a continuous square-integrable $(\mathcal{F}_t)$-martingale. Then for each $\epsilon > 0$,*

$$P\Big[\sup_{a\le t\le b} |\, M_t \,| \ge \epsilon\Big] \le \frac{1}{\epsilon^2} \|\, M_b \,\|^2_{L2(\Omega)}.$$

Lemma 5.5. *Let $f_n, f \in \mathcal{V}[a,b]$ be such that*

$$f_n \xrightarrow{\|\cdot\|_2} f \qquad \text{in } \mathcal{V}[a,b].$$

Furthermore, let $g_n, g \in \mathcal{W}[a,b]$ be such that

$$g_n(\cdot,\omega) \xrightarrow{\|\cdot\|_1} g(\cdot,\omega) \qquad \text{in } L^1[a,b] \text{ a.e. } \omega.$$

Set

$$X_n(t) := \underbrace{\int_a^t f_n(s)\,dB_s}_{=:S_n(t)} + \underbrace{\int_a^t g_n(s)\,ds}_{=:Z_n(t)}$$

and

$$X(t) := \underbrace{\int_a^t f(s)\,dB_s}_{=:S(t)} + \underbrace{\int_a^t g(s)\,ds}_{=:Z(t)}.$$

Then

$$X_n(t) \to X(t) \quad \text{uniformly on } [a,b], \text{ in probability.}$$

Proof. First consider the deterministic integral part. For all $a \le t \le b$ and all ω,

$$|Z_n(t,\omega) - Z(t)(t,\omega)| \le \int_a^t |g_n(s,\omega) - g(s,\omega)|\,ds \le \int_a^b |g_n(s,\omega) - g(s,\omega)|\,ds$$

and hence

$$\|\, Z_n(\cdot,\omega) - Z(\cdot,\omega) \,\|_u = \sup_{a\le t\le b} |Z_n(t,\omega) - Z(t)(t,\omega)| \le \|\, g_n(\cdot,\omega) - g(\cdot,\omega) \,\|_1 \to 0$$

a.e. ω, by assumption. It follows by theorem 1.23 that

$$\| Z_n(\cdot,\omega) - Z(\cdot,\omega) \|_u \to 0 \qquad \text{in probability.}$$

Next consider the Itô integral part. By theorems 4.11 and 4.12, $\{S_n(t)\}_{a\le t\le b}$ and $\{S(t)\}_{a\le t\le b}$ are continuous, square integrable martingales, hence by theorem 3.5, so is each $\{S_n(t) - S(t)\}_{a\le t\le b}$. Thus for all $\epsilon > 0$,

$$\begin{aligned} P\Big[\| S_n(\cdot,\omega) - S(\cdot,\omega) \|_u \ge \epsilon \Big] &= P\Big[\sup_{a\le t\le b} |(S_n - S)(\cdot,\omega)| \ge \epsilon \Big] \\ &\le \frac{1}{\epsilon^2} \| (S_n - S)(b) \|^2_{L^2(\Omega)} && \text{(by theorem 5.4)} \\ &= \frac{1}{\epsilon^2} \Big\| \int_a^b (f_n - f)(s)\, dB_s \Big\|^2_{L^2(\Omega)} \\ &= \frac{1}{\epsilon^2} \| f_n - f \|^2_{L^2([a,b]\times\Omega)} \to 0 && \text{(Itô isometry)} \end{aligned}$$

as $n \to \infty$ by assumption. That is

$$\| S_n(\cdot,\omega) - S(\cdot,\omega) \|_u \to 0 \qquad \text{in probability.}$$

Finally, for each $\epsilon > 0$,

$$\begin{aligned} &P\Big[\| X_n(\cdot,\omega) - X(\cdot,\omega) \|_u \ge \epsilon \Big] \\ &\quad = P\Big[\| (S_n(\cdot,\omega) - S(\cdot,\omega)) + (Z_n(\cdot,\omega) - Z(\cdot,\omega)) \|_u \ge \epsilon \Big] \\ &\quad \le P\Big[\| S_n(\cdot,\omega) - S(\cdot,\omega) \|_u \ge \frac{\epsilon}{2} \quad \text{or} \quad \| Z_n(\cdot,\omega) - Z(\cdot,\omega) \|_u \ge \frac{\epsilon}{2} \Big] \\ &\quad \le P\Big[\| S_n(\cdot,\omega) - S(\cdot,\omega) \|_u \ge \frac{\epsilon}{2} \Big] + P\Big[\| Z_n(\cdot,\omega) - Z(\cdot,\omega) \|_u \ge \frac{\epsilon}{2} \Big] \to 0 \end{aligned}$$

as $n \to \infty$. This proves the theorem. □

Remark 5.5. Theorem 1.24 deals with pointwise convergence of functions only. However, if we replace each line "$|f_n(\omega) - f(\omega)|$" by "$\|f_n(\omega) - f(\omega)\|_u$" in the proof of theorem 1.24, we immediately obtain the following statement:

Let $\{f_n\}_{a\le t\le b}$, $\{f\}_{a\le t\le b}$ be $(\mathcal{F}_t)$-progressive processes. If

$$f_n(t) \to f(t) \quad \textit{uniformly on } [a,b], \textit{ in probability}$$

then there exists a subsequence $\{f_{n_k}\}_{k=1}^\infty$ such that

$$f_{n_k}(t) \to f(t) \quad \textit{uniformly on } [a,b], \textit{ a.s.}$$

Now let $\{X_n\}$ be as in lemma 5.5. From the above remark we conclude that there exists a subsequence $\{X_{n_k}\}$ such that

$$X_{n_k}(t) \to X(t) \quad \text{uniformly on } [a,b], \text{ a.s.}$$

Proof of Itô's formula.
We prove the formula under the additional condition that the partial derivatives F_y, F_x and F_{xx} are bounded, say

$$|F_y(y,x)| \leq M, \qquad |F_x(y,x)| \leq M, \qquad |F_{xx}(y,x)| \leq M$$

for some $M > 0$ and all x and all $a \leq y \leq b$. (This is to avoid introduction of stopping times.) This assumption guarantees in particular that condition (A) is satisfied.

So let

$$X(t) = X(a) + \int_a^t f(s)\,dB_s + \int_a^t g(s)\,ds \qquad (a \leq t \leq b) \tag{5.16}$$

be an Itô process, where $f \in \mathcal{V}[a,b]$ and $g \in \mathcal{W}[a,b]$. Set

$$Y(t) := F\big(t, X(t)\big).$$

We need to show that

$$\begin{aligned} Y(b) - Y(a) &= F\big(b, X(b)\big) - F\big(a, X(a)\big) \\ &= \int_a^b \big[F_y + F_x g + \frac{1}{2} F_{xx} f^2\big]\big(s, X(s)\big)\,ds + \int_a^b [F_x f]\big(s, X(s)\big)\,dB_s \quad \text{a.s.} \end{aligned} \tag{5.17}$$

(Note that f and g are functions which depend on s, but not on $X(s)$. For ease of notation, we have written $f(s, X(s)), \ldots$ above.) The assertion is obvious if $b = a$, hence we may assume that $b > a$.

