Math 454 - Additional Material

Student edition with proofs

Michael Fochler
Department of Mathematics
Binghamton University

Last update: January 17, 2025

Version: **2025-01-17** ©Michael Fochler 2020-2025

Contents

History of Updates:

Date	Topic
2020-12-23	Created.

1 Before You Start

"All models are wrong, but some are useful".

Attributed to the statistician George E. P. Box (1919–2013)

This quote certainly applies to stochastic models in mathematical finance. The price of financial instruments such as stocks, bonds and stock options is usually assumed to be a Markov process, i.e., the future development of those prices does not depend on their past development, but only on their current value. As debatable as it is to completely ignore the history of a stock when predicting its future, those stochastic models are in wide use by institutions and individuals that trade financial securities.

Consider how far we have come in the last 120 years. In 1900 the French mathematician Louis Bachelier published his thesis, [2], Théorie de la spéculation, in which he modeled stock price as a Brownian motion. As a consequence, stock prices would be negative with positive probability. Today even the most basic models involving the pricing of stock options such as puts and calls are much improved in that they prevent stock prices from ever becoming negative.

This course attempts to convey the basics of continuous time stochastic models in mathematical finance. Unfortunately this is not possible in any reasonable manner without the concept of continous time martingales, and those again need a very sophisticated understanding of conditional probabilities and conditional expectations. Accordingly, a substantial part of these lecture notes is dedicated to conveying the necessary material. Much of which usually is taught in a probability theory for beginning graduate students. Thus proofs, even where they are given, are often considered optional.

1.1 About This Document

Remark 1.1 (The purpose of this document)**.** The intent is to put some core definitions and theorems into these lecture notes, in particular, if there is a substantial difference in notation and/or presentation to that used in the text for this class, [14] Shreve, Steven: Stochastic Calculus for Finance II: Continuous-Time Models. □

Remark 1.2 (Acknowledgements)**.** I am indepted to Prof. Dikran Karagueuzian from the Department of Mathematical Sciences at Binghamton University for sharing his notes from teaching this class at an earlier time. $\square$

2 Preliminaries about Sets, Numbers and Functions

Introduction 2.1. You find here a range of mathematical definitions and facts that you should be familiar with. □

The student should read this chapter carefully, with the expectation that it contains material that they are not familiar with, as much of it will be used in lecture without comment. Very likely candidates are power sets, a function $f : X \to Y$ where domain X and codomain Y are part of the definition.

2.1 Sets and Basic Set Operations

Introduction 2.2. This first subchapter of ch.2 is different from the following ones in that the treatment of sets given here is sufficiently exact for a PhD in math unless s/he works in the areas of logic or axiomatic set theory. The only exception is the end of the chapter where the preliminary definition of the size of a set (def.2.10 on p.16) needs to refer to finiteness.

Ask a mathematician how her or his Math is different from the kind of Math you learn in high school, in fact, from any kind of Math you find outside textbooks for mathematicians and theoretical physicists. One of the answers you are likely to get is that Math is not so much about numbers but also about other objects, among them sets and functions. Once you know about those, you can tackle sets of functions, set functions, sets of set functions, . . . □

An entire book can be filled with a mathematically precise theory of sets. [1] For our purposes the following "naive" definition suffices:

Definition 2.1 (Sets)**.**

[1]See remark 2.2 ("Russell's Antinomy") below.

A **set** is a collection of stuff called **members** or **elements** which satisfies the following rules: The order in which you write the elements does not matter and if you list an element two or more times then **it only counts once**.
We write a set by enclosing within curly braces the elements of the set. This can be done by listing all those elements or giving instructions that describe those elements. For example, to denote by X the set of all integer numbers between 18 and 24 we can write either of the following:

$$X := \{18, 19, 20, 21, 22, 23, 24\} \quad \text{or} \quad X := \{n : n \text{ is an integer and } 18 \le n \le 24\}$$

Both formulas clearly define the same collection of all integers between 18 and 24. On the left the elements of X are given by a complete list, on the right **setbuilder notation**, i.e., instructions that specify what belongs to the set, is used instead.
It is customary to denote sets by capital letters and their elements by small letters but this is not a hard and fast rule. You will see many exceptions to this rule in this document.
We write $x_1 \in X$ to denote that an item x_1 is an element of the set X and $x_2 \notin X$ to denote that an item x_2 is not an element of the set X. Occasionally we follow Shreve's example and write x_1 **in** X and x_2 **not in** X. [2]
For the above example we have $20 \in X$, $27 - 6 \in X$, $38 \notin X$, 'Jimmy' $\notin X$. $\square$

Example 2.1 (No duplicates in sets)**.** The following collection of alphabetic letters is a set:

$$S_1 = \{a, e, i, o, u\}$$

and so is this one:

$$S_2 = \{a, e, e, i, i, i, o, o, o, o, u, u, u, u, u\}$$

Did you notice that those two sets are equal? $\square$

Remark 2.1. The symbol n in the definition of $X = \{n : n \text{ is an integer and } 18 \le n \le 24\}$ is a **dummy variable** in the sense that it does not matter what symbol you use. The following sets all are equal to X:

$$\{x : x \text{ is an integer and } 18 \le x \le 24\},$$
$$\{\alpha : \alpha \text{ is an integer and } 18 \le \alpha \le 24\},$$
$$\{\mathfrak{Z} : \mathfrak{Z} \text{ is an integer and } 18 \le \mathfrak{Z} \le 24\} \quad \square$$

Remark 2.2 (Russell's Antinomy)**.** Care must be taken so that, if you define a set with the use of setbuilder notation, no inconsistencies occur. Here is an example of a definition of a set that leads to contradictions.

$$A := \{B : B \text{ is a set and } B \notin B\} \tag{2.1}$$

What is wrong with this definition? To answer this question let us find out whether or not this set A is a member of A. Assume that A belongs to A. The condition to the right of the colon states that $A \notin A$ is required for membership in A, so our assumption $A \in A$ must be wrong. In other words, we have established "by contradiction" that $A \notin A$ is true. But this is not the end of it: Now that we know that $A \notin A$ it follows that $A \in A$ because A contains **all** sets that do not contain themselves.

In other words, we have proved the impossible: both $A \in A$ and $A \notin A$ are true! There is no way out of this logical impossibility other than excluding definitions for sets such as the one given above. It is very important for mathematicians that their theories do not lead to such inconsistencies. Therefore, examples as the one above have spawned very complicated theories about "good sets". It is possible for a mathematician to specialize in the field of axiomatic set theory (actually, there are several set theories) which endeavors to show that the sets are of any relevance in mathematical theories do not lead to any logical contradictions.

The great majority of mathematicians take the "naive" approach to sets which is not to worry about accidentally defining sets that lead to contradictions and we will take that point of view in this document. □

Definition 2.2 (empty set)**.**

$\emptyset$ or {} denotes the **empty set**. It is the one set that does not contain any elements. □

Remark 2.3 (Elements of the empty set and their properties)**.** You can state anything you like about the elements of the empty sets as there are none. The following statements all are true:

a: If $x \in \emptyset$ then x is a positive number.
b: If $x \in \emptyset$ then x is a negative number.
c: Define $a \sim b$ if and only if both are integers and $a - b$ is an even number.
For any $x, y, z \in \emptyset$ it is true that
c1: $x \sim x$,
c2: if $x \sim y$ then $y \sim x$,
c3: if $x \sim y$ and $y \sim z$ then $x \sim z$.
d: Let A be any set. If $x \in \emptyset$ then $x \in A$.

As you will learn later, **c1**+**c2**+**c3** means that "$\sim$" is an equivalence relation (see def.**??** on p.**??**) and **d:** means that the empty set is a subset (see the next definition) of any other set. □

Definition 2.3 (subsets and supersets)**.**

We say that a set A is a **subset** of the set B and we write $A \subseteq B$ if any element of A also belongs to B. Equivalently we say that B is a **superset** of the set A and we write $B \supseteq A$. We also say that B includes A or A is included by B. Note that $A \subseteq A$ and $\emptyset \subseteq A$ is true for any set A.

If $A \subseteq B$ but $A \neq B$, i.e., there is at least one $x \in B$ such that $x \notin A$, then we say that A is a **strict subset** or a **proper subset** of B. We write "$A \subsetneq B$" or "$A \subset B$". Alternatively we say that B is a **strict superset** or a **proper superset** of A and we write "$B \supsetneq A$") or "$B \supset A$". □

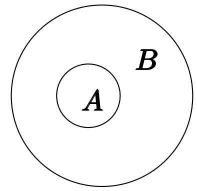


Figure 2.1: Set inclusion: $A \subseteq B,\ B \supseteq A$

Two sets A and B are equal means that they both contain the same elements. In other words, $A = B$ iff $A \subseteq B$ and $B \subseteq A$.

"**iff**" is a short for "if and only if": P iff Q for two statements P and Q means that if P is valid then Q is valid and vice versa. [3]

To show that two sets A and B are equal you show that
a. if $x \in A$ then $x \in B$,
b. if $x \in B$ then $x \in A$.

Definition 2.4 (unions, intersections and disjoint unions)**.**

Given are two arbitrary sets A and B. No assumption is made that either one is contained in the other or that either one contains any elements!
The **union** $A \cup B$ (pronounced "A union B") is defined as the set of all elements which belong to A or B or both.
The **intersection** $A \cap B$ (pronounced "A intersection B") is defined as the set of all elements which belong to both A and B.
We call A and B **disjoint** , also **mutually disjoint** , if $A \cap B = \emptyset$. We then usually write $A \uplus B$ (pronounced "A disjoint union B") rather than $A \cup B$. ☐

We could have shortened in the last definition the phrase "all elements which belong to A or B or both" to "all elements which belong to A or B". We will almost always do so because it is understood among mathematicians that "or" always means at least one of the choices. If they mean instead exactly one of the choices $\#1, \#2, \dots \#n$ then they will use the phrase "either $\#1$ or $\#2$ or $\dots$ or $\#n$. See rem**??** on p.**??**. We will also see in a moment that there is a special symbol $A \triangle B$ which denotes the items which belong to either A or B (but not both).

Remark 2.4. It is obvious from the definition of unions and intersections and the meaning of the phrases " all elements which belong to A or B or both", "all elements which belong to both A and B" and "$A \subseteq B$ if any element of A also belongs to B" that the following is true for any sets A, B

[3]A formal definition of "if and only if" will be given in def.**??** on p.**??** where we will also introduce the symbolic notation $P \Leftrightarrow Q$. Informally speaking, a statement is something that is either true or false.

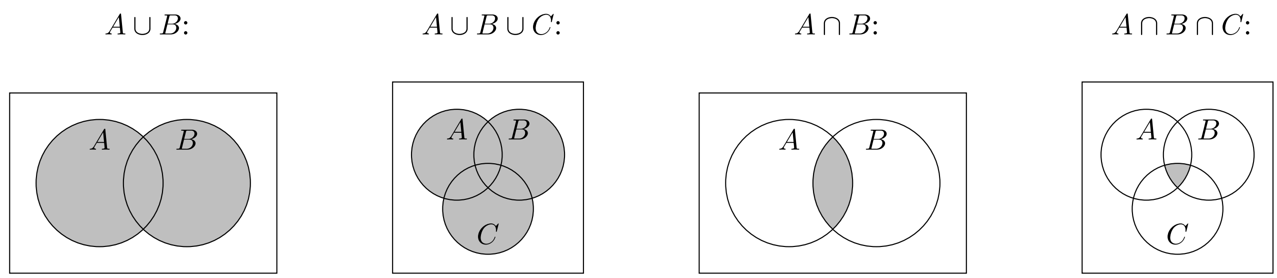


Figure 2.2: Union and intersection of sets

and C.

(2.2) $$A \cap B \subseteq A \subseteq A \cup B,$$

(2.3) $$A \subseteq B \Rightarrow A \cap B = A \text{ and } A \cup B = B,$$

(2.4) $$A \subseteq B \Rightarrow A \cap C \subseteq B \cap C \text{ and } A \cup C \subseteq B \cup C.$$

The symbol $\Rightarrow$ stands for "allows us to conclude that". So $A \subseteq B \Rightarrow A \cap B = A$ means "From the truth of $A \subseteq B$ we can conclude that $A \cap B = A$ is true". Shorter: "From $A \subseteq B$ we can conclude that $A \cap B = A$". Shorter: "If $A \subseteq B$ then it follows that $A \cap B = A$". Shorter: "If $A \subseteq B$ then $A \cap B = A$". More technical: $A \subseteq B$ implies $A \cap B = A$.

You will learn more about implication in ch.**??** of this document and in ch.3 (Some Points of Logic) of [5] Beck/Geoghegan: The Art of Proof. ☐

Definition 2.5 (set differences and symmetric differences)**.**

Given are two arbitrary sets A and B. No assumption is made that either one is contained in the other or that either one contains any elements!
The **difference set** or **set difference** $A \setminus B$ (pronounced "A minus B") is defined as the set of all elements which belong to A but not to B:

(2.5) $$A \setminus B := \{x \in A : x \notin B\}$$

The **symmetric difference** $A \triangle B$ (pronounced "A delta B") is defined as the set of all elements which belong to either A or B but not to both A and B:

(2.6) $$A \triangle B := (A \cup B) \setminus (A \cap B) \quad \square$$

Definition 2.6 (Universal set)**.**

Usually there always is a big set Ω that contains everything we are interested in and we then deal with all kinds of subsets $A \subseteq \Omega$. Such a set is called a **"universal" set**. ☐

For example, in this document, we often deal with real numbers and our universal set will then be $\mathbb{R}$. [4] If there is a universal set, it makes perfect sense to talk about the complement of a set:

Definition 2.7 (Complement of a set)**.**

Let Ω be a universal set. The **complement** of a set $A \subseteq \Omega$ consists of all elements of Ω which do not belong to A. We write $A^{\complement}$. or $\complement A$ In other words:

$$A^{\complement} := \complement A := \Omega \setminus A = \{\omega \in \Omega : x \notin A\} \quad \square \tag{2.7}$$

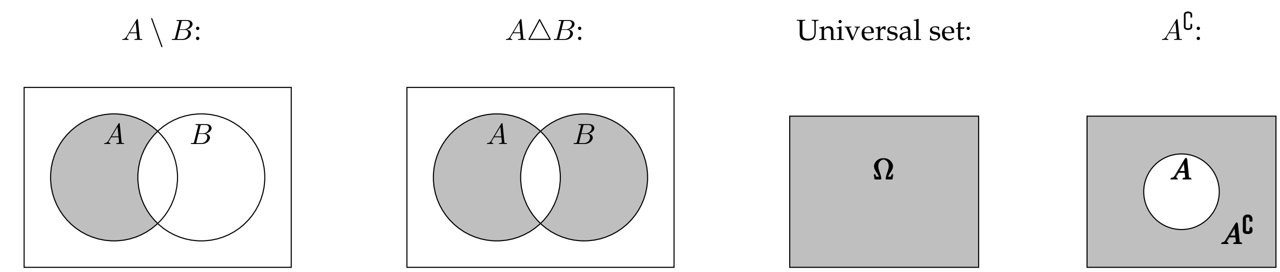


Figure 2.3: Difference, symmetric difference, universal set, complement

Remark 2.5. Note that for any kind of universal set Ω it is true that

$$\Omega^{\complement} = \emptyset, \qquad \emptyset^{\complement} = \Omega. \quad \square \tag{2.8}$$

Example 2.2 (Complement of a set relative to the unit interval)**.** Assume we are exclusively dealing with the unit interval, i.e., $\Omega = [0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$. Let $a \in [0,1]$ and $\delta > 0$ and

$$A = \{x \in [0,1] : a - \delta < x < a + \delta\} \tag{2.9}$$

the δ–neighborhood [5] of a (with respect to $[0,1]$ because numbers outside the unit interval are not considered part of our universe). Then the complement of A is

$$A^{\complement} = \{x \in [0,1] : x \leq a - \delta \text{ or } x \geq a + \delta\}. \quad \square$$

Draw some Venn diagrams to visualize the following formulas.

Proposition 2.1. *Let A, B, X be subsets of a universal set Ω and assume $A \subseteq X$. Then,*

[4] $\mathbb{R}$ is the set of all real numbers, i.e., the kind of numbers that make up the x-axis and y-axis in a beginner's calculus course (see ch.2.3 ("Classification of numbers") on p.17).