Since F_y, F_x and F_{xx} are continuous, and $s \to X_s(\omega)$ is continuous for all ω, the compositions

$$(y,s) \mapsto F_y\big(y, X(s)\big), \qquad (y,s) \mapsto F_x\big(y, X(s)\big), \qquad (y,s) \mapsto F_{xx}\big(y, X(s)\big)$$

are continuous, as are the compositions

$$s \mapsto F_y\big(s, X(s)\big), \qquad s \mapsto F_x\big(s, X(s)\big), \qquad s \mapsto F_{xx}\big(s, X(s)\big).$$

We will make use of this continuity throughout the proof. Furthermore, the maps

$$\omega \mapsto F_y\big(s, X(s,\omega)\big), \qquad \omega \mapsto F_x\big(s, X(s,\omega)\big), \qquad \omega \mapsto F_{xx}\big(s, X(s,\omega)\big)$$

are $(\mathcal{F}_s)$-measurable, hence the three processes $\{F_y(s, X_s)\}_{a \leq s \leq b}$, $\{F_x(s, X_s)\}_{a \leq s \leq b}$ and $\{F_{xx}(s, X_s)\}_{a \leq s \leq b}$ are progressive. (This was already observed in remark 5.3.)

Case 1: f and g are constant random variables. That is, $f(s,\omega) = f(\omega)$ and $g(s,\omega) = g(\omega)$ are independent of s, are $\mathcal{F}_a$-measurable, and $f(\omega) \in L^2(\Omega, \mathcal{F}_a, P)$.

Choose a sequence $\{P_n\}$ partitions of $[a,b]$ with $\|P_n\| \to 0$ as $n \to \infty$, say

$$P_n = \big\{a = t_0 < t_1 < t_2 < \cdots < t_m = b\big\}$$

where as usual, we omit the index (n) and set

$$\Delta B_j = B_{t_{j+1}} - B_{t_j}, \qquad \Delta t_j = t_{j+1} - t_j.$$

(Since we already know that all integrals involved exist, we can choose the partitions as we wish.) Similarly, given a function $h(t)$ defined on $[a,b]$, we set

$$\Delta h_j = h_{t_{j+1}} - h_{t_j}.$$

For each partition P_n, we have the telescoping sum

$$Y(b) - Y(a) = \sum_{j=0}^{m-1} \Big[\underbrace{F\big(t_{j+1}, X(t_{j+1})\big) - F\big(t_j, X(t_j)\big)}_{=:\Delta F_j}\Big] = \sum_{j=0}^{m-1} \Delta F_j. \tag{5.18}$$

Let a point (y_o, x_o) be given. We apply the one-dimensional Taylor formula at y_o to obtain

$$F(y,x) = F(y_o, x) + \underbrace{F_y(\tilde{y}, x)(y - y_o)}_{\text{remainder term}}$$

for some point $\tilde{y}$ between y and y_o (depending on x). Now choose $x = x_o$ and apply the Taylor formula to the second variable to obtain

$$F(y_o, x) = F(y_o, x_o) + F_x(y_o, x_o)(x - x_o) + \underbrace{\frac{1}{2}F_{xx}(y_o, \hat{x})(x - x_o)^2}_{\text{remainder term}}$$

for some point $\hat{x}$ between x and x_o. Combine both equations,

$$F(y,x) = F(y_o, x_o) + F_y(\tilde{y}, x)(y - y_o) + F_x(y_o, x_o)(x - x_o) + \frac{1}{2}F_{xx}(y_o, \hat{x})(x - x_o)^2.$$

Applying this identity to $y_o = t_j$, $y = t_{j+1}$, $x_o = X(t_j)$ and $x = X(t_{j+1})$ we obtain

$$\Delta F_j = F_y\big(\tilde{t}_j, X(t_{j+1})\big)\Delta t_j + F_x\big(t_j, X(t_j)\big)\Delta X_j + \frac{1}{2}F_{xx}\big(\tilde{t}_j, X(\hat{t}_j)\big)(\Delta X_j)^2$$

for some $\tilde{t}_j$ and $\hat{t}_j$ between t_j and t_{j+1}. Here we have used the fact that $X(t)$ is continuous; the intermediate value theorem guarantees that there exists $\hat{t}_j$ between t_j and t_{j+1} so that $\hat{x} = X(\hat{t}_j)$. Observe that $\tilde{t}_j$ and $\hat{t}_j$ depend on ω ! Equation (5.18) can now be written as

$$\begin{aligned} Y(b) - Y(a) &= \underbrace{\sum_{j=0}^{m-1} F_y\big(\tilde{t}_j, X(t_{j+1})\big)\Delta t_j}_{(I)} \\ &\quad + \underbrace{\sum_{j=0}^{m-1} F_x\big(t_j, X(t_j)\big)\Delta X_j}_{(II)} + \underbrace{\frac{1}{2}\sum_{j=0}^{m-1} F_{xx}\big(t_j, X(\hat{t}_j)\big)(\Delta X_j)^2}_{(III)}. \end{aligned} \tag{5.19}$$

We need to show: $(I) + (II) + \frac{1}{2}(III) \to$ right-hand side of (5.17) a.s. as $n \to \infty$. We consider each sum separately.

Claim:

$$\lim_{n\to\infty} (I) = \int_a^b F_y\big(s, X(s)\big)\,ds \qquad \forall\omega. \tag{5.20}$$

Since this is a deterministic integral, we let ω be arbitrary, but fixed. By continuity of $s \to F_y\big(s, X(s)\big)$, the above integral is in fact a limit of Riemann sums,

$$\int_a^b F_y\big(s, X(s)\big)\,ds = \lim_{n\to\infty} \underbrace{\sum_{j=0}^{m-1} F_y\big(t_j, X(t_j)\big)\Delta t_j}_{(Ia)}$$

So if we show that

$$\lim_{n\to\infty} \big|\,(I) - (Ia)\,\big| = 0, \tag{5.21}$$

then the triangle inequality will yield

$$\left|(I) - \int_a^b F_y\big(s, X(s)\big)\,ds\right| \le \big|\,(I) - (Ia)\,\big| + \left|(Ia) - \int_a^b F_y\big(s, X(s)\big)\,ds\right| \to 0$$

as $n \to \infty$, and the claim will be proved.

To prove (5.21), we consider

$$\begin{aligned}\big|\,(I) - (Ia)\,\big| &= \left|\sum_{j=0}^{m-1} F_y\big(\tilde{t}_j, X(t_{j+1})\big)\Delta t_j - \sum_{j=0}^{m-1} F_y\big(t_j, X(t_j)\big)\Delta t_j\right| \\ &\le \sum_{j=0}^{m-1} \Big|\underbrace{F_y\big(\tilde{t}_j, X(t_{j+1})\big) - F_y\big(t_j, X(t_j)\big)}_{(b)}\Big|\Delta t_j.\end{aligned}$$

Let $\epsilon > 0$ be given. Since $(y, s) \to F_y(y, X(s))$ is continuous on $[a, b] \times [a, b]$, it is uniformly continuous. Thus there exists $\delta > 0$ such that

$$|y_1 - y_2| < \delta \qquad \text{and} \qquad |s_1 - s_2| < \delta$$

imply

$$\Big|\,F_y\big(y_1, X(s_1)\big) - F_y\big(y_2, X(s_2)\big)\,\Big| < \frac{\epsilon}{b-a}.$$

Now if $\|P_n\| < \delta$ then we obtain from

$$|\tilde{t}_j - t_j| < \delta \qquad \text{and} \qquad |t_{j+1} - t_j| < \delta$$

that

$$\big|\,(b)\,\big| < \frac{\epsilon}{b-a}.$$

and hence

$$\big|\,(I) - (Ia)\,\big| \le \sum_{j=0}^{m-1} |(b)|\,\Delta t_j < \sum_{j=0}^{m-1} \frac{\epsilon}{b-a}\Delta t_j = \frac{\epsilon}{b-a}\sum_{j=0}^{m-1} \Delta t_j = \epsilon$$

provided that $\|P_n\| < \delta$. As ϵ was arbitrary, then (5.21) and hence the claim follow.

Claim:

$$\lim_{n\to\infty} (II) = \int_a^b F_x\big(s, X(s)\big) f(s)\, dB_s + \int_a^b F_x\big(s, X(s)\big) g(s)\, ds \qquad \text{a.e. } \omega. \tag{5.22}$$

In fact, since $X(t)$ is an Itô process, then

$$\begin{aligned} \Delta X_j = X(t_{j+1}) - X(t_j) &= \int_{t_j}^{t_{j+1}} f(s)\, dB_s + \int_{t_j}^{t_{j+1}} g(s)\, ds \\ &= f\Delta B_j + g\Delta t_j \qquad \text{a.e. } \omega \end{aligned} \tag{5.23}$$

where we have used the fact that f and g are constant with respect to s. Hence,

$$(II) = \underbrace{\sum_{j=0}^{m-1} F_x\big(t_j, X(t_j)\big) f\, \Delta B_j}_{(c1)} + \underbrace{\sum_{j=0}^{m-1} F_x\big(t_j, X(t_j)\big) g\, \Delta t_j}_{(c2)} \qquad \text{a.e. } \omega.$$

Now as $\{F_x(s, X_s)\}_{a\le s\le b}$ is $(\mathcal{F}_s)$-progressive, $f(\omega) \in L^2(\Omega, \mathcal{F}_a, P)$ and F_x is bounded, then $(s,\omega) \to F_x(s, X(s,\omega)) f(\omega) \in L^2([a,b]\times\Omega)$ and is also $(\mathcal{F}_s)$-progressive, that is, constitutes an element of $\mathcal{V}[a,b]$. Hence by exercise 5.1,

$$(c1) \xrightarrow{\|\cdot\|_2} \int_a^b F_x\big(s, X(s)\big) f\, dB_s$$

as $n \to \infty$. Replacing the sequence of partitions $\{P_n\}$ by as suitable subsequence, we have

$$(c1) \to \int_a^b F_x\big(s, X(s)\big) f\, dB_s \qquad \text{a.e. } \omega.$$

On the other hand, for each ω, $(c2)$ is simply a Riemann sum (recall here that $s \to F_x\big(s, X(s)\big)$ is continuous and g is constant with respect to s), hence

$$(c2) \to \int_a^b F_x\big(s, X(s)\big) g\, ds \qquad \forall\omega$$

as $n \to \infty$. This proves the claim.

Claim:

$$\lim_{n\to\infty} (III) = \int_a^b F_{xx}\big(s, X(s)\big) f(s)^2\, ds \qquad \text{a.e. } \omega. \tag{5.24}$$

This claim takes a fair amount of work to prove. In fact, by continuity of $s \to F_{xx}\big(s, X(s)\big)$, this integral is a limit of Riemann sums,

$$\int_a^b F_{xx}\big(s, X(s)\big) f(s)^2\, ds = \lim_{n\to\infty} \underbrace{\sum_{j=0}^{m-1} F_{xx}\big(t_j, X(t_j)\big) f^2\, \Delta t_j}_{(IIIa)}.$$

So if we can show that

$$\lim_{n\to\infty} \big|\, (III) - (IIIa) \,\big| = 0 \qquad \text{a.e. } \omega \tag{5.25}$$

then an application of the triangle inequality will yield the claim, just as in the case of sum (I) above.

To prove (5.25), we consider

$$\begin{aligned} \big|\, (III) - (IIIa) \,\big| &= \left| \sum_{j=0}^{m-1} F_{xx}(t_j, X(\hat{t}_j))(\Delta X_j)^2 - \sum_{j=0}^{m-1} F_{xx}(t_j, X(t_j)) f^2 \,\Delta t_j \right| \\ &\le \underbrace{\left| \sum_{j=0}^{m-1} \Big\{ F_{xx}(t_j, X(\hat{t}_j)) - F_{xx}(t_j, X(t_j)) \Big\} (\Delta X_j)^2 \right|}_{(d1)} \\ &\qquad + \underbrace{\left| \sum_{j=0}^{m-1} F_{xx}(t_j, X(t_j)) \Big[(\Delta X_j)^2 - f^2 \Delta t_j \Big] \right|}_{(d2)}. \end{aligned}$$

First consider $(d1)$. Here, we let ω be arbitrary but fixed, as we will make use of continuity of sample paths. Let $\epsilon > 0$ be given. Set

$$\hat{\epsilon} := \frac{\epsilon}{(f(\omega)^2 + 1)(b-a)}.$$

Now uniform continuity of the map $(y, s) \mapsto F_{xx}(y, X(s))$ on $[a, b] \times [a, b]$ implies that there exists $\delta > 0$ such that

$$|y_1 - y_2| < \delta \qquad \text{and} \qquad |s_1 - s_2| < \delta$$

imply

$$\Big| F_{xx}(y_1, X(s_1)) \;-\; F_{xx}(y_2, X(s_2)) \Big| < \hat{\epsilon}.$$

Now if $\|P_n\| < \delta$ then we obtain from

$$|\hat{t}_j - t_j| < \delta$$

that

$$\big|\, F_{xx}(t_j, X(\hat{t}_j)) - F_{xx}(t_j, X(t_j)) \,\big| < \hat{\epsilon}$$

and hence

$$\big|\, (d1) \,\big| \le \sum_{j=0}^{m-1} \Big| F_{xx}(t_j, X(\hat{t}_j)) - F_{xx}(t_j, X(t_j)) \Big| (\Delta X_j)^2 < \hat{\epsilon} \sum_{j=0}^{m-1} (\Delta X_j)^2. \tag{5.26}$$

We thus need to estimate the right-hand sum. Note that by (5.23)

$$\sum_{j=0}^{m-1} (\Delta X_j)^2 = \underbrace{\sum_{j=0}^{m-1} f^2 (\Delta B_j)^2}_{(e1)} + 2 \underbrace{\sum_{j=0}^{m-1} f g \Delta t_j \Delta B_j}_{(e2)} + \underbrace{\sum_{j=0}^{m-1} g^2 (\Delta t_j)^2}_{(e3)}. \tag{5.27}$$

Consider $(e1)$. By theorem 5.3, we have

$$\sum_{j=0}^{m-1} (\Delta B_j)^2 \xrightarrow{\|\cdot\|_2} \int_a^b 1\, ds = (b-a).$$

Replacing the sequence of partitions $\{P_n\}$ by a suitable subsequence, this becomes a.e. convergence as well, and hence

$$(e1) = f^2 \sum_{j=0}^{m-1} (\Delta B_j)^2 \to f^2(b-a) \qquad \text{a.e. } \omega$$

as $n \to \infty$.