[5] Neighborhoods of a point will be discussed in the chapter on the topology of $\mathbb{R}^n$ (see (**??**) on p.**??**). In short, the δ–neighborhood of a is the set of all points with distance less than δ from a.

(2.10a)	$A \cup \emptyset = A; \qquad A \cap \emptyset = \emptyset$
(2.10b)	$A \cup \Omega = \Omega; \qquad A \cap \Omega = A$
(2.10c)	$A \cup A^{\complement} = \Omega; \qquad A \cap A^{\complement} = \emptyset$
(2.10d)	$A \triangle B = (A \setminus B) \uplus (B \setminus A)$
(2.10e)	$A \setminus A = \emptyset$
(2.10f)	$A \triangle \emptyset = A; \qquad A \triangle A = \emptyset$
(2.10g)	$X \triangle A = X \setminus A$
(2.10h)	$A \cup B = (A \triangle B) \uplus (A \cap B)$
(2.10i)	$A \cap B = (A \cup B) \setminus (A \triangle B)$
(2.10j)	$A \triangle B = \emptyset$ *if and only if* $B = A$

PROOF: The proof is left as exercise 2.2. See p.35. ■

Next we give a very detailed and rigorous proof of a simple formula for sets. The reader should make an effort to understand it line by line.

Proposition 2.2 (Distributivity of unions and intersections for two sets)**.**

Let A, B, C be sets. Then

(2.11) $$(A \cup B) \cap C = (A \cap C) \cup (B \cap C),$$

(2.12) $$(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

PROOF: ⋆ We only prove (2.11). The proof of (2.12) is left as exercise 2.1.

PROOF of "⊆": Let $x \in (A \cup B) \cap C$. It follows from (2.2) on p.11 that $x \in (A \cup B)$, i.e., $x \in A$ or $x \in B$ (or both). It also follows from (2.2) that $x \in C$. We must show that $x \in (A \cap C) \cup (B \cap C)$ regardless of whether $x \in A$ or $x \in B$.

Case 1: $x \in A$. Since also $x \in C$, we obtain $x \in A \cap C$, hence, again by (2.2), $x \in (A \cap C) \cup (B \cap C)$, which is what we wanted to prove.

Case 2: $x \in B$. We switch the roles of A and B. This allows us to apply the result of case 1, and we again obtain $x \in (A \cap C) \cup (B \cap C)$.

PROOF of "⊇": Let $x \in (A \cap C) \cup (B \cap C)$, i.e., $x \in A \cap C$ or $x \in B \cap C$ (or both). We must show that $x \in (A \cup B) \cap C$ regardless of whether $x \in A \cap C$ or $x \in B \cap C$.

Case 1: $x \in A \cap C$. It follows from $A \subseteq A \cup B$ and (2.4) on p.11 that $x \in (A \cup B) \cap C$, and we are done in this case.

Case 2: $x \in B \cap C$. This time it follows from $A \subseteq A \cup B$ that $x \in (A \cup B) \cap C$. This finishes the proof of (2.11).

Epilogue: The proofs both of "⊆" and of "⊇" were **proofs by cases**, i.e., we divided the proof into several cases (to be exact, two for each of "⊆" and "⊇"), and we proved each case separately. For example we proved that $x \in (A \cup B) \cap C$ implies $x \in (A \cap C) \cup (B \cap C)$ separately for the cases $x \in A$ and $x \in B$. Since those two cases cover all possibilities for x the assertion "if $x \in (A \cup B) \cap C$

then $x \in (A \cap C) \cup (B \cap C)$" is proven. ■

Proposition 2.3 (De Morgan's Law for two sets)**.**

Let $A, B \subseteq \Omega$. Then the complement of the union is the intersection of the complements, and the complement of the intersection is the union of the complements:

(2.13) ***a.*** $(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}$ ***b.*** $(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$

PROOF of **a**:

1) First we prove that $(A \cup B)^{\complement} \subseteq A^{\complement} \cap B^{\complement}$:

Assume that $x \in (A \cup B)^{\complement}$. Then $x \notin A \cup B$, which is the same as saying that x does not belong to either of A and B. That in turn means that x belongs to both $A^{\complement}$ and $B^{\complement}$ and hence also to the intersection $A^{\complement} \cap B^{\complement}$.

2) Now we prove that $(A \cup B)^{\complement} \supseteq A^{\complement} \cap B^{\complement}$:

Let $x \in A^{\complement} \cap B^{\complement}$. Then x belongs to both $A^{\complement}, B^{\complement}$, hence neither to A nor to B, hence $x \notin A \cup B$. Therefore x belong to the complement of $A \cup B$. This completes the proof of formula **a**.

PROOF of **b**:

The proof is very similar to that of formula **a** and left as an exercise. ■

Formulas **a** through **g** of the next proposition are very useful. You are advised to learn them by heart and draw pictures to visualize them. You also should examine closely the proof of the next proposition. It shows how a proof which involves 3 or 4 sets can be split into easily dealt with cases.

Proposition 2.4.

Let A, B, C, Ω be sets such that $A, B, C \subseteq \Omega$. Then

a. $(A \triangle B) \triangle C = A \triangle (B \triangle C)$
b. $A \triangle \emptyset = \emptyset \triangle A = A$
c. $A \triangle A = \emptyset$
d. $A \triangle B = B \triangle A$

Further we have the following for the intersection operation:

e. $(A \cap B) \cap C = A \cap (B \cap C)$
f. $A \cap \Omega = \Omega \cap A = A$
g. $A \cap B = B \cap A$

And we have the following interrelationship between $\triangle$ and $\cap$:

h. $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$

PROOF: ⋆

Only the proof of **a** is given here. It is very tedious and there is a much more elegant proof, but that one requires knowledge of indicator functions [6] and of base 2 modular arithmetic (see, e.g., [5] B/G (Beck/Geoghegan) ch.6.2).

[6] Indicator functions will be discussed in ch.3.3 on p.44 and in ch.?? on p.??.

By definition $x \in U \triangle V$ if and only if either $x \in U$ or $x \in V$, i.e.,
(either) $\big[\ x \in U$ and $x \notin V\ \big]$ or $\big[\ x \in V$ and $x \notin U\ \big]$
Hence $x \in (A \triangle B) \triangle C$ means either $x \in (A \triangle B)$ or $x \in C$, i.e.,
either $\big[\ x \in A, x \notin B$ or $x \in B, x \notin A\ \big]$ or $x \in C$, i.e., we have one of the following four combinations:

a.	$x \in A$	$x \notin B$	$x \notin C$
b.	$x \notin A$	$x \in B$	$x \notin C$
c.	$x \in A$	$x \in B$	$x \in C$
d.	$x \notin A$	$x \notin B$	$x \in C$

and $x \in A \triangle (B \triangle C)$ means either $x \in A$ or $x \in (B \triangle C)$, i.e.,
either $x \in A$ or $\big[\ x \in B, x \notin C$ or $x \in C, x \notin B\ \big]$, i.e., we have one of the following four combinations:

1.	$x \in A$	$x \in B$	$x \in C$
2.	$x \in A$	$x \notin B$	$x \notin C$
3.	$x \notin A$	$x \in B$	$x \notin C$
4.	$x \notin A$	$x \notin B$	$x \in C$

We have a perfect match **a** $\leftrightarrow$ **2**, **b** $\leftrightarrow$ **3**, **c** $\leftrightarrow$ **1**, **d** $\leftrightarrow$ **4**. and this completes the proof of **a.**
■

Definition 2.8 (Partition)**.**

Let Ω be a set and $\mathfrak{A} \subseteq 2^\Omega$. We call $\mathfrak{A}$ a **partition** or a **partitioning** of Ω if

a. $A \cap B = \emptyset$ for any two $A, B \in \mathfrak{A}$ such that $A \neq B$, i.e., $\mathfrak{A}$ consists of mutually disjoint subsets of Ω (see def.2.4),

b. $\Omega = \biguplus \Big[A : A \in \mathfrak{A} \Big]$. □

Example 2.3.

a. For $n \in \mathbb{Z}$ let $A_n := \{n\}$. Then $\mathfrak{A} := \{A_n : n \in \mathbb{Z}\}$ is a partition of $\mathbb{Z}$. $\mathfrak{A}$ is not a partition of $\mathbb{N}$ because not all its members are subsets of $\mathbb{N}$ and it is not a partition of $\mathbb{Q}$ or $\mathbb{R}$. The reason: $\frac{1}{2} \in \mathbb{Q}$ and hence $\frac{1}{2} \in \mathbb{R}$, but $\frac{1}{2} \notin A_n$ for any $n \in \mathbb{Z}$, hence condition **b** of def.2.8 is not satisfied.

b. For $n \in \mathbb{N}$ let $B_n := [\, n^2, (n+1)^2 [\, = \{x \in \mathbb{R} : n^2 \leq x < (n+1)^2\}$. Then $\mathfrak{B} := \{B_n : n \in \mathbb{N}\}$ is a partition of $[1, \infty[$. □

Definition 2.9 (Power set)**.**

The **power set**

$$2^\Omega := \{A : A \subseteq \Omega\}$$

of a set Ω is the set of all its subsets. Note that many older texts also use the notatio
for the power set. □

Remark 2.6. Note that $\emptyset \in 2^\Omega$ for any set Ω, even if $\Omega = \emptyset$: $2^\emptyset = \{\emptyset\}$. It follows that the power set of the empty set is not empty. □

Definition 2.10 (Size of a set)**.**

a. Let X be a finite set, i.e., a set which only contains finitely many elements. We write $\left|X\right|$ for the number of its elements, and we call $\left|X\right|$ the **size** of the set X.
b. For infinite, i.e., not finite sets Y, we define $|Y| := \infty$. $\square$

A lot more will be said about sets once families are defined.

2.2 The Proper Use of Language in Mathematics: Any vs All, etc

Mathematics must be very precise in its formulations. Such precision is achieved not only by means of symbols and formulas, but also by its use of the English language. We will list some important points to consider early on in this document.

2.2.0.1 All vs. ANY

Assume for the following that X is a set of numbers. Do the following two statements mean the same?

(1) It is true for ALL $x \in X$ that x is an integer.
(2) It is true for ANY $x \in X$ that x is an integer.

You will hopefully agree that there is no difference and that one could rewrite them as follows:

(3) ALL $x \in X$ are integers.
(4) ANY $x \in X$ is an integer.
(5) EVERY $x \in X$ is an integer.
(6) EACH $x \in X$ is an integer.
(7) IF $x \in X$ THEN x is an integer.

Is it then always true that ALL and ANY means the same? Consider

(8a) It is NOT true for ALL $x \in X$ that x is an integer.
(8b) It is NOT true for ANY $x \in X$ that x is an integer.

Completely different things have been said: Statement **(8)** asserts that as few as one item and as many as all items in X are not integers, whereas **(9)** states that no items, i.e., exactly zero items in X, are integers.

My suggestion: Express formulations like **(8b)** differently. You could have written instead

(8c) There is no $x \in X$ such that x is an integer.

2.2.0.2 AND vs. IF ... THEN

Some people abuse the connective AND to also mean IF ... THEN. However, mathematicians use the phrase "p AND q" exclusively to mean that something applies to both p and q. Contrast the use of AND in the following statements:

(9) "Jane is a student AND Joe likes baseball". This phrase means that both are true: Jane is indeed a student and Joe indeed likes baseball.
(10) "You hit me again AND you'll be sorry". **Never, ever use the word AND in this context!** A mathematician would express the above as "IF you hit me again THEN you'll be sorry".

2.2.0.3 OR vs. EITHER ... OR

The last topic we address is the proper use of "OR". In mathematics the phrase

(11) "p is true OR q is true"

is always to be understood as

(12) "p is true OR q is true OR BOTH are true", i.e., at least one of p, q is true.

This is in contrast to everyday language where "p is true OR q is true" often means that exactly one of p and q is true, but not not both.

When referring to a collection of items then the use of "OR" also is inclusive If the items $a, b, c, \dots$ belong to a collection $\mathscr{C}$, e.g., if those items are elements of a set, then

(13) "a OR b OR c OR ..." means that we refer to at least one of $a, b, c, \dots$.

> Note that "OR" in mathematics always is an **inclusive or**, i.e., "A OR B" means "A OR B OR BOTH". More generally, "A OR B OR ..." means "at least one of A, B, ...".
> To rule out that more than one of the choices is true you must use a phrase like "EXACTLY ONE OF A, B, C, ..." or "EITHER A OR B OR C OR ...". We refer to this as an **exclusive or**.

2.3 Numbers

We start with an informal classification of numbers. It is not meant to be mathematically exact. We will give exact definitions of the integers, rational numbers and real numbers in chapter **??** (The Real Numbers).

Definition 2.11 (Integers and decimal numerals)**.** A **digit** or **decimal digit** Is one of the numbers $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$.

We call numbers that can be expressed as a finite string of digits, possibly preceded by a minus sign, **integers**. In particular we demand that an integer can be written without a decimal point. Examples of integers are

$$3,\ -29,\ 0,\ 3\cdot 10^6,\ -1,\ 2.\bar{9},\ 12345678901234567890,\ -2018. \tag{2.14}$$

Note that $3 \cdot 10^6 = 3000000$ is a finite string of digits and that $2.\bar{9}$ equals 3 (see below about the period of a decimal numeral). We write $\mathbb{Z}$ for the set of all integers.

Numbers in the set $\mathbb{N} = \{1, 2, 3, \dots\}$ of all strictly positive integers are called **natural numbers**.

An integer n is an **even** integer if it is a multiple of 2, i.e., there exists $j \in \mathbb{Z}$ such that $n = 2j$, and it is an **odd** integer otherwise. One can give a strict proof that n is odd if and only if there exists $j \in \mathbb{Z}$ such that $n = 2j + 1$.

A **decimal** or **decimal numeral** is a finite or infinite list of digits, possibly preceded by a minus sign, which is separated into two parts by a point, the **decimal point**. The list to the left of the decimal point must be finite or empty, but there may be an infinite number of digits to its right. Examples are

$$3.0,\ -29.0,\ 0.0,\ -0.75,\ .\bar{3},\ 2.74\bar{9},\ \pi = 3.141592.....,\ -34.56. \tag{2.15}$$

The bar on top of the rightmost part of a decimal such as ".$\bar{3}$" means that this part should be repeated over and over again, i.e., $.\bar{3} = 0.33333333333\dots$ and $1.234\overline{567} = 1.234567567567\dots$.

Any integer can be transformed into a decimal numeral of same value by appending the pattern ".0" to its right. For example, the integer 27 can be written as the decimal 27.0. □

Definition 2.12 (Real numbers)**.** We call any kind of number which can be represented as a decimal numeral, a **real number**. We write $\mathbb{R}$ for the set of all real numbers. It follows from what was remarked at the end of def.2.11 that integers, in particular natural numbers, are real numbers. Thus we have the following set relations:

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{R}. \quad \square \tag{2.16}$$

We next define rational numbers.

Definition 2.13 (Rational numbers)**.** A number that is an integer or can be written as a fraction of integers, i.e., as $\frac{m}{n}$ where $m, n \in \mathbb{Z}$ and $n \neq 0$, is called a **rational number**. We write $\mathbb{Q}$ for the set of all rational numbers. □

Examples of rational numbers are

$$\tfrac{3}{4},\ -0.75,\ -\tfrac{1}{3},\ .\bar{3},\ \tfrac{7}{1},\ 16,\ \tfrac{13}{4},\ -5,\ 2.99\bar{9},\ -37\tfrac{2}{7}.$$

Note that a mathematician does not care whether a rational number is written as a fraction

$$\frac{numerator}{denominator}$$

or as a decimal numeral. The following all are representations of one third:

$$0.\bar{3} \ = \ .\bar{3} \ = \ 0.33333333333\dots \ = \ \tfrac{1}{3} \ = \ \tfrac{-1}{-3} \ = \ \tfrac{2}{6}, \tag{2.17}$$

and here are several equivalent ways of expressing the number minus four:

$$-4 \ = \ -4.000 \ = \ -3.\bar{9} \ = \ -\tfrac{12}{3} \ = \ \tfrac{4}{-1} \ = \ \tfrac{-4}{1} \ = \ \tfrac{12}{-3} \ = \ -\tfrac{400}{100}. \tag{2.18}$$

There are real numbers which cannot be expressed as integers or fractions of integers.