Next consider $(e2)$. For future use, let use consider more general sums of the form

$$\sum_{j=0}^{m-1} H(t_j)\Delta t_j \Delta B_j$$

where $\{H(t)\}_{a \le t \le b}$ is an $(\mathcal{F}_t)$-adapted and bounded process. For each n, set

$$f_n := \sum_{j=0}^{m-1} H(t_j)\Delta t_j \mathbf{1}_{(t_j, t_{j+1}]} \in \mathcal{V}[a,b].$$

Since H is bounded, then $f_n \to 0$ uniformly on $[a,b] \times \Omega$, and hence $f_n \xrightarrow{\|\cdot\|_2} 0$ in $\mathcal{V}[a,b]$. Hence by the Itô isometry,

$$\sum_{j=0}^{m-1} H(t_j)\Delta t_j \Delta B_j = \int_a^b f_n(s)\, ds \xrightarrow{\|\cdot\|_2} \int_a^b 0 = 0$$

in $L^2(\Omega)$. Replacing $\{P_n\}$ by a suitable subsequence, this becomes a.e convergence as well. Now choosing $H = 1$ we obtain

$$(e2) = fg \sum_{j=0}^{m-1} \Delta t_j \Delta B_j \to 0 \qquad \text{a.e. } \omega.$$

Last we consider $(e3)$. A simple calculation gives

$$(e3) = g^2 \sum_{j=0}^{m-1} (\Delta t_j)^2 \le g^2 \|P_n\| \sum_{j=0}^{m-1} \Delta t_j = g^2 \|P_n\|(b-a) \to 0$$

as $n \to \infty$, for each ω. Combining $(e1)$, $(e2)$ and $(e3)$ we obtain from (5.26) that

$$\begin{aligned} \limsup_n \left|(d1)\right| &\le \hat{\epsilon} \limsup_n \left[(e1) + (e2) + (e3)\right] \\ &= \hat{\epsilon}\left[f^2(\omega)(b-a) + 0 + 0\right] < \epsilon \end{aligned}$$

for a.e. ω. As ϵ was arbitrary, we conclude that

$$|(d1)| \to 0 \qquad \text{a.e. } \omega \tag{5.28}$$

as $n \to \infty$.

The estimate for $(d2)$ is done similarly. By (5.23) and the triangle inequality we have

$$\begin{aligned} |(d2)| \leq\ & f^2 \Bigg| \underbrace{\sum_{j=0}^{m-1} F_{xx}\big(t_j, X(t_j)\big)\,(\Delta B_j)^2}_{(k1)} - \underbrace{\sum_{j=0}^{m-1} F_{xx}\big(t_j, X(t_j)\big)\,\Delta t_j}_{(k2)} \Bigg| \\ & + \Bigg| 2fg \underbrace{\sum_{j=0}^{m-1} F_{xx}\big(t_j, X(t_j)\big)\,\Delta t_j \Delta B_j}_{(k3)} \Bigg| + \Bigg| g^2 \underbrace{\sum_{j=0}^{m-1} F_{xx}\big(t_j, X(t_j)\big)\,(\Delta t_j)^2}_{(k4)} \Bigg| \end{aligned}$$

for all ω. Now by theorem 5.3,

$$(k1) \xrightarrow{\|\cdot\|_2} \int_a^b F_{xx}\big(s, X(s)\big)\,ds$$

and replacing $\{P_n\}$ by a suitable subsequence, this is also a.e. convergence. Now $(k2)$ is precisely a Riemann sum for this integral, hence

$$(k2) \to \int_a^b F_{xx}\big(s, X(s)\big)\,ds \qquad \forall \omega$$

so that $(k1) - (k2) \to 0$ a.e. Applying the argument for $(e2)$ above (to the process $H(t) = F_{xx}\big(t, X(t)\big)$) we obtain that (replacing $\{P_n\}$ by a suitable subsequence)

$$(k3) \to 0 \qquad \text{a.e. } \omega.$$

Finally,

$$|(k4)| \leq \sum_{j=0}^{m-1} M(\Delta t_j)^2 \leq M\|P_n\| \sum_{j=0}^{m-1} \Delta t_j = M\|P_n\|(b-a) \to 0 \qquad \forall \omega$$

as $n \to \infty$. Combining the last four convergence results, we conclude that

$$(d2) \to 0 \qquad \text{a.e. } \omega$$

as $n \to \infty$. Together with (5.28) we now obtain that (5.25) holds; thus the claim holds.

By combining all three claims the theorem is now proved in the case of constant f and g.

Case 2: f and g are $(\mathcal{F}_t)$-adapted simple functions. That is (choosing a common refinement) there exists a partition

$$Q = \big\{a = u_o < u_1 < u_2 < \cdots < u_k = b\big\}$$

of $[a, b]$ such that

$$\begin{aligned} f &= \sum_{i=0}^{k-1} f_i \,\mathbf{1}_{(u_i, u_{i+1}]}, \qquad && f_i \in L^2(\Omega, \mathcal{F}_{u_i}, P) \\ g &= \sum_{i=0}^{k-1} g_i \,\mathbf{1}_{(u_i, u_{i+1}]}, \qquad && g_i \text{ is } \mathcal{F}_{u_i}\text{-measurable.} \end{aligned} \tag{5.29}$$

Then by case 1, for a.e. ω,

$$\begin{aligned} Y(b) - Y(a) &= \sum_{i=0}^{k-1} \Big[Y(u_{i+1}) - Y(u_i) \Big] \\ &= \sum_{i=0}^{k-1} \left\{ \int_{u_i}^{u_{i+1}} \big[F_y + F_x g_i + \frac{1}{2} F_{xx} f_i^2\big]\big(s, X(s)\big)\, ds + \int_{u_i}^{u_{i+1}} [F_x f_i]\big(s, X(s)\big)\, dB_s \right\} \\ &= \sum_{i=0}^{k-1} \left\{ \int_{u_i}^{u_{i+1}} \big[F_y + F_x g + \frac{1}{2} F_{xx} f^2\big]\big(s, X(s)\big)\, ds + \int_{u_i}^{u_{i+1}} [F_x f]\big(s, X(s)\big)\, dB_s \right\} \\ &= \int_a^b \big[F_y + F_x g + \frac{1}{2} F_{xx} f^2\big]\big(s, X(s)\big)\, ds + \int_a^b [F_x f]\big(s, X(s)\big)\, dB_s \end{aligned}$$

where we have used the fact that all integrals are independent of the values of f and g at the partition points u_i, and that f and g are constant on the interior of the partition intervals.

Case 3: General f and g. The idea is to approximate f and g by sequences of simple functions, and to show that all integrals converge as required. Since $f \in \mathcal{V}[a, b]$, there exists a sequence $\{f_n\}$ of simple functions in $\mathcal{V}[a, b]$ with

$$\| f - f_n \|_{L^2([a,b]\times\Omega)} \to 0.$$

Similarly, since $g \in \mathcal{W}[a, b]$, there exists a sequence $\{g_n\}$ of simple functions in $\mathcal{W}[a, b]$ with

$$\| g(\cdot, \omega) - g_n(\cdot, \omega) \|_{L^1[a,b]} \to 0 \qquad \text{a.e. } \omega.$$

Now consider the Itô processes

$$X_n(t) := X(a) + \int_a^t f_n(s)\, dB_s + \int_0^t g_n(s)\, ds$$

and

$$X(t) := X(a) + \int_a^t f(s)\, dB_s + \int_0^t g(s)\, ds$$

for $a \le t \le b$. By case 2, for each n,

$$Y_n(b) - Y_n(a) = F\big(b, X_n(b)\big) - F\big(a, X_n(a)\big) \tag{5.30}$$

$$= \int_a^b \big[F_y + F_x g_n + \frac{1}{2} F_{xx} f_n^2\big]\big(s, X_n(s)\big)\, ds + \int_a^b [F_x f_n]\big(s, X_n(s)\big)\, dB_s. \tag{5.31}$$

We must show:

$$
\begin{array}{lll}
(5.30) & \to & \text{left-hand side of (5.17) a.e. } \omega \\
(5.31) & \to & \text{right-hand side of (5.17) a.e. } \omega.
\end{array}
$$

Consider first equation (5.30). Replacing $\{f_n\}$ and $\{g_n\}$ by subsequences if necessary, remark 5.5 guarantees that for almost all ω,

$$X_n(t) \to X(t) \qquad \text{uniformly on } [a,b]. \tag{5.32}$$

Since uniform convergence implies pointwise convergence and by continuity of F,

$$
\begin{aligned}
Y_n(b) - Y_n(a) = F\big(b, X_n(b)\big) - F\big(a, X_n(a)\big) \\
\to F\big(b, X(b)\big) - F\big(a, X(a)\big). = Y(b) - Y(a)
\end{aligned}
$$

Next consider equation (5.31). We discuss the two integrals separately. By continuity of F_y, F_x, F_{xx} and $X(t)$ and uniform convergence (5.32), it follows that

$$
\begin{array}{rcl}
F_y\big(s, X_n(s)\big) & \to & F_y\big(s, X(s)\big) \\
F_x\big(s, X_n(s)\big) & \to & F_x\big(s, X(s)\big) \\
F_{xx}\big(s, X_n(s)\big) & \to & F_{xx}\big(s, X(s)\big)
\end{array}
$$

uniformly on $[a,b]$, for almost all ω. Hence the corresponding sequences of integrals will converge, for example,

$$
\begin{aligned}
&\left| \int_a^b F_y\big(s, X_n(s)\big)\, ds - \int_a^b F_y\big(s, X(s)\big)\, ds \right| \\
&\qquad \le \int_a^b \underbrace{\big\| F_y\big(s, X_n(s)\big) - F_y\big(s, X(s)\big) \big\|_u}_{=:M_n \to 0 \text{ a.e. } \omega} ds = M_n(b-a) \to 0 \quad \text{a.e. } \omega.
\end{aligned} \tag{5.33}
$$

Similar computations hold for F_x and F_{xx} with constants K_n and L_n (depending on ω of course), respectively.

Consider first the deterministic integral in (5.31). Since

$$\int_\Omega \int_a^b (f_n - f)^2\, ds\, dP = \big\| f_n - f \big\|_{L^2([a,b]\times\Omega)} \to 0$$

then (replacing $\{f_n\}$ by a subsequence if necessary)

$$\big\| f_n(\cdot,\omega) - f(\cdot,\omega) \big\|^2_{L^2[a,b]} = \int_a^b \big[f_n(s,\omega) - f(s,\omega)\big]^2 ds \to 0 \qquad \text{a.e. } \omega$$

so by continuity of the norm,

$$\big\| f_n(\cdot,\omega) \big\|^2_{L^2[a,b]} \to \big\| f(\cdot,\omega) \big\|^2_{L^2[a,b]} \qquad \text{a.e. } \omega.$$

We have (note that all terms also depend on ω !),

$$
\begin{aligned}
\Big| \int_a^b & [F_y + F_x g_n + \frac{1}{2} F_{xx} f_n^2]\big(s, X_n(s)\big)\, ds - \int_a^b [F_y + F_x g + \frac{1}{2} F_{xx} f^2]\big(s, X(s)\big)\, ds \Big| \\
&\le \int_a^b \Big| F_y\big(s, X_n(s)\big) - F_y\big(s, X(s)\big) \Big|\, ds \\
&\quad + \int_a^b \Big| F_x\big(s, X_n(s)\big) \Big|\, \big| g_n(s) - g(s) \big|\, ds \\
&\quad + \int_a^b \Big| F_x\big(s, X_n(s)\big) - F_x\big(s, X(s)\big) \Big|\, \big| g(s) \big|\, ds \\
&\quad + \frac{1}{2} \int_a^b \Big| F_{xx}\big(s, X_n(s)\big) \Big|\, \big| f_n(s)^2 - f(s)^2 \big|\, ds \\
&\quad + \int_a^b \Big| F_{xx}\big(s, X_n(s)\big) - F_{xx}\big(s, X(s)\big) \Big|\, \big| f(s)^2 \big|\, ds \\
&\le M_n(b-a) + M \, \|g_n - g\|_{L^1[a,b]} + K_n \|g\|_{L^1[a,b]} \\
&\quad + \frac{M}{2} \Big[\|f_n\|^2_{L^2[a,b]} - \|f\|^2_{L^2[a,b]} \Big] + \frac{L_n}{2} \|f\|^2_{L^2[a,b]} \to 0 \qquad \text{a.e. } \omega.
\end{aligned}
$$

Thus, the deterministic integral in (5.31) converges to the deterministic integral on the right side of (5.17) a.s.

Now look at the Itô integral in (5.31). By linearity of the integral and Itô isometry,

$$
\begin{aligned}
\Big\| \int_a^b & [F_x f_n]\big(s, X_n(s)\big)\, dB_s - \int_a^b [F_x f]\big(s, X(s)\big)\, dB_s \Big\|_{L^2(\Omega)} \\
&= \Big\| F_x\big(s, X_n(s)\big) f_n(s) - F_x\big(s, X(s)\big) f(s) \Big\|_{L^2([a,b]\times\Omega)} \\
&\le \Big\| F_x\big(s, X_n(s)\big) [f_n(s) - f(s)] \Big\|_{L^2([a,b]\times\Omega)} \\
&\qquad + \Big\| \Big[F_x\big(s, X_n(s)\big) - F_x\big(s, X(s)\big) \Big] f(s) \Big\|_{L^2([a,b]\times\Omega)} \\
&\le M \big\| f_n - f \big\|_{L^2([a,b]\times\Omega)} + K_n \big\| f \big\|_{L^2([a,b]\times\Omega)} \to 0
\end{aligned}
$$

as $n \to \infty$. Replacing $\{f_n\}$ by a suitable subsequence, it follows that

$$
\int_a^b [F_x f_n]\big(s, X_n(s)\big)\, dB_s \to \int_a^b [F_x f]\big(s, X(s)\big)\, dB_s \qquad \text{a.e. } \omega.
$$

as $n \to \infty$. That is, the stochastic integral in (5.31) converges to the stochastic integral on the right side of (5.17) a.s. The theorem is now proved. □

Chapter 6

Stochastic Differential Equations

In this chapter we apply Itô's formula to find solutions of stochastic differential equation. For lack of time, we will touch on the theory only briefly, and focus on particularly simple linear equations.

Throughout this chapter, $(\Omega, \mathcal{F}, P)$ will denote a *complete* probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. We assume that F_0 (and hence each F_t) contains all null sets. Furthermore, $\{B_t\}_{t\geq 0}$ will be *standard* $(\mathcal{F}_t)$-Brownian motion.

6.1 Existence and Uniqueness of Solutions

We consider the *stochastic differential equation* (SDE)

$$dX_t = f(t, X_t)\, dB_t + g(t, X_t)\, dt \tag{6.1}$$

which is to be understood as the integral equation

$$X_t = X_0 + \int_0^t f(s, X_s)\, dB_s + \int_0^t g(s, X_s)\, ds. \tag{6.2}$$

with the initial value X_0 assumed to be $\mathcal{F}_0$-measurable.

Definition 6.1. By a solution of (6.1) we mean a continuous $(\mathcal{F}_t)$-adapted (hence progressive) process $\{X_t\}_{t\geq 0}$ such that (6.2) holds a.s., for all $t \geq 0$.