Definition 2.14 (Irrational numbers)**.** We call real numbers that are not rational **irrational numbers**. They hence fill the gaps that exist between the rational numbers. In fact, there is a simple way (but not easy to prove) of characterizing irrational numbers: Rational numbers are those that can be expressed with at most finitely many digits to the right of the decimal point, including repeating decimals. You can find the underlying theory and exact proofs in ch.**??** (Decimal Expansions of Real and Rational Numbers). Irrational numbers must then be those with infinitely many decimal digits without a continually repeating pattern. □

Example 2.4. To illustrate that repeating decimals are in fact rational numbers we convert $x = 0.1\overline{45}$ into a fraction:

$$99x \ = 100x - x \ = \ 14.5\overline{45} - 0.1\overline{45} = \ 14.4$$

It follows that $x = 144/990$, and that is certainly a fraction. □

Remark 2.7. Examples of irrational numbers are $\sqrt{2}$ *and* $\boldsymbol{\pi}$. A proof that $\sqrt{2}$ is irrational (actually that $\sqrt[n]{2}$ is irrational for any integer $n \geq 2$) is given in prop.**??** on p.**??**. $\square$

Definition 2.15 (Types of numbers)**.** We summarize what was said sofar about the classification of numbers:

$\mathbb{N} := \{1, 2, 3, \dots\}$ denotes the set of **natural numbers**.
$\mathbb{Z} := \{0, \pm 1, \pm 2, \pm 3, \dots\}$ denotes the set of all **integers**.
$\mathbb{Q} := \{n/d : n \in \mathbb{Z}, d \in \mathbb{N}\}$ denotes the set of all **rational numbers**.
$\mathbb{R} :=$ {all integers or decimal numbers with finitely or infinitely many decimal digits} denotes the set of all **real numbers**.
$\mathbb{R} \setminus \mathbb{Q} =$ {all real numbers which cannot be written as fractions of integers} denotes the set of all **irrational numbers**. There is no special symbol for irrational numbers. Example: $\sqrt{2}$ and π are irrational. $\square$

Here are some customary abbreviations of some often referenced sets of numbers:

$\mathbb{N}_0 := \mathbb{Z}_+ := \mathbb{Z}_{\geq 0} := \{0, 1, 2, 3, \dots\}$ denotes the set of nonnegative integers,
$\mathbb{R}_+ := \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geq 0\}$ denotes the set of all nonnegative real numbers,
$\mathbb{R}^+ := \mathbb{R}_{>0} := \{x \in \mathbb{R} : x > 0\}$ denotes the set of all positive real numbers,
$\mathbb{R}_{\neq 0} := \{x \in \mathbb{R} : x \neq 0\}$. $\square$

Definition 2.16 (Intervals of Numbers)**.** For $a, b \in \mathbb{R}$ we have the following intervals.

- $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ is the **closed interval** with endpoints a and b.
- $]a, b[:= \{x \in \mathbb{R} : a < x < b\}$ is the **open interval** with endpoints a and b.
- $[a, b[:= \{x \in \mathbb{R} : a \leq x < b\}$ and $]a, b] := \{x \in \mathbb{R} : a < x \leq b\}$ are **half-open intervals** with endpoints a and b.

The symbol "∞" stands for an object which itself is not a number but is larger than any (real) number, and the symbol "$-\infty$" stands for an object which itself is not a number but is smaller than any number. We thus have $-\infty < x < \infty$ for any number x. This allows us to define the following intervals of "infinite length":

$$\begin{aligned} &]-\infty, a] := \{x \in \mathbb{R} : x \leq a\}, \quad]-\infty, a[:= \{x \in \mathbb{R} : x < a\}, \\ &]a, \infty[:= \{x \in \mathbb{R} : x > a\}, \quad [a, \infty[:= \{x \in \mathbb{R} : x \geq a\}, \quad]-\infty, \infty[:= \mathbb{R} \end{aligned} \tag{2.19}$$

You should always work with $a < b$. In case you don't, you get

- $[a, a] = \{a\}$; $[a, a[=]a, a[=]a, a] = \emptyset$
- $[a, b] = [a, b[=]a, b[=]a, b] = \emptyset$ for $a \geq b$ $\square$

Notation 2.1 (Notation Alert for intervals of integers or rational numbers)**.**

It is at times convenient to also use the notation $[\dots]$, $]\dots[$, $[\dots[$, $]\dots]$, for intervals of integers or rational numbers. We will subscript them with $\mathbb{Z}$ or $\mathbb{Q}$. For example,

$$\begin{aligned} [\,3, n\,]_{\mathbb{Z}} &= [\,3,\ n] \cap \mathbb{Z} \;=\; \{k \in \mathbb{Z} : 3 \le k \le n\}, \\ ]-\infty, 7\,]_{\mathbb{Z}} &=]-\infty, 7\,] \cap \mathbb{Z} \;=\; \{k \in \mathbb{Z} : k \le 7\} \;=\; \mathbb{Z}_{\le 7}, \\ ]a, b[_{\mathbb{Q}} &=]a, b[\cap \mathbb{Q} \;=\; \{q \in \mathbb{Q} : a < q < b\}. \end{aligned}$$

An interval which is not subscripted always means an interval of real numbers, but we will occasionally write, e.g., $[a, b]_{\mathbb{R}}$ rather than $[a, b]$, if the focus is on integers or rational numbers and an explicit subscript helps to avoid confusion. $\square$

Definition 2.17 (Absolute value, positive and negative part)**.** For a real number x we define its

absolute value: $|x| \;=\; \begin{cases} x & \text{if } x \ge 0, \\ -x & \text{if } x < 0. \end{cases}$

positive part: $x^+ \;=\; \max(x, 0) \;=\; \begin{cases} x & \text{if } x \ge 0, \\ 0 & \text{if } x < 0. \end{cases}$

negative part: $x^- \;=\; \max(-x, 0) \;=\; \begin{cases} -x & \text{if } x \le 0, \\ 0 & \text{if } x > 0. \end{cases}$

If f is a real–valued function then we define the functions $|f|, f^+, f^-$ argument by argument:

$$|f|(x) \;:=\; |f(x)|, \qquad f^+(x) \;:=\; \big(f(x)\big)^+, \qquad f^-(x) \;:=\; \big(f(x)\big)^-. \;\square$$

For completeness we also give the definitions of min and max.

Definition 2.18 (Minimum and maximum)**.** For two real number x, y we define

maximum: $x \vee y \;=\; \max(x, y) \;=\; \begin{cases} x & \text{if } x \ge y, \\ y & \text{if } x \le y. \end{cases}$

minimum: $x \wedge y \;=\; \min(x, y) \;=\; \begin{cases} y & \text{if } x \ge y, \\ x & \text{if } x \le y. \end{cases}$

If f and g is are real–valued function then we define the functions $f \vee g = \max(f, g)$ and $f \wedge g = \min(f, g)$ argument by argument:

$$f \vee g(x) \;:=\; f(x) \vee g(x) \;=\; \max\big(f(x), g(x)\big), \quad f \wedge g(x) \;:=\; f(x) \wedge g(x) \;=\; \min\big(f(x), g(x)\big). \;\square$$

Remark 2.8. You are advised to compute $|x|, x^+, x^-$ for $x = -5, x = 5, x = 0$ and convince yourself that the following is true:

$$\begin{aligned} x &= x^+ - x^-, \\ |x| &= x^+ + x^-, \end{aligned}$$

Thus any real–valued function f satisfies

$$\begin{aligned} f &= f^+ - f^-, \\ |f| &= f^+ + f^-, \end{aligned}$$

Get a feeling for the above by drawing the graphs of $|f|, f^+, f^-$ for the functon $f(x) = 2x$. □

Remark 2.9. For any real number x we have

$$\sqrt{x^2} = |x|. \quad \square \tag{2.20}$$

Assumption 2.1 (Square roots are always assumed nonnegative)**.** Remember that for any number a it is true that

$$a \cdot a = (-a)(-a) = a^2, \qquad \text{e.g.,} \quad 2^2 = (-2)^2 = 4,$$

or that, expressed in form of square roots, for any number $b \geq 0$

$$(+\sqrt{b})(+\sqrt{b}) = (-\sqrt{b})(-\sqrt{b}) = b.$$

We will always assume that "$\sqrt{b}$" is the **positive** value unless the opposite is explicitly stated.

Example: $\sqrt{9} = +3$, not -3. □

Proposition 2.5 (The Triangle Inequality for real numbers)**.** *The following inequality is used all the time in mathematical analysis to show that the size of a certain expression is limited from above:*

$$\textit{Triangle Inequality}: \ |\boldsymbol{a} + \boldsymbol{b}| \leq |\boldsymbol{a}| + |\boldsymbol{b}| \tag{2.21}$$

This inequality is true for any two real numbers a and b.

PROOF:

It is easy to prove this: just look separately at the three cases where both numbers are nonnegative, both are negative or where one of each is positive and negative. ■

2.4 A First Look at Functions and Sequences

The material on functions presented in this section will be discussed again and in greater detail in chapter **??** (Functions and Relations) on p.**??**.

Introduction 2.3. You are familiar with functions from calculus. Examples are $f_1(x) = \sqrt{x}$ and $f_2(x,y) = \ln(x-y)$. Sometimes $f_1(x)$ means the entire graph, i.e., the entire collection of pairs $\left(x, \sqrt{x}\right)$ and sometimes it just refers to the function value $\sqrt{x}$ for a "fixed but arbitrary" number x. In case of the function $f_2(x)$: Sometimes $f_2(x,y)$ means the entire graph, i.e., the entire collection of pairs $\left((x,y), \ln(x-y)\right)$ in the plane. At other times this expression just refers to the function value $\ln(x-y)$ for a pair of "fixed but arbitrary" numbers (x,y).

To obtain a usable definition of a function there are several things to consider. In the following $f_1(x)$ and $f_2(x,y)$ again denote the functions $f_1(x) = \sqrt{x}$ and $f_2(x,y) = \ln(x-y)$.

a. The source of all allowable arguments (x–values in case of $f_1(x)$ and (x,y)–values in case of $f_2(x,y)$) will be called the **domain** of the function. The domain is explicitly specified as part of a function definition and it may be chosen for whatever reason to be only a subset of all arguments for which the function value is a valid expression. In case of the function $f_1(x)$ this means that the domain must be restricted to a subset of the interval $[0, \infty[$ because the square root of a negative number cannot be taken. In case of the function $f_2(x,y)$ this means that the domain must be restricted to a subset of $\{\ (x,y) : x, y \in \mathbb{R}\ and\ x - y > 0\}$ because logarithms are only defined for strictly positive numbers.

b. The set to which all possible function values belong will be called the **codomain** of the function. As is the case for the domain, the codomain also is explicitly specified as part of a function definition. It may be chosen as any superset of the set of all function values for which the argument belongs to the domain of the function.

For the function $f_1(x)$ this means that we are OK if the codomain is a superset of the interval $[0, \infty[$. Such a set is big enough because square roots are never negative. It is OK to specify the interval $]-3.5, \infty[$ or even the set $\mathbb{R}$ of all real numbers as the codomain. In case of the function $f_2(x,y)$ this means that we are OK if the codomain contains $\mathbb{R}$. Not that it would make a lot of sense, but the set $\mathbb{R} \cup \{$ all inhabitants of Chicago $\}$ also is an acceptable choice for the codomain.

c. A function $y = f(x)$ is not necessarily something that maps (assigns) numbers or pairs of numbers to numbers. Rather domain and codomain can be a very different kind of animal. In chapter **??** on logic you will learn about statement functions $A(x)$ which assign arguments x from some set $\mathscr{U}$, called the universe of discourse, to statements $A(x)$, i.e., sentences that are either true or false.

d. Considering all that was said so far one can think of the graph of a function $f(x)$ with domain D and codomain C (see earlier in this note) as the set

$$\Gamma_f := \{\left(x, f(x)\right) : x \in D\}.$$

Alternatively one can characterize this function by the assignment rule which specifies how $f(x)$ depends on any given argument $x \in D$. We write "$x \mapsto f(x)$" to indicate this. You can also write instead $f(x) =$ whatever the actual function value will be.

This is possible if one does not write about functions in general but about specific functions such as $f_1(x) = \sqrt{x}$ and $f_2(x,y) = \ln(x-y)$. We further write

$$f : D \longrightarrow C$$

as a short way of saying that the function $f(x)$ has domain D and codomain C.
In case of the function $f_1(x) = \sqrt{x}$ for which we might choose the interval $X := [\,2.5, 7\,]$ as the domain (small enough because $X \subseteq [0, \infty[$) and $Y :=]1, 3[$ as the codomain (big enough because $1 < \sqrt{x} < 3$ for any $x \in X$) we specify this function as

$$\text{either } \; f_1 : [\,2.5, 7\,] \to]1, 3[; \quad x \mapsto \sqrt{x} \qquad \text{or } \; f_1 : [\,2.5, 7\,] \to]1, 3[; \quad f(x) = \sqrt{x}.$$

Let us choose $U := \{(x,y) : x, y \in \mathbb{R} \text{ and } 1 \leq x \leq 10 \text{ and } y < -2\}$ as the domain and $V := [0, \infty[$ as the codomain for $f_2(x,y) = \ln(x-y)$. These choices are OK because $x - y \geq 1$ for any $(x,y) \in U$ and hence $ln(x-y) \geq 0$, i.e., $f_2(x,y) \in V$ for all $(x, y \in U$. We specify this function as

$$\text{either } \; f_2 : U \to V, \quad (x,y) \mapsto \ln(x-y) \qquad \text{or } \; f_2 : U \to V, \quad f(x,y) = \ln(x-y). \;\square$$

We incorporate what we noted above into this definition of a function.