We say that *pathwise uniqueness* holds if whenever $\{X_t\}_{t\geq 0}$ and $\{\widetilde{X}_t\}_{t\geq 0}$ are two solutions of (6.1) satisfying the same initial condition, $X_0 = \widetilde{X}_0$ a.s., then

$$X_t = \widetilde{X}_t \quad \forall\, t \geq 0 \qquad \text{a.s.}$$

(That is $\{\omega : X_t(\omega) \neq \widetilde{X}_t(\omega) \text{ for some } t\}$ is a null set.)

Theorem 6.1. (Global existence and uniqueness) *Suppose,*

1. $f(t, x)$ *and* $g(t, x)$ *are Borel measurable on* $[0, T] \times \mathbb{R}$,

2. *There exists $K > 0$ such that for all $t \in [0, T]$ and $x, y \in \mathbb{R}$,*

$$|f(t,x) - f(t,y)| + |g(t,x) - g(t,y)| \leq K|x-y| \tag{6.3}$$

("Lipschitz condition") and

$$f(t,x)^2 + g(t,x)^2 \leq K^2(1+x^2) \tag{6.4}$$

("growth condition").

3. $X_0 \in L^2(\Omega, \mathcal{F}_0, P)$.

Then there exists a solution $X(t)$ of (6.1) defined on $[0,T]$ with $X(t) \in L^2(\Omega)$ for all t, and in fact,

$$\sup_{0 \leq t \leq T} \|X(t)\|_2 < \infty. \tag{6.5}$$

Furthermore, this solution is pathwise unique.

Proof. See [4], theorem 3.1. □

Remark 6.1. 1. Inequality (6.5) implies that $X(t) \in \mathcal{V}[0,T]$. Furthermore, together with growth condition (6.4) it implies that $f(s, X(s))$ and $g(s, X(s))$ satisfy the conditions of definition 5.1, so that $\{X_t\}$ is an Itô process.

2. One can show that the growth condition is not required, but the proof is difficult.

6.2 Reducing Stochastic Differential Equations

A common technique used to solve differential equations is the transformation of an equation to a simpler equation by a change of variables. This idea applies to stochastic differential equations as well, as we shall see now.

Consider a stochastic differential equation

$$dX_t = f(t, X_t)\, dB_t + g(t, X_t)\, dt \tag{6.6}$$

where $f \neq 0$ and f_t and f_{xx} exist and are continuous. Our goal is to transform it into a simpler equation

$$dY_t = \tilde{f}(t)\, dB_t + \tilde{g}(t)\, dt \tag{6.7}$$

(which can be solved through integration) by a change of variables

$$y = h(t,x) \qquad (t \geq 0,\ x, y \in \mathbb{R}) \tag{6.8}$$

where $x \mapsto h(t,x)$ is an invertible, sufficiently smooth function for each t, say its inverse is

$$x = k(t,y). \tag{6.9}$$

Then

$$y = h(t, k(t,y)) \qquad \text{and} \qquad x = k(t, h(t,x)).$$

So let a change of variables (6.8) be given. Set

$$Y(t) := h\big(t, X(t)\big).$$

Then by the general Itô formula,

$$Y(t) = Y(0) + \int_0^t \tilde{f}\big(s, X(s)\big)\, dB_s + \int_0^t \tilde{g}\big(s, X(s)\big)\, ds \tag{6.10}$$

where

$$\tilde{f}(t,x) = \big[h_x f\big](t,x) \tag{6.11}$$

$$\tilde{g}(t,x) = \big[h_t + h_x g + \frac{1}{2} h_{xx} f^2\big](t,x). \tag{6.12}$$

(Throughout, we will assume that condition (A) is satisfied.)

Definition 6.2. Equation (6.6) is called *reducible* if there exists h as in (6.8) such that the functions $\tilde{f}$ and $\tilde{g}$ are *independent* of x, for (t,x) in some open subset D of the right half-plane plane $\{(t,x) : t \geq 0\}$ and where $f \neq 0$ on D.

Suppose, (6.6) is reducible by some h. That is, (6.11) and (6.12) are of the form

$$\tilde{f}(t) = \big[h_x f\big](t,x) \tag{6.13}$$

$$\tilde{g}(t) = \big[h_t + h_x g + \frac{1}{2} h_{xx} f^2\big](t,x) \tag{6.14}$$

on D. Differentiate (6.14) with respect to x,

$$h_{tx} + \frac{\partial}{\partial x}\big[h_x g + \frac{1}{2} h_{xx} f^2\big] = 0. \tag{6.15}$$

On the other hand, by (6.13),

$$h_x(t,x) = \frac{\tilde{f}(t)}{f(t,x)} \tag{6.16}$$

as f is continuous and does not vanish on D. Differentiating further,

$$h_{tx} = \frac{\tilde{f}'(t) f(t,x) - \tilde{f}(t) f_t(t,x)}{f(t,x)^2}$$

and also

$$h_{xx} = \frac{-\tilde{f}(t) f_x(t,x)}{f(t,x)^2}.$$

Substitute the last three equations into (6.15),

$$\frac{\tilde{f}' f - \tilde{f} f_t}{f^2} + \frac{\partial}{\partial x}\left(\frac{\tilde{f} g}{f} - \frac{1}{2}\tilde{f} f_x\right) = 0. \tag{6.17}$$

As $\tilde{f}$ is independent of x, this equation is equivalent to

$$\frac{\tilde{f}'}{f} - \frac{\tilde{f} f_t}{f^2} + \tilde{f}\frac{\partial}{\partial x}\Big(\frac{g}{f}\Big) - \frac{1}{2}\tilde{f} f_{xx} = 0$$

or

$$\frac{\tilde{f}'}{\tilde{f}} = \underbrace{f\Big[\frac{f_t}{f^2} - \frac{\partial}{\partial x}\Big(\frac{g}{f}\Big) + \frac{1}{2}f_{xx}\Big]}_{=:M(t)} \tag{6.18}$$

Note here that $\tilde{f}(t) \neq 0$ as $f(t,x) \neq 0$ on D and $h(t,\cdot)$ is invertible. Now since the left-hand side of (6.18) is independent of x, the right-hand side must also be constant with respect to x.

Conversely, suppose the right-hand side of (6.18) is defined and independent of x for (t,x) in some open set D. Then solving the differential equation

$$\frac{\tilde{f}'}{\tilde{f}} = M(t) \tag{6.19}$$

(which always has a nonzero solution by assumptions on f and g) we obtain a function $\tilde{f}(t)$ which does not vanish on some open interval and satisfies (6.18). Using equation (6.16) we have now found a pair $\tilde{f}$ and h_x satisfying (6.13). Observe that since $h_x(t,x) \neq 0$, any choice of h must be (at least locally) invertible with respect to the variable x.

Now we can find $\tilde{g}$: Since equations (6.17) and (6.18) are equivalent – as $\tilde{f}$ is a function of t only – then (6.17) holds. But the left-hand side of (6.17) was obtained by differentiating the right-hand side of (6.14) with respect to x and using identity (6.16) which holds by our choice of h_x; hence

$$\frac{d}{dx}\big[h_t + h_x g + \frac{1}{2}h_{xx}f^2\big] = 0.$$

That is, if we use equation (6.14) to define $\tilde{g}$, then $\tilde{g}$ will be a function of t only, for all possible choices of h satisfying (6.16).

We have shown:

Theorem 6.2. *Equation (6.6) is reducible if and only if*

$$f\Big[\frac{f_t}{f^2} - \frac{\partial}{\partial x}\Big(\frac{g}{f}\Big) + \frac{1}{2}f_{xx}\Big] \tag{6.20}$$

is independent of x.

Example 6.1. (Autonomous equation.) Consider an equation

$$dX_t = f(X_t)\,dB_t + g(X_t)\,dt \tag{6.21}$$

where f and g are independent of t. Such an equation is called *autonomous*. By theorem 6.2, equation (6.21) is reducible $\Leftrightarrow$

$$M(t,x) := f\Big[\frac{f_t}{f^2} - \frac{\partial}{\partial x}\Big(\frac{g}{f}\Big) + \frac{1}{2}f_{xx}\Big]$$

is independent of x for (t,x) in some open set. However, as f and g do not depend on t, then M cannot depend on t either and hence $f_t = 0$; it follows that (6.21) is reducible $\Leftrightarrow$

$$M = f\left[-\frac{\partial}{\partial x}\left(\frac{g}{f}\right) + \frac{1}{2}f_{xx}\right]$$

is constant for x in some open interval I.

Suppose this is the case. Then solving (6.19) we obtain one particular solution

$$\boxed{\tilde{f}(t) = e^{Mt}}.$$

Then by (6.16),

$$h_x = \frac{e^{Mt}}{f(x)}.$$

Integrate,

$$\boxed{h(t,x) = e^{Mt}\int_a^x \frac{ds}{f(s)}} \tag{6.22}$$

where $a \in I$ is arbitrary. Now $\tilde{g}$ is computed by (6.14),

$$\boxed{\tilde{g} = e^{Mt}\left[M\int_a^x \frac{ds}{f(s)} + \frac{g(x)}{f(x)} - \frac{1}{2}f'(x)\right]}.$$

6.3 Linear Equations

A stochastic differential equation of the form

$$dX_t = \left[f_1(t) + f_2(t)X_t\right] dB_t + \left[g_1(t) + g_2(t)X_t\right] dt \tag{6.23}$$

is called *linear*. So here,

$$\begin{aligned} f(t,x) &= f_1(t) + f_2(t)x \\ g(t,x) &= g_1(t) + g_2(t)x. \end{aligned}$$

Equation (6.23) is called

- *homogeneous* if $f_1(t) = f_2(t) = 0$,
- *narrow-sense* if $f_2(t) = 0$.

Suppose also that we are given an initial condition $X(0) = X_0 \in L^2(\Omega, \mathcal{F}_0, P)$.

In general, a linear equation is not reducible. In fact the condition of theorem 6.2 becomes

$$f\left[\frac{f_1'(t) + f_2'(t)x}{f(t,x)^2} - \frac{\partial}{\partial x}\left(\frac{g_1(t) + g_2(t)x}{f(t,x)}\right)\right] \quad \text{is independent of } x,$$

since $f_{xx} = 0$. Computing the derivative, this is equivalent to

$$\frac{f_1'(t) + f_2'(t)x}{f(t,x)} - \frac{g_2(t)[f_1(t) + f_2(t)x] - \big[g_1(t) + g_2(t)x\big] f_2(t)}{f(t,x)} \quad \text{is independent of } x,$$

or

$$M(t,x) := \frac{f_1'(t) + f_2'(t)x - f_1(t)g_2(t) + f_2(t)g_1(t)}{f_1(t) + f_2(t)x} \quad \text{is independent of } x. \tag{6.24}$$

Differentiating, this equation is equivalent to

$$f_2' f_1 - (f_1' - f_1 g_2 + f_2 g_1) f_2 = 0.$$

This is obviously not satisfied in general. Let us consider two special cases:

Case 1: The equation (6.23) is *homogeneous*,

$$dX_t = f_2(t) X_t \, dB_t + g_2(t) X_t \, dt$$

with $f_2 \in C^1[0,\infty)$ and $g_2 \in L^1[0,t]$ for all $t \geq 0$. Then the fraction in (6.24) becomes

$$M(t,x) = \frac{f_2'(t)}{f_2(t)}$$

and is independent of x. That is, a homogeneous linear equation is reducible. Equation (6.19) now becomes

$$\frac{\tilde{f}'(t)}{\tilde{f}(t)} = M(t) = \frac{f_2'(t)}{f_2(t)};$$

hence we can choose

$$\boxed{\tilde{f}(t) = f_2(t)}.$$

Equation (6.16) gives us

$$h_x = \frac{\tilde{f}(t)}{f(t,x)} = \frac{f_2(t)}{f_2(t)x} = \frac{1}{x}$$

so that we can choose

$$\boxed{y = h(t,x) = \ln|x|}. \tag{6.25}$$

Also, by (6.14)

$$\tilde{g}(t) = 0 + \frac{1}{x}\big[g_2(t)x\big] - \frac{1}{2x^2}\big[f_2(t)x\big]^2$$

that is,

$$\boxed{\tilde{g}(t) = g_2(t) - \frac{1}{2} f_2(t)^2}.$$

The transformed equation (6.10) is thus

$$\begin{cases} dY_t = f_2(t)\, dB_t + \Big(g_2(t) - \frac{1}{2} f_2(t)^2\Big)\, dt \\ Y(0) = h\big(0, X(0)\big) = \ln(X_0) \end{cases} \tag{6.26}$$

(assuming that $X_0(\omega) > 0$ for all ω), and its solution is

$$Y(t) = Y(0) + \int_0^t f_2(s)\,dB_s + \int_0^t \Big(g_2(s) - \frac{1}{2}f_2(s)^2\Big)\,ds.$$

Now we have to revert to the original variables. Solving (6.25) for x we obtain $x = e^y$ (assuming that $x > 0$.) Hence, $X(t) = e^{Y(t)}$, that is

$$X(t) = X_0 \exp\Big[\int_0^t f_2(s)\,dB_s\Big] \exp\Big[\int_0^t \Big(g_2(s) - \frac{1}{2}f_2(s)^2\Big)\,ds\Big]. \tag{6.27}$$

Observe that if $X_0(\omega) < 0$ for all ω then we choose $y = \ln(-x)$ and in the resubstitution, $x = -e^y$ to arrive at the same formula (6.27).

Case 2: The equation (6.23) is *narrow-sense*,

$$dX_t = f_1(t)\,dB_t + \big[g_1(t) + g_2(t)X_t\big]\,dt \tag{6.28}$$

with $f_1 \neq 0 \in C^1[0,\infty)$ and $g_1(t), g_2(t)L^1[0,t]$ for all $t \geq 0$. The fraction in (6.24) becomes

$$M(t,x) = \frac{f_1'(t) - f_1(t)g_2(t)}{f_1(t)} = \frac{f_1'(t)}{f_1(t)} - g_2(t)$$

and is independent of x. That is, a narrow-sense linear equation is reducible. Equation (6.19) now becomes

$$\frac{\tilde{f}'(t)}{\tilde{f}(t)} = M(t) = \frac{f_1'(t)}{f_1(t)} - g_2(t).$$

Solving for $\tilde{f}$, we obtain one choice

$$\boxed{\tilde{f}(t) = f_1(t)p(t)}, \qquad \text{where} \qquad p(t) = e^{-\int_0^t g_2(s)\,ds}.$$

Equation (6.16) gives us

$$h_x = \frac{\tilde{f}(t)}{f(t,x)} = \frac{f_1(t)p(t)}{f_1(t)} = p(t).$$

Hence we can choose

$$\boxed{y = h(t,x) = p(t)x}. \tag{6.29}$$

By (6.14)

$$\tilde{g}(t) = -g_2(t)p(t)x + p(t)\big[g_1(t) + g_2(t)x\big] - 0$$

that is,

$$\boxed{\tilde{g}(t) = p(t)g_1(t)}.$$

The transformed equation (6.10) is thus

$$\begin{cases} dY_t = f_1(t)p(t)\,dB_t + g_1(t)p(t)\,dt \\ Y(0) = h\big(0, X(0)\big) = p(0)X_0 = X_0 \end{cases} \tag{6.30}$$

and its solution is

$$Y(t) = Y(0) + \int_0^t f_1(s)p(s)\,dB_s + \int_0^t g_1(s)p(s)\,ds.$$

Now we have to revert to the original variables. Solving (6.29) for x we obtain $x = \frac{1}{p(t)}y$. Hence, $X(t) = \frac{1}{p(t)}Y(t)$, that is

$$X(t) = \frac{1}{p(t)}\left[X_0 + \int_0^t f_1(s)p(s)\,dB_s + \int_0^t g_1(s)p(s)ds\right] \tag{6.31}$$

where

$$p(t) = e^{-\int_0^t g_2(s)\,ds}.$$

(Compare these solutions with the solutions of deterministic first order linear differential equations !)