Definition 2.19 (Function)**.**

A **function** f consists of two nonempty sets X and Y and an assignment rule $x \mapsto f(x)$ which assigns any $x \in X$ uniquely to some $y \in Y$. We write $f(x)$ for this assigned value and call it the **function value** of the **argument** x. X is called the **domain** and Y is called the **codomain** of f. We write

$$f : X \to Y, \qquad x \mapsto f(x). \tag{2.22}$$

We read "$a \mapsto b$" as "a is assigned to b" or "a maps to b" and refer to $\mapsto$ as the **maps to operator** or **assignment operator**. The **graph** of such a function is the collection of pairs

$$\Gamma_f := \{(x, f(x)) : x \in X\}. \;\square \tag{2.23}$$

Remark 2.10. The name given to the argument variable is irrelevant. Let f_1, f_2, X, Y, U, V be as defined in **d** of the introduction to ch.2.4 (A First Look at Functions and Sequences). The function

$$g_1 : X \to Y, \quad p \mapsto \sqrt{p}$$

is identical to the function f_1. The function

$$g_2 : U \to V, \quad (t,s) \mapsto \ln(t-s)$$

is identical to the function f_2 and so is the function

$$g_3 : U \to V, \quad (s,t) \mapsto \ln(s-t).$$

The last example illustrates the fact that you can swap function names as long as you do it consistently in all places. □

We all know what it means that $f(x) = \sqrt{x}$ has the function $g(x) = x^2$ as its inverse function: f and f^{-1} cancel each other, i.e.,

$$g\big(f(x)\big) \;=\; f\big(g(x)\big) \;=\; x.$$

Definition 2.20 (Inverse function)**.**

Given are two nonempty sets X and Y and a function $f : X \to Y$ with domain X and codomain Y. We say that f has an **inverse function** if it satisfies all of the following conditions which uniquely determine this inverse function, so that we are justified to give it the symbol f^{-1}:

- **a.** $f^{-1} : Y \to X$, i.e., f^{-1} has domain Y and codomain X.
- **b.** $f^{-1}\big(f(x)\big) \;=\; x$ for all $x \in X$, and $f\big(f^{-1}(y)\big) \;=\; y$ for all $y \in Y$. □

Remark 2.11. You may recall that a function f has an inverse f^{-1} if and only if f is "onto" or **surjective**: for each $y \in Y$ there is at least one $x \in X$ such that $f(x) = y$, and if f is "one–one" or **injective**: for each $y \in Y$ there is at most one $x \in X$ such that $f(x) = y$. □

Example 2.5. Be sure you understand the following:

- **a.** $f : \mathbb{R} \to \mathbb{R};\ x \to e^x$ does not have an inverse $f^{-1}(y) = \ln(y)$ since its domain would have to be the codomain $\mathbb{R}$ of f and $\ln(y)$ is not defined for $y \leq 0$.
- **b.** $g : \mathbb{R} \to]0, \infty[;\ x \to e^x$ has the inverse $g^{-1} :]0, \infty[\to \mathbb{R};\ g^{-1}(y) = \ln(y)$ since

$$\begin{aligned} &Dom_{g^{-1}} = Cod_g =]0, \infty[, &&Cod_{g^{-1}} = Dom_g = \mathbb{R},\\ &e^{\ln(y)} = y \text{ for } 0 < y < \infty, &&\ln(e^x) = x \text{ for all } x \in \mathbb{R}. \ \square \end{aligned}$$

Definition 2.21 (Restriction/Extension of a function)**.**

Given are three nonempty sets A, X and Y such that $A \subseteq X$, and a function $f : X \to Y$ with domain X. We define the **restriction of f to A** as the function

(2.24) $$f\big|_A : A \to Y \quad \text{defined as} \quad f\big|_A(x) := f(x) \text{ for all } x \in A.$$

Conversely let $f : A \to Y$ and $\varphi : X \to Y$ be functions such that $f = \varphi\,|_A$. We then call φ an **extension** of f to X. □

2.5 Cartesian Products

We next define cartesian products of sets. [7] Those mathematical objects generalize rectangles

$$[a_1, b_1] \times [a_2, b_2] \;=\; \{(x, y) : x, y \in \mathbb{R}, a_1 \leq x \leq b_1 \text{ and } a_2 \leq y \leq b_2\}$$

[7] See ch.**??** (Cartesian Products and Relations) on p.**??** for the real thing and examples.

and quads

$$[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \;=\; \{(x, y, z) : x, y, z \in \mathbb{R}, a_1 \leq x \leq b_1, a_2 \leq y \leq b_2 \text{ and } a_3 \leq z \leq b_3\}.$$

Definition 2.22 (Cartesian Product)**.**

Let X and Y be two sets The set

$$X \times Y \;:= \{(x, y) : x \in X, y \in Y\} \tag{2.25}$$

is called the **cartesian product** of X and Y.
Note that the order is important: (x, y) and (y, x) are different unless $x = y$.
We write X^2 as an abbreviation for$X \times X$.
This definition generalizes to more than two sets: Let $X_1, X_2, \dots, X_n$ be sets. The set

$$X_1 \times X_2 \cdots \times X_n \;:= \{(x_1, x_2, \dots, x_n) : x_j \in X_j \text{ for each } j = 1, 2, \dots n\} \tag{2.26}$$

is called the cartesian product of $X_1, X_2, \dots, X_n$.
We write X^n as an abbreviation for$X \times X \times \cdots \times X$. □

Example 2.6. The graph Γ_f of a function with domain X and codomain Y (see def.2.23) is a subset of the cartesian product $X \times Y$. □

Example 2.7. The domains given in **a** and **d** of the introduction to ch.2.4 (A First Look at Functions and Sequences) are subsets of the cartesian product

$$\mathbb{R}^2 \;=\; \mathbb{R} \times \mathbb{R} \;=\; \{(x, y) : x, y \in \mathbb{R}\} \;\square$$

.

2.6 Sequences and Families

We now briefly discuss (infinite) sequences, subsequences, finite sequences and families.

Definition 2.23.

Let $n_\star$ be an integer and let let there be an item x_j for each integer $j \geq n_\star$ Such an item can be a number or a set (the only items we are looking at for now). In other words, we have an item x_j assigned to each $j \in [n_\star, \infty[_\mathbb{Z}$. We write $(x_n)_{n \geq n_\star}$ or $(x_j)_{j=n_\star}^\infty$ or $x_{n_\star}, x_{n_\star+1}, x_{n_\star+2}, \dots$ for such a collection of items and we call it a **sequence** with **start index** $n_\star$.

For example if $u_k = k^2$ for $k \in \mathbb{Z}$ then then $(u_k)_{k \geq -2}$ is the sequence of integers $4, 1, 0, 1, 4, 9, 16, \dots$.
The second example is a sequence of sets. If $A_j = [-1 - \frac{1}{j}, 1 + \frac{1}{j}] = \{x \in \mathbb{R} : -1 - \frac{1}{j} \leq x \leq 1 + \frac{1}{j}\}$ then $(A_j)_{j \geq 3}$ is the sequence of intervals (of real numbers) $[-\frac{4}{3}, \frac{4}{3}],\ [-\frac{5}{4}, \frac{5}{4}],\ [-\frac{6}{5}, \frac{6}{5}], \dots$.
One can think of a sequence $(x_i)_{i \geq n_\star}$ in terms of the assignment $i \mapsto x_i$ and this sequence can then be interpreted as the function

$$x : [n_\star, \infty[_\mathbb{Z} \longrightarrow \text{ suitable codomain}; \quad i \mapsto x(i) := x_i,$$

where that "suitable codomain" depends on the nature of the items x_i. In example 1 ($u_k = k^2$ for $k \in \mathbb{Z}$) we could chose $\mathbb{Z}$ as that codomain, in example 2 ($A_j = [-1 - \frac{1}{j}, 1 + \frac{1}{j}]$) we could choose $2^{\mathbb{R}}$, the power set of $\mathbb{R}$.

We will occasionally also admit an "ending index" $n^\star$ instead of ∞, i.e., there will be an indexed item x_j for each $j \in [n_\star, n^\star]_{\mathbb{Z}}$. We then talk of a **finite sequence**, and we write $(x_n)_{n_\star \leq n \leq n^\star}$ or $(x_j)_{j=n_\star}^{n^\star}$ or $x_{n_\star}, x_{n_\star+1}, \dots, x_{n^\star}$ for such a finite collection of items. If we refer to a sequence $(x_n)_n$ without qualifying it as finite then we imply that we deal with an **infinite sequence**, $(x_n)_{n=n_\star}^{\infty}$.

If one pares down the full set of indices $\{n_\star,\ n_\star+1,\ n_\star+2, \dots\}$ to a subset $\{n_1,\ n_2,\ n_3, \dots\}$ such that $n_\star \leq n_1 < n_2 < n_3 < \dots$ then we call the corresponding thinned out sequence $(x_{n_j})_{j \in \mathbb{N}}$ a **subsequence** of the sequence $(x_n)_{n \geq m}$.
If this subset of indices is finite, i.e., we have $n_\star \leq n_1 < n_2 < \dots < n_K$ for some suitable $K \in \mathbb{N}$ then we call $(x_{n_j})_{j=1}^{K}$ a **finite subsequence** of the original sequence. $\square$

We will later define a stochastic process as a "family" $(Z_t)_{t \in I}$ where I is an interval of real numbers and each indexed item Z_t is a random variable. Typical choices for I would be

$$I = [0, T] \text{ (where } T > 0), \quad I = [0, \infty[, \quad I = [t_0, T] \text{ (where } 0 \leq t \leq T), \dots$$

Here is the formal definition of a family.

Definition 2.24 (Indexed families)**.**

Let J and X be nonempty sets and assume that

for each $j \in J$ there exists **exactly one** indexed item $x_j \in X$.

- **a.** $(x_j)_{j \in J}$ is called an **indexed family** aka **family** in X.
- **b.** J is called the **index set** of the family.
- **c.** For each $j \in J$, x_j is called a **member of the family** $(x_j)_{j \in J}$. $\square$

Some remarks:

- A family is completely defined by the assignment $j \mapsto x_j$. In that sense a family behaves like a function
$$F: J \to X, \qquad j \mapsto F(j) := x_j.$$
- j is a dummy variable: $(x_j)_{j \in J}$ and $(x_k)_{k \in J}$ describe the same family as long as $j \mapsto x_j$ and $k \mapsto x_k$ describe the same assignment.
- Sequences $(x_n) : n \in \mathbb{N}$ are families with index set $\mathbb{N}$.

2.7 Proofs by Induction and Definitions by Recursion

Introduction 2.4. The integers have a property which makes them fundamentally different from the rational numbers (fractions) and the real numbers: Given any two integers $m < n$, there are only finitely many integers between m and n. To be precise, there are exactly $n-m-1$ of them. For example, there are only 4 integers between 12 and 17: the numbers $13, 14, 15, 16$. [8]

Therefore, given an integer n, we have the concept of its predecessor, $n-1$, and its successor, $n+1$. This has some profound consequences. If we know what to do for a certain starting number $k_0 \in \mathbb{Z}$ (we call this number the base case), and if we can figure out for each integer $k \geq k_0$ what to do for $k+1$ if only we know what to do for k, then we know what to do for **any** $k \geq k_0$! $\square$

We make use of the above when defining a sequence by **recursion**. Here is a simple example.

Example 2.8. Let $k_0 = -2$, $x_{k_0} = 5$ (base case), and $x_{k+1} = x_k + 3$ (i.e., we know how to obtain x_{k+1} just from the knowledge of x_k), then we know how to build the entire sequence

$$x_{-2} = 5,\ x_{-1} = x_{-2} + 3 = 8,\ x_0 = x_{-1} + 3 = 11,\ x_1 = x_0 + 3 = 14,\ \dots,$$

The equation $x_{k+1} = x_k + 3$ which tells us how to obtain the next item from the given one is the **recurrence relation** for that recursive definition. $\square$

Example 2.9. Given is a sequence of sets $A_1, A_2, \dots$. For $n \in \mathbb{N}$ we define $\bigcup\limits_{j=1}^{n} A_j$ and $\bigcap\limits_{j=1}^{n} A_j$ recursively as follows. [9]

$$\bigcup_{j=1}^{1} A_j \ := \ A_1, \qquad \bigcup_{j=1}^{n+1} A_j \ := \left(\bigcup_{j=1}^{n} A_j \right) \cup A_{n+1}, \tag{2.27}$$

$$\bigcap_{j=1}^{1} A_j \ := \ A_1, \qquad \bigcap_{j=1}^{n+1} A_j \ := \left(\bigcap_{j=1}^{n} A_j \right) \cap A_{n+1}. \tag{2.28}$$

this tells us the meaning of $\bigcup\limits_{j=1}^{n} A_j$ and $\bigcap\limits_{j=1}^{n} A_j$ for any natural number n. For example, $\bigcap\limits_{j=1}^{4} A_j$ is computed as follows.

$$\begin{aligned}
\bigcap_{j=1}^{1} A_j &= A_1, \\
\bigcap_{j=1}^{2} A_j &= \left(\bigcap_{j=1}^{1} A_j \right) \cap A_2 \ = \ A_1 \cap A_2, \\
\bigcap_{j=1}^{3} A_j &= \left(\bigcap_{j=1}^{2} A_j \right) \cap A_3 \ = \ (A_1 \cap A_2) \cap A_3, \\
\bigcap_{j=1}^{4} A_j &= \left(\bigcap_{j=1}^{3} A_j \right) \cap A_4 \ = \ \big((A_1 \cap A_2) \cap A_3\big) \cap A_4. \ \square
\end{aligned}$$

[8] All of this will be made mathematically precise in ch.**??** on p.**??**.

[9] An "official" definition for unions and intersections of arbitrarily many sets (not just for finitely many) will be given in def.3.2 on p.38.

Remark 2.12. The discrete structure of the integers can also be used as a means to prove a collection of mathematical statements $P(k_0), P(k_0+1), P(k_0+2), \dots$ which is defined for all integers k, starting at $k_0 \in \mathbb{Z}$. Each $P(k)$ might be an equation or an inequality for two numeric expressions that depend on k. It could also be a relation between sets or it could be something entirely different. For example, $P(k)$ could be the statement $\left(\bigcup_{j=1}^{k} A_j\right) \cap B = \bigcup_{j=1}^{k} (A_j \cap B)$. An extremely important tool for proofs of this kind is the following principle. Its mathematical justification will be given later in thm.**??** on p.**??**.

Principle of Mathematical Induction
Assume that for each integer $k \geq k_0$ there is an associated statement $P(k)$ such that the following is valid:

A. Base case. The statement $P(k_0)$ is true.

B. Induction Step. For each $k \geq k_0$ we have the following: Assuming that $P(k)$ is true ("**Induction Assumption**"), it can be shown that $P(k+1)$ also is true.

It then follows that $P(k)$ is true for **each** $k \geq k_0$.

Here is an example which generalizes prop.2.2 on p.13.

Proposition 2.6 (Distributivity of unions and intersections for finitely many sets)**.**

Let $A_1, A_2, \dots$ and B be sets. If $n \in \mathbb{N}$, then

$$\left(\bigcup_{j=1}^{n} A_j\right) \cap B = \bigcup_{j=1}^{n} (A_j \cap B), \tag{2.29}$$

$$\left(\bigcap_{j=1}^{n} A_j\right) \cup B = \bigcap_{j=1}^{n} (A_j \cup B). \tag{2.30}$$

PROOF: We only prove (2.29), and this will be done by induction on n, i.e., the number of sets A_j. The proof of (2.30) is left as exercise 2.11

A. Base case: $k_0 = 1$. The statement $P(1)$ is (2.29) for $n = 1$: $\left(\bigcup_{j=1}^{1} A_j\right) \cap B = \bigcup_{j=1}^{1} (A_j \cap B)$. We must prove that $P(1)$ is true. According to our recursive definition of finite unions which was given in example 2.8 this is the same as $(A_1) \cap B = (A_1 \cap B)$, and this is a true statement. We have proven the base case.

B. Induction step:

(2.31) **Induction assumption:** $P(k) : \left(\bigcup_{j=1}^{k} A_j\right) \cap B = \bigcup_{j=1}^{k} (A_j \cap B)$ is true for some $k \geq 1$.

Under this assumption

$$\text{(2.32)}\qquad \text{we must prove the truth of } P(k+1):\ \Big(\bigcup_{j=1}^{k+1} A_j\Big)\cap B \ = \ \bigcup_{j=1}^{k+1}(A_j\cap B).$$

The trick is to manipulate $P(k+1)$ in a way that allows us to "plug in" the induction assumption. For (2.32) one way to do this is to take the left–hand side and transform it repeatedly until we end up with the right–hand side, and doing so in such a manner that (2.31) will be used at some point.

$$\begin{aligned}
\Big(\bigcup_{j=1}^{k+1} A_j\Big)\cap B &= \Big(\Big(\bigcup_{j=1}^{k} A_j\Big)\cup A_{n+1}\Big)\cap B && \text{we used (2.27)}\\
&= \Big(\Big(\bigcup_{j=1}^{k} A_j\Big)\cap B\Big)\cup(A_{n+1}\cap B) && \text{we used (2.11) on p. 13}\\
&= \bigcup_{j=1}^{k}(A_j\cap B)\cup(A_{n+1}\cap B) && \textbf{we used the induction assumption!}\\
&= \bigcup_{j=1}^{k+1}(A_j\cap B) && \text{we used (2.27)}
\end{aligned}$$

We have managed to establish the truth of $P(k+1)$, and this concludes the proof.

Epilogue: It is crucial that you understand in what way the induction assumption was used to get from the left–hand side of (2.32) to the right–hand side, and that you first had to find a base from which to proceed by proving the base case. ■

Proposition 2.7 (The Triangle Inequality for n real numbers)**.**

Let $n \in \mathbb{N}$ such that $n \geq 2$. Let $a_1, a_2, \dots, a_n \in \mathbb{N}$. Then,

$$\text{(2.33)}\qquad |a_1 + a_2 + \dots + a_n| \ \leq \ |a_1| + |a_2| + \dots + |a_n|$$

PROOF: Note that this proposition generalizes prop.2.5 on p.21 from 2 terms to n terms. The proof will be done by induction on n, the number of terms in the sum.

A. Base case: For $k_0 = 2$, inequality 2.33 was already shown (see (2.21) on p.21).

B. Induction step: Let us assume that 2.33 is true for some $k \geq 2$. This is our induction assumption. We now must prove the inequality for $k+1$ terms $a_1, a_2, \dots, a_k, a_{k+1} \in \mathbb{N}$. We abbreviate

$$A \ := \ a_1 + a_2 + \dots + a_k; \qquad B \ := \ |a_1| + |a_2| + \dots + |a_k|$$

then our induction assumption for k numbers is that $|A| \leq B$. We know from (2.21) that the triangle inequality is valid for the two terms A and a_{k+1}. It follows that $|A+a_{k+1}| \leq |A|+|a_{k+1}|$. We combine those two inequalities and obtain

$$\text{(2.34)}\qquad |A + a_{k+1}| \ \leq \ |A| + |a_{k+1}| \ \leq \ B + |a_{k+1}|$$

In other words,

$$\text{(2.35)}\qquad \big|(a_1 + a_2 + \dots + a_k) + a_{k+1}\big| \ \leq \ B + |a_{k+1}| \ = \ \big(|a_1| + |a_2| + \dots + |a_k|\big) + |a_{k+1}|,$$

and this is (2.33) for $k+1$ rather than k numbers: We have shown the validity of the triangle inequality for $k+1$ items under the assumption that it is valid for k items. It follows from the induction principle that the inequality is valid for any $k \geq k_0 = 2$. ■

To summarize what we did in all of part B: We were able to show the validity of the triangle inequality for $k+1$ numbers under the assumption that it was valid for k numbers.

Remark 2.13 (Why induction works)**.** But how can we from all of the above conclude that the distributivity formulas of prop.2.6 and the triangle inequality of prop.2.7 work for all $n \in \mathbb{N}$ such that $n \geq k_0$? We illustrate this for the triangle inequality.

Step 1:	We know that the statement is true for $k_0 = 2$ because that was proven in the base case.
Step 2:	But according to the induction step, if it is true for $k_0 = 2$, it is also true for the successor $k_0 + 1 = 3$ of 2.
Step 3:	But according to the induction step, if it is true for $k_0 + 1$, it is also true for the successor $(k_0 + 1) + 1 = 4$ of $k_0 + 1$.
Step 4:	But according to the induction step, if it is true for $k_0 + 2$, it is also true for the successor $(k_0 + 2) + 1 = 5$ of $k_0 + 2$.
......	....
Step $53,920$:	But according to the induction step, if it is true for $k_0 + 53,918$, it is also true for the successor $(k_0 + 53,918) + 1 = 53,921$ of $k_0 + 53,918$.
......	....

And now we see why the statement is true for any natural number $n \geq k_0$. ☐

2.8 Some Preliminaries From Calculus

Remark 2.14. To understand this remark you need to be familiar with the concepts of continuity, differentiability and antiderivatives (integrals) of functions of a single variable. Just skip the parts where you lack the background.

The following is known from calculus (see [15] Stewart, J: Single Variable Calculus): Let $a \in \mathbb{R} \cup \{-\infty\}$ and $b \in \mathbb{R} \cup \{\infty\}$ and let $X :=]a,b[$ be the open (end points a, b are excluded) interval of all real numbers between a and b. Let $x_0 \in]a,b[$ be "fixed but arbitrary".

Let $f :]a,b[\to \mathbb{R}$ be a function which is continuous on $]a,b[$. Then

a. f is integrable for any $\alpha, \beta \in \mathbb{R}$ such that $a < \alpha < \beta < b$, i.e., the **definite integral** $\int_\alpha^\beta f(u)du$ exists. For a definition of integrability see, e.g., [15] Stewart, J: Single Variable Calculus.

b. Integration is "linear", i.e., it is additive: $\displaystyle\int_\alpha^\beta \big(f(u) + g(u)\big)du = \int_\alpha^\beta f(u)du + \int_\alpha^\beta g(u)du$, and you also can "pull out" constant $\lambda \in \mathbb{R}$: $\displaystyle\int_\alpha^\beta \lambda f(u)du = \lambda \int_\alpha^\beta f(u)du$.

c. Integration is "monotonic":
If $f(x) \leq g(x)$ for all $\alpha \leq x \leq \beta$ then $\displaystyle\int_\alpha^\beta \big(f(u)\big)du \leq \int_\alpha^\beta g(u)du$.

d. f has an **antiderivative**: There exists a function $F :]a,b[\to \mathbb{R}$ whose derivative $F'(\cdot)$ exists on all of $]a,b[$ and coincides with f, i.e., $F'(x) = f(x)$ for all $x \in]a,b[$.

e. This antiderivative satisfies $F(\beta) - F(\alpha) = \int_\alpha^\beta f(u)du$ for all $a < \alpha < \beta < b$ and it is **not** uniquely defined: If $C \in \mathbb{R}$ then $F(\cdot) + C$ is also an antiderivative of f.
On the other hand, if both F_1 and F_2 are antiderivatives for f then their difference $G(\cdot) := F_2(\cdot) - F_1(\cdot)$ has the derivative $G'(\cdot) = f(\cdot) - f(\cdot)$ which is constant zero on $]a, b[$. It follows that any two antiderivatives only differ by a constant.
To summarize the above: If we have one antiderivative F of f then any other antiderivative $\tilde{F}$ is of the form $\tilde{F}(\cdot) = F(\cdot) + C$ for some real number C.
This fact is commonly expressed by writing $\int f(x)dx = F(x) + C$ for the **indefinite integral** (an integral without integration bounds).

f. It follows from **e** that if some $c_0 \in \mathbb{R}$ is given then there is only one antiderivative F such that $F(x_0) = c_0$.
Here is a quick proof: Let G be another antiderivative of f such that $G(x_0) = c_0$. Because $G - F$ is constant we have for all $x \in]a, b[$ that

$$G(x) - F(x) \;=\; \text{const} \;=\; G(x_0) - F(x_0) \;=\; 0,$$

i.e., $G = F$. $\square$

2.9 Convexity ⋆

Note that this chapter is starred, hence optional.

Definition 2.25 (Concave-up and convex functions)**.** ⋆

Let $-\infty \le \alpha < \beta \le \infty$ and let $I :=]\alpha, \beta[$ be the open interval of real numbers with endpoints α and β. Let $f : I \to \mathbb{R}$.

- **a.** The **epigraph** of f is the set $epi(f) := \{(x_1, x_2) \in I \times \mathbb{R} : x_2 \ge f(x_1)\}$ of all points in the plane that lie above the graph of f.
- **b.** f is **convex** if for any two vectors $\vec{a}, \vec{b} \in epi(f)$ the entire line segment $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \le \lambda \le 1$ is contained in $epi(f)$. See Figure 2.4.
- **c.** Let f be differentiable at all points $x \in I$. Then f is **concave-up**, if the function $f' : x \mapsto f'(x) = \frac{df}{dx}(x)$ is increasing. $\square$

Convexity is illustrated in the figure below. [10]

Proposition 2.8 (Convexity criterion)**.**

[10] Source: Wikipedia, https://upload.wikimedia.org/wikipedia/commons/c/c7/ConvexFunction.svg.

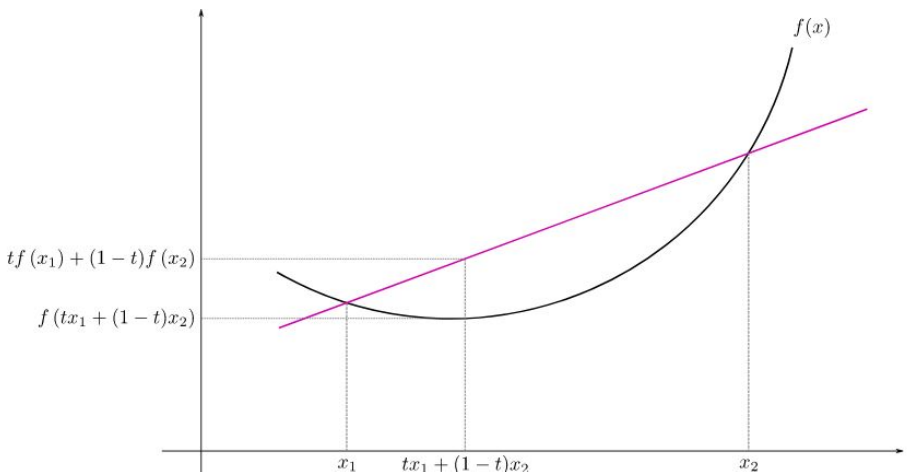


Figure 2.4: Convex function

f is convex if and only if the following is true: For any

$$\alpha \ < \ a \ \le \ x_0 \ \le \ b \ < \ \beta$$

let $S(x_0)$ be the unique number such that the point $(x_0, S(x_0))$ is on the line segment that connects the points $(a, f(a))$ and $(b, f(b))$. Then,

(2.36) $$f(x_0) \ \le \ S(x_0).$$

Note that any x_0 between a and b can be written as $x_0 = \lambda a + (1-\lambda)b$ for some $0 \le \lambda \le 1$ and that the corresponding y-coordinate $S(x_0) = S(\lambda a + (1-\lambda)b)$ on the line segment that connects $(a, f(a))$ and $(b, f(b))$ then is $S(\lambda a + (1-\lambda)b) = \lambda f(a) + (1-\lambda)f(b)$. Hence we can rephrase the above as follows:
f is convex if and only if for any $a < b$ such that $a, b \in I$ and $0 \le \lambda \le 1$ it is true that

(2.37) $$f(\lambda a + (1-\lambda)b) \ \le \lambda f(a) + (1-\lambda)f(b).$$

PROOF of "⇒": Any line segment S that connects the points $(a, f(a))$ and $(b, f(b))$ in such a way that S is entirely contained in the epigraph of f will satisfy $(x_0, S(x_0)) \in epi(f)$ and hence $f(x_0) \le S(x_0)$ for all $a \le x_0 \le b$. It follows that convexity of f implies (2.36).

PROOF of "⇐": Let (2.36) be valid for all $a, b \in I$. Let $\vec{a} = (a_1, a_2), \vec{b} = (b_1, b_2) \in epi(f)$. Then

(2.38) $$a_2 \ge f(a_1) \ \text{ and } \ b_2 \ge f(b_1).$$

We must show that the entire line segment $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \le \lambda \le 1$ is contained in $epi(f)$.

Let $\vec{a}' := (a_1, f(a_1))$. Let $S' := \{\lambda\vec{a}' + (1-\lambda)\vec{b} : 0 \le \lambda \le 1\}$ be the line segment obtained by leaving the right endpoint $\vec{b}$ unchanged and pushing the left one downward until a_2 matches $f(a_1)$. Clearly, S' nowhere exceeds S.

Let $\vec{b}'' := (b_1, f(b_1))$. Let $S'' := \{\lambda\vec{a}' + (1-\lambda)\vec{b}' : 0 \le \lambda \le 1\}$ be the line segment obtained by leaving the left endpoint $\vec{a}'$ unchanged and pushing the right one downward until the b_2 matches $f(b_1)$. Clearly, S'' nowhere exceeds S'.

We view any line segment T between two points with abscissas a_1 and b_1 as a function $T(\cdot) : [a_1, b_1] \to \mathbb{R}$ which assigns to $x \in [a_1, b_1]$ that unique value $T(x)$ for which the point $\big(x, T(x)\big)$ lies on T.

The segment S'' connects the points $(a, f(a))$ and $(b, f(b))$ and it follows from assumption **b** that for any $a \leq x_0 \leq b$ we have $f(x_0) \leq S''(x_0)$. We conclude from $S(\cdot) \geq S'(\cdot) \geq S''(\cdot)$ that $S(x_0) \geq f(x_0)$, i.e., $(x_0, S(x_0)) \in \mathit{epi}(f)$. As this is true for any $a \leq x_0 \leq b$ it follows that the line segment S is entirely contained in the epigraph of f. ■

Proposition 2.9 (Convex vs concave-up).

Let $f : \mathbb{R} \to \mathbb{R}$ be concave-up. Then f is convex.

PROOF: Assume to the contrary that f is (differentiable and) concave-up and that there are $a, b, x_0 \in I$ such that $a < x_0 < b$ and $f(x_0) > S(x_0)$. Here $S(x_0)$ denotes the unique number such that the point $(x_0, S(x_0))$ is on the line segment that connects the points $(a, f(a))$ and $(b, f(b))$.

Let m be the slope of the linear function $S(\cdot) : x \mapsto S(x)$, i.e.,

$$m = \frac{S(b) - S(a)}{b - a}.$$

It follows that

$$m = \frac{S(b) - S(x_0)}{b - x_0} > \frac{S(b) - f(x_0)}{b - x_0} = \frac{f(b) - f(x_0)}{b - x_0} = f'(\xi) \tag{2.39}$$

for some $x_0 < \xi < b$ (according to the mean value theorem for derivatives). Further

$$m = \frac{S(x_0) - S(a)}{x_0 - a} < \frac{f(x_0) - S(a)}{x_0 - a} = \frac{f(x_0) - f(a)}{x_0 - a} = f'(\eta) \tag{2.40}$$

for some $a < \eta < x_0$ (according to the mean value theorem for derivatives).

Because f is concave up we have

$$f'(a) \leq f'(\eta) \leq f'(x_0) \leq f'(\xi) \leq f'(b).$$

From (2.39) and (2.40) we obtain

$$m < f'(\eta) \leq f'(x_0) \leq f'(\xi) < m,$$

and we have reached a contradiction. ■

If a convex function f is differentiable at some argument x, i.e., f possesses a tangent at x, then the graph of this tangent will stay below the graph of f. (Draw a picture!) The following proposition generalizes this convex functions in general, without any differentiability requirements.

Proposition 2.10.

Let $-\infty \leq \alpha < \beta \leq \infty$, I an interval with endpoints α and β where α and/or β may or may not belong to I, and let $f : I \to \mathbb{R}$ be convex. Let

$$\mathscr{L} := \{ I \xrightarrow{L} \mathbb{R} : L(x) = mx + b \text{ for suitable } m, b \in \mathbb{R} \text{ and } L \leq f \}, \tag{2.41}$$

i.e., the graph of L is a straight line and it is dominated by the graph of f. Then

$$f(x) = \sup\{ L(x) : L \in \mathscr{L} \} \quad \text{for all } x \in I. \tag{2.42}$$

PROOF: Can be found, e.g., in [4] Bauer, Heinz: Measure and Integration Theory. ■

Proposition 2.11 (Sublinear functions are convex)**.**

Let $f : \mathbb{R} \to \mathbb{R}$ be sublinear. Then f is convex.

PROOF: Let $0 \leq \lambda \leq 1$ and $x, y \in \mathbb{R}$. Then

$$p(\lambda x + (1-\lambda)y) \leq p(\lambda x) + p((1-\lambda)y) = \lambda p(x) + (1-\lambda)p(y). \tag{2.43}$$

It follows from prop.2.8 that f is concave-up. ■

2.10 Miscellaneous

Proposition 2.12. ⋆

Let $A = \big((a_{ij})\big)$; $(i = 1, \dots, m;\ j = 1 \dots, n)$, be an $m \times n$ matrix. We can think of A as a function

$$A : \mathbb{R}^n \to \mathbb{R}^m; \quad \vec{x} \mapsto A\,\vec{x},$$

which assigns to the column vector $\vec{x} \in \mathbb{R}^n$, viewed as a $n \times 1$ matrix, the matrix product $\vec{y} = A\vec{x}$, an $m \times 1$ matrix which we view as an element of $\mathbb{R}^m$.
Let $A^\top = \big((a^*_{k\ell})\big)$ denote the transpose of A, i.e., the $n \times m$ matrix one obtains by switching rows and columns. In other words, $a^*_{k\ell} = a_{\ell k}$. Matrix multiplication with $m \times 1$ vectors $\vec{\eta}$ makes $A^\top$ a function

$$A^\top : \mathbb{R}^m \to \mathbb{R}^n; \quad \vec{\eta} \mapsto A^\top\,\vec{\eta}.$$

The following is true. [11]

$$A \text{ is surjective} \quad \Leftrightarrow \quad A^\top \text{ is injective.}$$

PROOF: Consult a book on linear algebra. ■

Corollary 2.1. *Let $A = \big((a_{ij})\big)$ be a matrix with m rows and n columns. Then* ***(a)*** $\Leftrightarrow$ ***(b)****, where*

(a) *The set of m linear equations in n unknowns $\vec{x} = (x_1, \dots, x_n)^\top$,*

$$A\,\vec{x} = \vec{y},$$

has a solution $\vec{x}$ for any choice of right hand side $\vec{y} = (y_1, \dots, y_m)^\top$.