Example 6.2. The equation

$$dX_t = \sigma\,dB_t - bX_t\,dt \qquad (\sigma, b \text{ constant}) \tag{6.32}$$

is called the *Langevin equation.* Here the deterministic part looks like the equation of exponential growth/decay, This equation is narrow-sense linear, with $f_1(t) = \sigma$, $g_1(t) = 0$ and $g_2(t) = -b$. It is also autonomous. The equation models, for example, the velocity of a grain of pollen in a viscous liquid. The deterministic part reflects reduction in velocity due to water resistance, and the stochastic part introduces "random acceleration".

Its solution is given by (6.31). Since

$$p(t) = e^{-\int_0^t g_2(s)\,ds} = e^{\int_0^t b\,ds} = e^{bt}$$

then

$$X(t) = e^{-bt}\left[X_0 + \sigma\int_0^t e^{bs}\,dB_s\right].$$

The solution $\{X_t\}_{t\geq 0}$ is called an *Ornstein-Uhlenbeck process.*

Assume that the initial condition is deterministic, $X(0) = x_0$. Then the solution becomes

$$X(t) = \underbrace{x_o e^{-bt}}_{=:\mu(t)} + \underbrace{\sigma e^{-bt}\int_0^t e^{bs}\,dB_s}_{=:K(t)}.$$

Now by theorem 4.15,

$$\int_0^t e^{bs}\,dB_s$$

is an $N\big(0, r(t)\big)$-random variable, where

$$r(t) = \|\, e^{bs} \,\|^2_{L^2[0,t]} = \int_0^t e^{2bs}\,ds = \frac{1}{2b}(e^{2bt} - 1),$$

so that by exercise 2.2,

$$K(t) \text{ is } N\Big(0, \frac{\sigma^2}{2be^{2bt}}(e^{2bt}-1)\Big) = N\Big(0, \frac{\sigma^2}{2b}(1-e^{-2bt})\Big).$$

Now as $\mu(t)$ is constant with respect to ω, it is an easy exercise to verify that

$$X(t) \text{ is } N\Big(\mu(t), \frac{\sigma^2}{2b}(1-e^{-2bt})\Big).$$

(In general, if $Y = c =$ constant and Z is $N(0,r)$, then $X+Y$ is $N(c,r)$.) Observe that

$$\lim_{t\to\infty} \mu(t) = 0 \qquad \text{and} \qquad \lim_{t\to\infty} \operatorname{var}(X_t) = \lim_{t\to\infty} \frac{\sigma^2}{2b}\big(1-e^{-2bt}\big) = \frac{\sigma^2}{2b}.$$

So as time passes on, the mean velocity of the pollen will decrease to zero (which is expected because of friction) while the variance will increase from zero to $\frac{\sigma^2}{2b}$.

Example 6.3. Consider the equation

$$dX_t = \alpha X_t\, dB_t + \beta X_t\, dt \qquad (\alpha, \beta \text{ constant}). \tag{6.33}$$

This is a linear, homogeneous equation (with $f_2(t) = \alpha$ and $g_2(t) = \beta$). It is used to model population growth. The stochastic part reflects random growth, while the deterministic part coincides with the usual equation of exponential growth/decay. Since the population is always positive, we have $X_0(\omega) > 0$ for all ω. The solution of this equation is given by (6.27),

$$X(t) = X_0 \exp\Big[\int_0^t \alpha\, dB_s\Big] \exp\Big[\int_0^t \underbrace{\Big(\beta - \frac{1}{2}\alpha^2\Big)}_{=:b} ds\Big].$$

Since the integrands are all constant,

$$X(t) = X_0\, e^{\alpha B_t + bt}.$$

The process $\{X_t\}_{t\geq 0}$ is called *geometric Brownian motion.*

Note that by exercise 3.2,

$$\Big\{e^{\alpha B_t - \frac{\alpha^2}{2}t}\Big\}_{t\geq 0}$$

is an $(\mathcal{F}_t)$-martingale. Now by theorem 3.5, martingales have constant expectation. Hence,

$$\begin{aligned} E\big[X(t)\big] &= E\big[X_0\, e^{\alpha B_t + bt}\big] = E\big[e^{\beta t} X_0\, e^{\alpha B_t - \frac{\alpha^2}{2}t}\big] = e^{\beta t} E\big[X_0\, e^{\alpha B_t - \frac{\alpha^2}{2}t}\big] \\ &= e^{\beta t} E\big[X_0\big]\, E\big[e^{\alpha B_t - \frac{\alpha^2}{2}t}\big] = e^{\beta t} E\big[X_0\big]\, E\big[e^{\alpha B_0 - \frac{\alpha^2}{2}\cdot 0}\big] \\ &= e^{\beta t} E\big[X_0\big]\, E\big[e^0\big] = e^{\beta t} E\big[X_0\big]. \end{aligned}$$

where we have used the fact that B_t is independent of $\mathcal{F}_0$, hence of X_0, and $\{B_t\}_{t\geq 0}$ is *standard* Brownian motion. Thus, the expected value of $X(t)$ increases exponentially without limit.

Exercise 6.1. Consider an autonomous equation

$$dX_t = cX_t\, dB_t + g(X_t)\, dt \qquad (c \text{ constant}).$$

with $X(0) = X_0 > 0$. Show:

1. This equation is reducible $\Leftrightarrow$ g is for the form

$$g(x) = Lx - Kx \ln |x|$$

   for some constants K, L.

2. If the equation is reducible and $K \neq 0$, then its solution is

$$X(t) = [X_o]^{e^{-Kt}} \exp\left[\frac{cN}{K}(1 - e^{-Kt})\right] \exp\left[ce^{-Kt} \int_0^t e^{Ks}\, dB_s\right]$$

   where $N = \frac{L}{c} - \frac{c}{2}$.

Bibliography

[1] Robert B. Ash and Catherine A. Doléans-Dade, *Probability & Measure Theory*, 2nd ed., Academic Press, 2000.

[2] Donald L. Cohn, *Measure Theory*, Birkhäuser, 1980.

[3] Gerald B. Folland, *Real Analaysis – Modern Techniques and Applications*, 2nd ed., John Wiley & Sons, 1999.

[4] Thoms C. Gard, *Introduction to Stochastic Differential Equations*, Marcel Dekker, 1988.

[5] Zenghu Li, *Stochastic Differential Equations*, `http://math.bnu.edu.cn/~lizh`.

[6] Thomas Mikosch, Elementary Stochastic Calculus with Finance in View, World Scientific, 1998.