(b) *the set of n linear equations in m unknowns $\vec{\xi} = (\xi_1, \dots, \xi_m)^\top$,*

$$A^\top\,\vec{\xi} = \vec{\eta},$$

has at most one solution $\vec{\xi}$ for any $\vec{\eta} = (\eta_1, \dots, \eta_n)^\top$.

PROOF: This is a direct translation of Proposition 2.12 from the language of matrix multiplication to that of systems of linear equations. ■

2.11 Exercises for Ch.2

2.11.1 Exercises for Sets

Exercise 2.1. Prove (2.12) of prop.2.2 on p.13.

Exercise 2.2. Prove the set identities of prop.2.1.

Exercise 2.3. Prove that for any three sets A, B, C it is true that $(A \setminus B) \setminus C \;=\; A \setminus (B \cup C)$.
Hint: use De Morgan's formula (2.13.**a**). ■

Exercise 2.4. Let $X = \{x, y, \{x\}, \{x, y\}\ \}$. True or false?

a. $\{x\} \in X$ **c.** $\{\ \{x\}\ \} \in X$ **e.** $y \in X$ **g.** $\{y\} \in X$
b. $\{x\} \subseteq X$ **d.** $\{\ \{x\}\ \} \subseteq X$ **f.** $y \subseteq X$ **h.** $\{y\} \subseteq X$ □

For the subsequent exercises refer to example **??** for the definition of the size $|A|$ of a set A and to def.**??** (Cartesian Product of Two Sets) for the definition of Cartesian product. You find both in ch.**??** (Cartesian Products and Relations) on p.**??**

Exercise 2.5. Find the size of each of the following sets:

a. $A = \{x, y,\ \{x\},\ \{x, y\}\ \}$ **c.** $C = \{u, v, v, v, u\}$ **e.** $E = \{\sin(k\pi/2) : k \in \mathbb{Z}\}$
b. $B = \{1,\ \{0\},\ \{1\}\ \}$ **d.** $D = \{3z - 10 : z \in \mathbb{Z}\}$ **f.** $F = \{\pi x : x \in \mathbb{R}\}$ □

Exercise 2.6. Let $X = \{x, y, \{x\}, \{x, y\}\ \}$ and $Y = \{x, \{y\}\ \}$. True or false?

a. $x \in X \cap Y$ **c.** $x \in X \cup Y$ **e.** $x \in X \setminus Y$ **g.** $x \in X \Delta Y$
b. $\{y\} \in X \cap Y$ **d.** $\{y\} \in X \cup Y$ **f.** $\{y\} \in X \setminus Y$ **h.** $\{y\} \in X \Delta Y$ □

Exercise 2.7. Let $X = \{1, 2, 3, 4\}$ and let $Y = \{x, y\}$.

a. What is $X \times Y$? **c.** What is $|X \times Y|$? **e.** Is $(x, 3) \in X \times Y$? **g.** Is $3 \cdot x \in X \times Y$?
b. What is $Y \times X$? **d.** What is $|X \times Y|$? **f.** Is $(x, 3) \in Y \times X$? **h.** Is $2 \cdot y \in Y \times X$? □

Exercise 2.8. Let $X = \{8\}$. What is $2^{(2^X)}$?

Exercise 2.9. Let $A = \{1, \{1, 2\}, 2, 3, 4\}$ and $B = \{\{2, 3\}, 3, \{4\}, 5\}$. Compute the following.

a. $A \cap B$ **b.** $A \cup B$ **c.** $A \setminus B$ **d.** $B \setminus A$ **e.** $A \triangle B$ □

Exercise 2.10. Let A, X be sets such that $A \subseteq X$ and let $x \in X$. Prove the following:

a. If $a \in A$ then $A \;=\; (A \setminus \{a\}) \uplus \{a\}$.
b. If $a \notin A$ then $A \;=\; (A \uplus \{a\}) \setminus \{a\}$.
□

2.11.2 Exercises for Proofs by Induction

Exercise 2.11. Use induction on n to prove (2.30) of prop.2.6 on p.28 of this document: Let $A_1, A_2, \dots$ and B be sets. If $n \in \mathbb{N}$ then $\left(\bigcap\limits_{j=1}^{n} A_j\right) \cup B \;=\; \bigcap\limits_{j=1}^{n} (A_j \cup B)$. □

Exercise 2.12. [12]

Let $K \in \mathbb{N}$ such that $K \geq 2$ and $n \in \mathbb{Z}_{\geq 0}$. Prove that $K^n > n$. $\square$

Exercise 2.13. Let $n \in \mathbb{N}$. Then $n^2 + n$ is even, i.e., this expression is an integer multiple of 2. $\square$

PROOF: The proof is given in this instructor's edition.

The proof is done by induction on n.

The base case ($n_0 = 1$) holds because $1^2 + 1 = 2$, and this is an even number.

Induction step: Let $k \in \mathbb{N}$.

(2.44) Induction assumption: $k^2 + k$ is even, i.e., $k^2 + k = 2j$ for some suitable $j \in \mathbb{Z}$.

We must show that there exists $j' \in \mathbb{Z}$ such that $(k+1)^2 + k + 1 = 2j'$. We have

$$(k+1)^2 + k + 1 = k^2 + 2k + 1 + k + 1 = (k^2 + k) + 2(k+1) \overset{(2.44)}{=} 2j + 2(k+1).$$

Let $j' := j + k + 1$. Then $(k+1)^2 + k + 1 = 2j'$ and this finishes the proof. $\blacksquare$

Exercise 2.14. Use the result from exercise 2.13 above to prove by induction that $n^3 + 5n$ is an integer multiple of 6 for all $n \in \mathbb{N}$. $\square$

PROOF: The proof is given in this instructor's edition.

The proof is done by induction on n.

The base case ($n_0 = 1$) holds because $1^3 + 5 = 6 = 1 \cdot 6$.

Induction step: Let $k \in \mathbb{N}$.

(2.45)
Induction assumption: $k^3 + 5k$ is an integer multiple of 6, i.e., $k^3 + 5k = 6j$ for some $j \in \mathbb{Z}$.

We must show that there exists $j' \in \mathbb{Z}$ such that $(k+1)^3 + 5(k+1) = 6j'$. We know frome exercise 2.13 that $3(k^2 + k) = 3 \cdot 2 \cdot i$ for a suitable $i \in \mathbb{Z}$, hence

$$\begin{aligned}(k+1)^3 + 5(k+1) &= k^3 + 3k^2 + 3k + 1 + 5k + 5 = (k^3 + 5k) + 3(k^2 + k) + 6 \\ &= (k^3 + 5k) + 6i + 6 \overset{(2.45)}{=} 6(j + i + 1).\end{aligned}$$

Let $j' := j + i + 1$. Then $(k+1)^3 + 5(k+1) = 6j'$ and this finishes the proof. $\blacksquare$

Exercise 2.15. Let $x_1 = 1$ and $x_{n+1} = x_n + 2n + 1$. Prove by induction that $x_n = n^2$ for all $n \in \mathbb{N}$. $\square$

[12] Note that this exercise generalizes B/G prop.7.1: If $n \in \mathbb{N}$ then $n < 10^n$. Also note that if you allow K to be a real number rather than an integer then it will not be true for all $K > 1$ and $n \in \mathbb{Z}_{\geq 0}$. For example $K^n > n$ is false for $K = 1.4$ and $n = 2$ (but it is true for $K = 1.5$ and $n = 2$).

2.12 Blank Page after Ch.2

This page is intentionally left blank!

3 More on Sets and Functions

3.1 More on Set Operations

We will not deal with limits of sequences of sets except for the following since it is so suggestive.

Definition 3.1 (Notation for limits of monotone sequences of sets)**.**

Let (A_n) be a **increasing sequence of sets**, i.e., $A_1 \subseteq A_2 \subseteq \dots$ and let $A := \bigcup_n A_n$.
Further let B_n be a **decreasing sequence of sets**, i.e., $B_1 \supseteq B_2 \supseteq \dots$ and let $B := \bigcap_n B_n$.
We write suggestively

$$A_n \uparrow A \; (n \to \infty), \quad A = \lim_{n\to\infty} A_n, \qquad B_n \downarrow B \; (n \to \infty), \quad B = \lim_{n\to\infty} B_n. \; \square$$

We adopt the following convention.

Let $\mathfrak{E}$ be a set of sets, e.g., $\mathfrak{E}$ is a subset of the power set 2^Ω of a set Ω. Then a phrase such as
- "Let $U_n \uparrow$ in $\mathfrak{E}$" is shorthand notation for
  "Let $U_n \subseteq \mathfrak{E} \, (n \in \mathbb{N})$" be a increasing sequence."
- "Let $U_n \downarrow$ in $\mathfrak{E}$" is shorthand notation for
  "Let $U_n \subseteq \mathfrak{E} \, (n \in \mathbb{N})$" be a decreasing sequence."

Definition 3.2 (Arbitrary unions and intersections)**.**

Let J be a nonempty set and let $(A_i)_{i \in J}$ be a family of sets.
We define its **union**, $\bigcup\limits_{i \in I} A_i$, and its **intersection** $\bigcap\limits_{i \in I} A_i$, as follows:

$$\bigcup_{i \in I} A_i := \bigcup \left[A_i : i \in I \right] := \{x : x \in A_{i_0} \text{ for some } i_0 \in I\}, \tag{3.1}$$

$$\bigcap_{i \in I} A_i := \bigcap \left[A_i : i \in I \right] := \{x : x \in A_{i_0} \text{ for each } i_0 \in I\}. \tag{3.2}$$

It is convenient to allow unions and intersections for the empty index set $J = \emptyset$. For intersections this requires the existence of a universal set Ω. We define

$$\bigcup_{i \in \emptyset} A_i := \emptyset, \qquad \bigcap_{i \in \emptyset} A_i := \Omega. \; \square \tag{3.3}$$

Note that any statement concerning arbitrary families of sets such as the definition above covers finite lists $A_1, A_2, \dots, A_n$ of sets ($J = \{1, 2, \dots, n\}$) and also sequences $A_1, A_2, \dots,$ of sets ($J = \mathbb{N}$).

We give some examples of non-finite unions and intersections.

Example 3.1. For any set A we have $A = \bigcup_{a \in A} \{a\}$. According to (3.3) this also is true if $A = \emptyset$. □

The following trivial lemma is useful if you need to prove statements of the form $A \subseteq B$ or $A = B$ for sets A and B. Be sure to understand what it means if you choose $J = \{1, 2\}$ (draw one or two Venn diagrams).

Lemma 3.1 (Inclusion lemma). *Let J be an arbitrary, nonempty index set. Let U, X_j, Y, Z_j, W $(j \in J)$ be sets such that $U \subseteq X_j \subseteq Y \subseteq Z_j \subseteq W$ for all $j \in J$. Then*

$$U \subseteq \bigcap_{j \in J} X_j \subseteq Y \subseteq \bigcup_{j \in J} Z_j \subseteq W. \tag{3.4}$$

PROOF: Draw pictures! ■

Definition 3.3 (Disjoint families).

Let J be a nonempty set. We call a family of sets $(A_i)_{i \in J}$ a **mutually disjoint family** if for any two different indices $i, j \in J$ it is true that $A_i \cap A_j = \emptyset$, i.e., if any two sets in that family with different indices are mutually disjoint. □

Definition 3.4 (Partition).

Let $\mathfrak{A} \subseteq 2^\Omega$. We call $\mathfrak{A}$ a **partition** or a **partitioning** of Ω if

a. $A \cap B = \emptyset$ for any two $A, B \in \mathfrak{A}$ such that $A \neq B$, **b.** $\Omega = \biguplus \left[A : A \in \mathfrak{A} \right]$.

We reformulate the above for arbitrary families and hence finite collections and sequences of subsets of Ω: Let J be an arbitrary nonempty set, let $(A_j)_{j \in J}$ be a family of subsets of Ω. We call $(A_j)_{j \in J}$ a partition of Ω if it is a mutually disjoint family which satisfies

$$\Omega = \biguplus \left[A_j : j \in J \right],$$

in other words, if $\mathfrak{A} := \{A_j : j \in J\}$ is a partition of Ω.

Note that duplicate nonempty sets cannot occur in a disjoint family of sets because otherwise the condition of mutual disjointness does not hold. □

Example 3.2. Here are some examples of partitions.

a. For any set Ω the collection $\{ \{\omega\} : \omega \in \Omega\}$ is a partition of Ω.

b. The empty set is a partition of the empty set and it is its only partition. Do you see that this is a special case of **a**?

c. This example is important for stochastic processes. [13]

Let

$$t_0 < t_1 < \cdots < t_{n-1} < t_n$$

be a list of real numbers. It lets us create a variety of partitions. Here are four possibilities.

[13] Stochastic processes will be central to stochastic finance. See Definition 4.17 on p.74.

- $[t_0, t_1[,\ [t_1, t_2[,\ \dots,\ [t_{n-1}, t_n[$ partitions $[t_0, t_n[$,
- $]t_0, t_1],\ ]t_1, t_2],\ \dots,\ ]t_{n-1}, t_n]$ partitions $]t_0, t_n]$,
- $[t_0, t_1[,\ [t_1, t_2[,\ \dots,\ [t_{n-2}, t_{n-1}[,\ [t_{n-1}, t_n]$ partitions $[t_0, t_n]$,
- $[t_0, t_1[,\ [t_1, t_2[,\ \dots,\ [t_{n-1}, t_n[\,[t_n, \infty[$ partitions $[t_0, \infty[$. □

Theorem 3.1 (De Morgan's Law)**.** *Let there be a universal set Ω (see (2.6) on p.11). Then the following "duality principle" holds for any indexed family $(A_\alpha)_{\alpha\in I}$ of sets:*

(3.5) $$\textbf{a.}\ \Big(\bigcup_\alpha A_\alpha\Big)^{\complement} = \bigcap_\alpha A_\alpha^{\complement} \qquad \textbf{b.}\ \Big(\bigcap_\alpha A_\alpha\Big)^{\complement} = \bigcup_\alpha A_\alpha^{\complement}$$

To put this in words, the complement of an arbitrary union is the intersection of the complements, and the complement of an arbitrary intersection is the union of the complements.

PROOF: ⋆ Left as an exercise. ■

The following generalizes prop.2.6 (Distributivity of unions and intersections for finitely many sets)

Proposition 3.1 (Distributivity of unions and intersections)**.** *Let $(A_i)_{i\in I}$ be an arbitrary family of sets and let B be a set. Then*

(3.6) $$\bigcup_{i\in I}(B\cap A_i) = B\cap\bigcup_{i\in I}A_i,$$

(3.7) $$\bigcap_{i\in I}(B\cup A_i) = B\cup\bigcap_{i\in I}A_i.$$

PROOF: ■

Proposition 3.2 (Rewrite unions as disjoint unions)**.**

Let $(A_j)_{j\in\mathbb{N}}$ be a sequence of sets which all are contained within the universal set Ω. Let

$$B_n := \bigcup_{j=1}^{n} A_j = A_1\cup A_2\cup\dots\cup A_n\ (n\in\mathbb{N}),$$
$$C_1 := A_1 = B_1, \quad C_{n+1} := A_{n+1}\setminus B_n\ (n\in\mathbb{N}).$$

Then

a. *The sequence* $(B_j)_j$ *is increasing:* $m < n \Rightarrow B_m \subseteq B_n$.
b. *For each* $n \in \mathbb{N}$, $\bigcup_{j=1}^{n} A_j = \bigcup_{j=1}^{n} B_j$.
c. *The sets* C_j *are mutually disjoint and* $\bigcup_{j=1}^{n} A_j = \biguplus_{j=1}^{n} C_j$.
d. *The sets* C_j $(j \in \mathbb{N})$ *form a partitioning of the set* $\bigcup_{j=1}^{\infty} A_j$.

PROOF: ■

3.2 Direct Images and Preimages of a Function

Introduction 3.1. We continue with yet another example. It leads to the very important definition of the direct images of subsets of the domain, and of the preimages of subsets of the codomain of a function. □

Example 3.3. Let X and Y be nonempty sets and $f : X \to Y$. We define two functions $f_\star$ and $f^\star$ which are associated with f and for which both arguments and function values are sets(!) as follows.

a. $f_\star : 2^X \to 2^Y$; $\quad A \mapsto f_\star(A) := \{f(a) : a \in A\}$,
b. $f^\star : 2^Y \to 2^X$; $\quad B \mapsto f^\star(B) := \{x \in X : f(x) \in B\}$.

You should convince yourself that indeed $f_\star$ maps any subset of X to a subset of Y, and that $f^\star$ maps any subset of Y to a subset of X. □

The sets $f_\star(A)$ and $f^\star(B)$ are used pervasively in abstract mathematics, but it is much more common nowadays to write $f(A)$ rather than $f_\star(A)$ and $f^{-1}(B)$ rather than $f^\star(B)$. We will follow this convention.

Definition 3.5.

Let X, Y be two nonempty sets and $f : X \to Y$. We associate with f the functions

(3.8) $\quad f : 2^X \to 2^Y; \quad A \mapsto f(A) := \{f(a) : a \in A\},$

(3.9) $\quad f^{-1} : 2^Y \to 2^X; \quad B \mapsto f^{-1}(B) := \{x \in X : f(x) \in B\}.$

We call $f : 2^X \to 2^Y$ the **direct image function** and $f^{-1} : 2^Y \to 2^X$ the **indirect image function** or **preimage function** associated with $f : X \to Y$.
For each $A \subseteq X$ we call $f(A)$ the **direct image** of A under f, and for each $B \subseteq Y$ we call $f^{-1}(B)$ the **indirect image** or **preimage** of B under f. □

Note that the range $f(X)$ of f is a special case of a direct image.

Notational conveniences I:

If we have a set that is written as $\{\dots\}$ then we may write $f\{\dots\}$ instead of $f(\{\dots\})$ and $f^{-1}\{\dots\}$ instead of $f^{-1}(\{\dots\})$. Specifically for singletons $\{x\} \subseteq X$ and $\{y\} \subseteq Y$ we obtain $f\{x\}$ and $f^{-1}\{y\}$.
Many mathematicians will write $f^{-1}(y)$ instead of $f^{-1}\{y\}$ but this author sees no advantages doing so whatsoever. There seemingly are no savings with respect to time or space for writing that alternate form but we are confounding two entirely separate items: a subset $f^{-1}\{y\}$ of X v.s. the function value $f^{-1}(y)$ of $y \in Y$ which is an element of X. We are allowed to talk about the latter only in case that the inverse function f^{-1} of f exists.

The same symbol f is used for the original function $f : X \to Y$ and the direct image function $f : 2^X \to 2^Y$, and the symbol f^{-1} which is used here for the indirect image function $f^{-1} : 2^Y \to 2^X$ will also be used to define the inverse function $f^{-1} : Y \to X$ of f in case this can be done. Be careful not to let this confuse you! $\square$

Example 3.4 (Direct images)**.** Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$.

a. $f(]-4,2[) = \{ x^2 : x \in]-4,2[\} = \{ x^2 : -4 < x < 2 \} =]4,16[$.
b. $f([1,3]) = \{ x^2 : x \in [1,3] \} = \{ x^2 : 1 \le x \le 3 \} = [1,9]$.
c. $f(]-4,2[\cap [1,3]) = \{ x^2 : x \in]-4,2[\text{ and } x \in [1,3] \} = \{ x^2 : 1 \le x < 2 \} = [1,4[$. $\square$

And here are the results for the preimages of the same sets with respect to the same function $x \mapsto x^2$.

Example 3.5 (Preimages)**.** Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$. Determine
a. $f^{-1}(]-4,-2[)$, **b.** $f^{-1}([1,2])$, **c.** $f^{-1}([5,6])$, **d.** $\{-4 < f < -2 \textbf{ or } 1 \le f \le 2 \textbf{ or } 5 \le f < 6]\}$.

Solution:
a. $f^{-1}(]-4,-2[) = \{ x \in \mathbb{R} : x^2 \in]-4,-2[\} = \{ -4 < f < -2 \} = \emptyset$.
b. $f^{-1}([1,2]) = \{ x \in \mathbb{R} : x^2 \in [1,2] \} = \{ 1 \le f \le 2 \} = [-\sqrt{2},-1] \cup [1,\sqrt{2}]$.
c. $f^{-1}([5,6]) = \{ x \in \mathbb{R} : x^2 \in [5,6] \} = \{ 5 \le f \le 6 \} = [-\sqrt{6},-\sqrt{5}] \cup [\sqrt{5},\sqrt{6}]$.
d. $\{-4 < f < -2 \textbf{ or } 1 \le f \le 2 \textbf{ or } 5 \le f < 6]\} = f^{-1}(]-4,-2[\cup [1,2] \cup [5,6])$
$= \{ x \in \mathbb{R} : x^2 \in]-4,-2[\text{ or } x^2 \in [1,2] \textbf{ or } x^2 \in [5,6] \}$
$= [-\sqrt{2},-1] \cup [1,\sqrt{2}] \cup [-\sqrt{6},-\sqrt{5}] \cup [\sqrt{5},\sqrt{6}]$. $\square$

Example 3.6 (Preimages)**.** Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$. Determine
a. $f^{-1}(]-4,2[)$, **b.** $f^{-1}([1,3])$, **c.** $\{-4 < f < 2 \textbf{ and } 1 \le f \le 3\}$.

Solution:
a. $f^{-1}(]-4,2[) = \{ x \in \mathbb{R} : x^2 \in]-4,2[\} = \{ x \in \mathbb{R} : -4 < x^2 < 2 \} =]-\sqrt{2},\sqrt{2}[$.
b. $f^{-1}([1,3]) = \{ x \in \mathbb{R} : x^2 \in [1,3] \} = \{ x \in \mathbb{R} : 1 \le x^2 \le 3 \} = [-\sqrt{3},-1] \cup [1,\sqrt{3}]$.
c. $\{-4 < f < 2 \textbf{ and } 1 \le f \le 3\} = f^{-1}(]-4,2[\cap [1,3])$
$= \{ x \in \mathbb{R} : x^2 \in]-4,2[\textbf{ and } x^2 \in [1,3] \}$
$= \{ x \in \mathbb{R} : 1 \le x^2 < 2 \} =]-\sqrt{2},-1] \cup [1,\sqrt{2}[$. $\square$

Example 3.7 (Direct images)**.** Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$.

a. $f(]-4,-2[) \;=\; \{\, x^2 : x \in]-4,-2[\,\} \;=\; \{\, x^2 : -4 < x < -2 \,\} \;=]4,16[.$
b. $f([1,2]) \;=\; \{\, x^2 : x \in [1,2] \,\} \;=\; \{\, x^2 : 1 \le x \le 2 \,\} \;=\; [1,4].$
c. $f([5,6]) \;=\; \{\, x^2 : x \in [5,6] \,\} \;=\; \{\, x^2 : 5 \le x \le 6 \,\} \;=\; [25,36].$
d. $f(]-4,-2[\cup [1,2] \;\cup\; [5,6]) \;=\; \{\, x^2 : x \in]-4,-2[$ or $x \in [1,2]$ or $x \in [5,6] \,\}$
$=]4,16[\cup [1,4] \;\cup\; [25,36] \;=\; [1,16[\cup [25,36].$ $\square$

Proposition 3.3. *Some simple properties:*

(3.10) $f(\emptyset) \;=\; f^{-1}(\emptyset) \;=\; \emptyset$

(3.11) $A_1 \subseteq A_2 \subseteq X \Rightarrow f(A_1) \subseteq f(A_2)$ (***monotonicity of*** $f\{\dots\}$)

(3.12) $B_1 \subseteq B_2 \subseteq Y \Rightarrow f^{-1}(B_1) \subseteq f^{-1}(B_2)$ (***monotonicity of*** $f^{-1}\{\dots\}$)

(3.13) $x \in X \Rightarrow f(\{x\}) \;=\; \{f(x)\}$

(3.14) $f(X) \;=\; Y \;\Leftrightarrow\; f$ *is "surjective" (see Remark 2.11 on p.24*

(3.15) $f^{-1}(Y) \;=\; X$ *always!*

PROOF: Left as exercise **??** on p.**??**. ■

Notational conveniences II:

In measure theory and probability theory the following notation is also very common:
$\{f \in B\} := f^{-1}(B),\ \{f = y\} := f^{-1}\{y\}.$
Let R be an ordered integral domain with associated order relation "$<$". Let $a, b \in R$ such that $a < b$. We write $\{a \le f \le b\} := f^{-1}([a,b]_R),\ \{a < f < b\} := f^{-1}(]a,b[_R),$
$\{a \le f < b\} := f^{-1}([a,b[_R),\ \{a < f \le b\} := f^{-1}(]a,b]_R),\ \{f \le b\} := f^{-1}(]-\infty,b]_R)$, etc.

Proposition 3.4 (f^{-1} is compatible with all basic set ops)**.** *Assume that X, Y be nonempty, $f : X \to Y$, J is an arbitrary index set, $B \subseteq Y$, $B_j \subseteq Y$ for all j. Then*

(3.16) $$f^{-1}\Big(\bigcap_{j \in J} B_j\Big) \;=\; \bigcap_{j \in J} f^{-1}(B_j)$$

(3.17) $$f^{-1}\Big(\bigcup_{j \in J} B_j\Big) \;=\; \bigcup_{j \in J} f^{-1}(B_j)$$

(3.18) $$f^{-1}(B^{\complement}) \;=\; \big(f^{-1}(B)\big)^{\complement}$$

(3.19) $$f^{-1}(B_1 \setminus B_2) \;=\; f^{-1}(B_1) \setminus f^{-1}(B_2)$$

(3.20) $$f^{-1}(B_1 \Delta B_2) \;=\; f^{-1}(B_1) \Delta f^{-1}(B_2)$$

PROOF: ⋆ MF330 notes, ch.8 ■

Proposition 3.5 (Properties of the direct image). *Assume that X, Y be nonempty, $f : X \to Y$, J is an arbitrary index set, $B \subseteq Y$, $B_j \subseteq Y$ for all j. Then*

$$f\Big(\bigcap_{j \in J} A_j\Big) \subseteq \bigcap_{j \in J} f(A_j) \tag{3.21}$$

$$f\Big(\bigcup_{j \in J} A_j\Big) = \bigcup_{j \in J} f(A_j) \tag{3.22}$$

PROOF: ⋆ MF330 notes, ch.8 ■

Remark 3.1. In general you will not have equality in (3.21). Counterexample: $f(x) = x^2$ with domain $\mathbb{R}$: Let $A_1 :=]-\infty, 0]$ and $A_2; = [0, \infty[$. Then $A_1 \cap A_2 = \{0\}$, hence $f(A_1 \cap A_2) = f(\{0\}) = \{0\}$. On the other hand, $f(A_1) = f(A_2) = [0, \infty]$, hence $f(A_1) \cap f(A_2) = [0, \infty]$. Clearly, $\{0\} \subsetneq [0, \infty]$. □

Proposition 3.6 (Direct images and preimages of function composition). *Let X, Y, Z be arbitrary, nonempty sets. Let $f : X \to Y$ and $g : Y \to Z$, and let $U \subseteq X$ and $W \subseteq Z$. Then*

$$(g \circ f)(U) = g\big(f(U)\big) \text{ for all } U \subseteq X. \tag{3.23}$$

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}, \text{ i.e., } (g \circ f)^{-1}(W) = f^{-1}\big(g^{-1}(W)\big) \text{ for all } W \subseteq Z. \tag{3.24}$$

PROOF: ⋆ MF330 notes, ch.8 ■

3.3 Indicator Functions

Indicator functions often are convenient when working with integrals and expected values. They will allow us, e.g., to write "$E[\mathbf{1}_A X] = \dots$" rather than having to state all of this: "Let $Y(\omega) := X(\omega$ on A and 0 else. Then $E[Y] = \dots$"

Definition 3.6 (indicator function for a set).

Let Ω be a nonempty set and $A \subseteq \Omega$. Let $\mathbf{1}_A : \Omega \to \{0, 1\}$ be the function defined as

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases} \tag{3.25}$$

$\mathbf{1}_A$ is called the **indicator function** of the set A.

Some authors call $\mathbf{1}_A$ the **characteristic function** of A and some choose to write χ_A or $\mathbb{1}_A$ instead of $\mathbf{1}_A$. □

Let $m, n \in \mathbb{Z}$. We recall that $m+n \mod 2$ (the sum mod 2 of m and n) is given by

$$m+n \mod 2 \;=\; \begin{cases} 0 & \Leftrightarrow \quad (m+n)/2 \text{ has remainder } 0 \text{, i.e., } m+n \text{ is even,} \\ 1 & \Leftrightarrow \quad (m+n)/2 \text{ has remainder } 1 \text{, i.e., } m+n \text{ is odd.} \end{cases} \tag{3.26}$$

Proposition 3.7.

Let A, B, C be subsets of Ω. Then

$$\mathbf{1}_{A \cup B} \;=\; \max(\mathbf{1}_A, \mathbf{1}_B), \tag{3.27}$$

$$\mathbf{1}_{A \cap B} \;=\; \min(\mathbf{1}_A, \mathbf{1}_B), \tag{3.28}$$

$$\mathbf{1}_{A^\complement} \;= 1 - \mathbf{1}_A, \tag{3.29}$$

$$\mathbf{1}_{A \triangle B} \;= \mathbf{1}_A + \mathbf{1}_B \mod 2. \tag{3.30}$$

PROOF: The proof of the first three equations is left as an exercise.
PROOF of (3.30): This follows easily from the the fact that

$$(A \triangle B)^\complement \;=\; \{\omega \in \Omega : \,[\text{either } \omega \in A \cap B] \text{ or } [\text{neither } \omega \in A \text{ nor } \omega \in B\,]\} \quad \blacksquare$$

Prop.**??** above helps us to prove associativity of symmetric set differences.

Proposition 3.8 (Symmetric set differences $A \triangle B$ are associative)**.**

Let $A, B, C \subseteq \Omega$. Then

$$(A \triangle B) \triangle C \;=\; A \triangle (B \triangle C). \tag{3.31}$$

PROOF: We will write for convenience $m \oplus n$ as shorthand notation for $m+n \mod 2$.
Formula (3.31) follows easily from (3.30) and and the associativity of $a \oplus b := a+b \mod 2$ as follows. Let $\omega \in \Omega$. Then

$$\begin{aligned} \omega \in (A \triangle B) \triangle C \;&\Leftrightarrow \mathbf{1}_{(A \triangle B) \triangle C}(\omega) = 1 \\ &\Leftrightarrow \big(\mathbf{1}_A(\omega) \oplus \mathbf{1}_B(\omega)\big) \oplus \mathbf{1}_C(\omega) = 1 \\ &\Leftrightarrow \mathbf{1}_A(\omega) \oplus \big(\mathbf{1}_B(\omega) \oplus \mathbf{1}_C(\omega)\big) = 1 \\ &\Leftrightarrow \mathbf{1}_{A \triangle (B \triangle C)}(\omega) = 1 \;\Leftrightarrow\; \omega \in A \triangle (B \triangle C). \end{aligned}$$

We obtained the equivalence in the middle from the fact that modular arithmetic is associative. $\blacksquare$

4 Basic Measure and Probability Theory

Introduction:
The following are the best known examples of measures ($a_j, b_j \in \mathbb{R}$):

$$\begin{aligned}
&\text{Length}: \quad \lambda^1([a_1, b_1]) := b_1 - a_1,\\
&\text{Area}: \quad \lambda^2([a_1, b_1] \times [a_2, b_2]) := (b_1 - a_1)(b_2 - a_2),\\
&\text{Volume}: \quad \lambda^3([a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]) := (b_1 - a_1)(b_2 - a_2)(b_3 - a_3).
\end{aligned}$$

Then there also are probability measures: $P\{$ a die shows a 1 or a $6\} = 1/3$.
We will explore in this chapter some of the basic properties of measures.

4.1 Measure Spaces and Probability Spaces

Notation 4.1. By augmenting certain sets of real numbers with $\pm\infty$ we obtain the sets

$$\begin{aligned}
\overline{\mathbb{R}} &:= [-\infty, \infty] := \mathbb{R} \cup \{-\infty, \infty\} \qquad \text{(extended real numbers)}\,,\\
\overline{\mathbb{R}}_+ &:= [\,0, \infty\,] := \mathbb{R}_+ \cup \{+\infty\} \qquad (4.1)\\
[\,a, \infty\,] &:= [\,a, \infty[\, \cup \{+\infty\} \qquad (\text{here } -\infty \leq a < \infty) \ \square
\end{aligned}$$

Definition 4.1 (Extended real–valued functions)**.**

Let X be an arbitrary, nonempty set and $Y \subseteq \overline{\mathbb{R}}$. A function $F : X \to Y$ whose codomain is a subset of the extended real numbers is called an **extended real–valued function**. $\square$

Remark 4.1 (Extended real numbers arithmetic)**.** To work with extended real–valued functions we must be clear about the rules of arithmetic where $\pm\infty$ is involved. In the following assume that $c \in \mathbb{R}$ and $0 < p < \infty$.

Rules for Addition:

$$c \pm \infty = \infty \pm c = \infty, \qquad (4.2)$$
$$c \pm (-\infty) = -\infty \pm c = -\infty, \qquad (4.3)$$
$$\infty + \infty = \infty, \qquad (4.4)$$
$$-\infty - \infty = -\infty, \qquad (4.5)$$
$$(\pm\infty) \mp \infty = \textbf{UNDEFINED}. \qquad (4.6)$$

Rules for Multiplication:

$$
\begin{aligned}
&(4.7) && p\cdot(\pm\infty) = (\pm\infty)\cdot p = \pm\infty,\\
&(4.8) && (-p)\cdot(\pm\infty) = (\pm\infty)\cdot(-p) = \mp\infty,\\
&(4.9) && 0\cdot(\pm\infty) = (\pm\infty)\cdot 0 = 0 \quad\text{and}\quad \frac{1}{\infty} = 0,\\
&(4.10) && (\pm\infty)\cdot(\pm\infty) = \infty,\\
&(4.11) && (\pm\infty)\cdot(\mp\infty) = -\infty,
\end{aligned}
$$

Be clear about the ramifications of those rules. Rule (4.6) implies that if we have two extended real–valued functions f, g defined on a domain A then $f + g$ is only defined on

$$A \setminus \{x \in A : \text{ either } [f(x) = \infty \text{ and } g(x) = -\infty] \text{ or } [f(x) = -\infty \text{ and } g(x) = \infty]\},$$

and $f - g$ is only defined on

$$A \setminus \{x \in A : \text{ either } [f(x) = g(x) = \infty] \text{ or } [f(x) = g(x) = -\infty]\}.$$

That is easy to understand and remember, but the real danger comes from rule (4.9) which you might not have expected:

$$0 \cdot \pm\infty = \pm\infty \cdot 0 = 0.$$

This convention is very convenient, but it comes at a price: it is no longer true that all sequences $(a_n)_n$ and $(b_n)_n$ of real numbers that have limits $a = \lim\limits_{n\to\infty} a_n$, $b = \lim\limits_{n\to\infty} b_n$, satisfy $\lim\limits_{n\to\infty} a_n b_n = ab$. Such a counterexample would be: $a_n = n, b_n = \frac{1}{n}$. □

We give some convenient definitions and notations for monotone sequences of numbers, functions and sets.

Definition 4.2.

(a) Let x_n be a sequence of extended real–valued numbers.
- We call x_n a **nondecreasing** or **increasing** sequence, if $j < n \Rightarrow x_j \leq x_n$.
- We call x_n a **strictly increasing** sequence, if $j < n \Rightarrow x_j < x_n$.
- We call x_n a **nonincreasing** or **decreasing** sequence, if $j < n \Rightarrow x_j \geq x_n$.
- We call x_n a **strictly decreasing** sequence, if $j < n \Rightarrow x_j > x_n$.
- We write $x_n \uparrow$ for nondecreasing x_n, and $x_n \uparrow x$ to indicate that $\sup_n x_n = x$,
- We write $x_n \downarrow$ for nonincreasing x_n, $x_n \downarrow x$ to indicate that $\inf_n x_n = x$.

(b) Let $X \neq \emptyset$ and $f_n : X \to \bar{R}$ a sequence of extended real–valued functions. We call f_n a **nondecreasing** or **increasing function sequence**

$$\text{and we write } f_n \uparrow, \text{ if } \quad j < n \Rightarrow f_j(x) \leq f_n(x) \text{ for all } x \in X.$$

We call f_n a **nonincreasing** or **decreasing function sequence**

$$\text{and we write } f_n \downarrow, \text{ if } \quad j < n \Rightarrow f_j(x) \geq f_n(x) \text{ for all } x \in X.$$

Strictly increasing and **strictly decreasing function sequences** are defined by replacing $\leq$ with $<$ and $\geq$ with $>$ in those last definitions.

(c) Let $X \neq \emptyset$ and $A_n \subseteq X$ a sequence of subsets of X We call A_n a **nondecreasing** (resp. **strictly increasing** resp. ....) **sequence of sets**, if the corresponding sequence $\mathbf{1}_{A_n}$ of indicator functions is a nondecreasing (resp. strictly increasing resp. ....) function sequence. We write $A_n \uparrow$ if A_n is nondecreasing and $A_n \downarrow$ if A_n is nonincreasing. □

Remark 4.2. (A) In Definition 4.2, we made no assumptions about the domain X of the functions f_n besides not being empty. In particular, X can be the power set 2^Ω of some arbitrary set Ω. Then a sequence of functions

$$\mu_n : 2^\Omega \to \bar{\mathbb{R}}; \qquad A \mapsto \mu_n(A)$$

would take subsets of Ω as arguments and map them to real numbers. You are familiar with the following example: Probabilities are functions which assign numbers to events, i.e., sets.

(B) You should convince yourself of the following. If X is a nonempty set and $A_n \in X$, then

(4.12) $$A_n \uparrow \Leftrightarrow A_1 \subseteq A_2 \subseteq \dots; \qquad A_n \text{ is strictly increasing } \Leftrightarrow A_1 \subsetneq A_2 \subsetneq \dots;$$

(4.13) $$A_n \downarrow \Leftrightarrow A_1 \supseteq A_2 \supseteq \dots; \qquad A_n \text{ is strictly decreasing } \Leftrightarrow A_1 \supsetneq A_2 \supsetneq \dots.$$

(4.14) $$A_n \uparrow \Rightarrow \mathbf{1}_{A_n} \uparrow \mathbf{1}_{\bigcup_j A_j}, \qquad A_n \downarrow \Rightarrow \mathbf{1}_{A_n} \downarrow \mathbf{1}_{\bigcap_j A_j},$$

(C) Also note in Definition 4.2**(a)** that

(4.15) $$x_n \uparrow \Rightarrow \sup_{n \in \mathbb{N}} x_n = \lim_{n \to \infty} x_n, \quad \text{i.e., } x_n \uparrow \lim_{j \to \infty} x_j;$$

(4.16) $$x_n \downarrow \Rightarrow \inf_{n \in \mathbb{N}} x_n = \lim_{n \to \infty} x_n, \quad \text{i.e., } x_n \downarrow \lim_{j \to \infty} x_j.$$

Thus, if for $f_n, f : X \to \bar{\mathbb{N}}$ we define f to be the **(pointwise) limit** of the functions f_n, i.e.,

$$f := \lim_{n \to \infty} f_n \quad \Leftrightarrow \quad f(x) = \lim_{n \to \infty} f_n(x) \text{ for all } x \in X,$$

then we obtain from (4.15) and (4.16) the following.

(4.17) $$f_n \uparrow \Rightarrow \sup_{n \in \mathbb{N}} f_n(x) = \lim_{n \to \infty} f_n(x) \text{ for all } x \in X, \quad \text{i.e., } f_n \uparrow \lim_{j \to \infty} f_j;$$

(4.18) $$f_n \downarrow \Rightarrow \inf_{n \in \mathbb{N}} f_n(x) = \lim_{n \to \infty} f_n(x) \text{ for all } x \in X, \quad \text{i.e., } f_n \downarrow \lim_{j \to \infty} f_j.$$

Finally, note the following for $X \neq \emptyset$ and $A_n \subseteq X$.

$$A_n \uparrow \overset{(4.14)}{\Rightarrow} \mathbf{1}_{A_n} \uparrow \mathbf{1}_{\bigcup_j A_j}, \overset{(4.17)}{\Rightarrow} \mathbf{1}_{\bigcup_j A_j} = \lim_{j\to\infty} \mathbf{1}_{A_j}\,, \tag{4.19}$$

$$A_n \downarrow \overset{(4.14)}{\Rightarrow} \mathbf{1}_{A_n} \uparrow \mathbf{1}_{\bigcap_j A_j}, \overset{(4.18)}{\Rightarrow} \mathbf{1}_{\bigcup_j A_j} = \lim_{j\to\infty} \mathbf{1}_{A_j}\,. \tag{4.20}$$

It thus makes sense to speak of limits of sequences of sets in those two cases: [14]

$$A_n \uparrow \Rightarrow \bigcup_j A_j = \lim_{j\to\infty} A_j\,, \qquad \text{and} \qquad A_n \downarrow \Rightarrow \bigcap_j A_j = \lim_{j\to\infty} A_j\,. \ \square$$

For the following see SCF2 Definition 1.1.1.

Definition 4.3 (σ–algebras)**.**

Let $\Omega \neq \emptyset$. let $\mathfrak{F}$ be a set that contains some, but not necessarily all, subsets of Ω.
$\mathfrak{F}$ is called a **σ–algebra** or **σ–field** for or on Ω if it satisfies the following:

$$\emptyset \in \mathfrak{F}, \tag{4.21a}$$

$$A \in \mathfrak{F} \quad \Rightarrow \quad A^{\complement} \in \mathfrak{F} \tag{4.21b}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \quad \Rightarrow \quad \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F} \tag{4.21c}$$

- The pair $(\Omega, \mathfrak{F})$ is called a **measurable space**.
- The elements of $\mathfrak{F}$ (these elements are sets!) are called **$\mathfrak{F}$–measurable sets**. or also just **measurable sets** if it is clear what σ–algebra is referred to. $\square$

We do not consider $\Omega = \emptyset$ with σ–algebra $\{\emptyset\}$ a measurable space since it cannot carry a probability P which would have to satisfy $P(\emptyset) = 0$ and $P(\Omega) = 1$. See Chapter 4.2 (Measurable Functions and Random Elements).

Remark 4.3. If $\mathfrak{F}$ is a σ–algebra then

$$\Omega \in \mathfrak{F} \tag{4.22a}$$

$$A \in \mathfrak{F} \quad \Rightarrow \quad A^{\complement} \in \mathfrak{F} \tag{4.22b}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \quad \Rightarrow \quad \bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{F} \tag{4.22c}$$

The last assertion is a consequence of De Morgan's laws (Theorem 3.1 on p.40).

[14] and to make the following general definition: If $B, B_n \subseteq X$, we say that

$$B = \lim_{n\to\infty} B_n \quad \Leftrightarrow \quad \mathbf{1}_B = \lim_{n\to\infty} \mathbf{1}_{B_n}\,.$$

If countably many (i.e., a finite or infinite sequence of) operations are performed involving
• unions, • intersections, • complements, • set differences, • symmetric differences
of elements of a σ–algebra $\mathfrak{F}$ then the resulting set also belongs to $\mathfrak{F}$. □

Example 4.1. Here are two trivial σ–algebras of a nonempty set Ω.

(1) $\{\emptyset, \Omega\}$ is the smallest possible σ–algebra.
(2) The power set 2^Ω of Ω is the largest possible σ–algebra. □

Proposition 4.1 (Minimal sigma–algebras)**.** *Let Ω be a nonempty set.*

A: *The intersection of arbitrarily many σ–algebras is a σ–algebra.*

B: *Let $\mathfrak{E} \subseteq 2^\Omega$, i.e., $\mathfrak{E}$ is a set which contains subsets of Ω. It is not assumed that $\mathfrak{E}$ is a σ–algebra. Then there exists a σ–algebra which contains $\mathfrak{E}$ and is minimal in the sense that it is contained in any other σ–algebra that also contains $\mathfrak{E}$. We name this σ–algebra $\sigma(\mathfrak{E})$ because it clearly is uniquely determined by $\mathfrak{E}$. It is constructed as follows:*

$$\sigma(\mathfrak{E}) \;=\; \bigcap \{\mathfrak{F} : \mathfrak{F} \supseteq \mathfrak{E} \text{ and } \mathfrak{F} \text{ is a } \sigma\text{–algebra for } \Omega\}.$$

PROOF: ⋆ ■

That last proposition allows us to make the next definition.

Definition 4.4.

Let Ω be a nonempty set and let $\mathfrak{E} \subseteq 2^\Omega$. We call the σ–algebra

$$\sigma(\mathfrak{E}) \;=\; \bigcap \{\mathfrak{G} : \mathfrak{G} \supseteq \mathfrak{E} \text{ and } \mathfrak{G} \text{ is a } \sigma\text{–algebra for } \Omega\}. \tag{4.23}$$

of Proposition 4.1 the **σ–Algebra generated by $\mathfrak{E}$** □

Remark 4.4.

(1) You are familiar with this construct from linear algebra:
Given a subset A of a vector space V, its linear span

$$span(A) \;=\; \{\, \sum_{j=1}^{k} \alpha_j x_j : k \in \mathbb{N}, \alpha_j \in \mathbb{R}, x_j \in A \; (1 \leq j \leq k) \,\}.$$

of all linear combinations of vectors in A is obtained as follow:

$$\text{Let } \mathfrak{V} := \{W \subseteq V : W \supseteq A \text{ and } W \text{ is a subspace of } V\}.$$
$$\text{Then } span(A) = \bigcap \left[W : W \in \mathfrak{V}\right].$$

In other words, $span(A)$ is the (linear) subspace generated by A.

(2) Note that if $\mathfrak{E} \subseteq \mathfrak{F}$ then $\sigma(\mathfrak{E}) \subseteq \mathfrak{F}$, since $\mathfrak{F}$ is one of the σ–algebras $\mathfrak{G}$ which occur on the right–hand side of (4.23). □

You should visualize the next proposition for the case of one, two, three, and four events A_j.

Proposition 4.2. ⋆

Let $(\Omega, \mathfrak{F})$ be a measurable space in which a finite or infinite sequence of events $A_1, A_2, \dots$ is a partition of Ω and generates $\mathfrak{F}$. Let $J := \{1, 2, \dots, n\}$ in case of a finite sequence $A_j : 1 \le j \le n$, and let $J := \mathbb{N}$ in case of a sequence $A_j : j \in \mathbb{N}$. Then our assumptions can be stated as follows.

$$A_i \cap A_j = \emptyset \text{ for } i \neq j, \quad \biguplus_{j \in J} A_j = \Omega, \quad \mathfrak{F} = \sigma\{A_j : j \in J\}. \tag{4.24}$$

Under those assumptions it is true that $\mathfrak{F}$ consists of all countable unions $A_{n_1} \uplus A_{n_2} \uplus \dots$.

PROOF: Left as an exercise.

Hint: What is the complement of the union $A_{n_1} \uplus A_{n_2} \uplus \dots$? ■

Proposition 4.3 (Monotonicity of generated σ–algebras)**.** *Let Ω be a nonempty set and let $\mathfrak{E}_1$ and $\mathfrak{E}_2$ be two collections of subsets of Ω.*

$$\text{If} \quad \mathfrak{E}_1 \subseteq \mathfrak{E}_2 \quad \text{then} \quad \sigma(\mathfrak{E}_1) \subseteq \sigma(\mathfrak{E}_2)\,. \tag{4.25}$$

PROOF: Any σ–algebra $\mathfrak{G}$ that contains $\mathfrak{E}_2$ also contains $\mathfrak{E}_1$ Thus more sets are intersected in

$$\sigma(\mathfrak{E}_1) = \bigcap\{\mathfrak{G} : \mathfrak{G} \supseteq \mathfrak{E}_1 \text{ and } \mathfrak{G} \text{ is a } \sigma\text{–algebra for } \Omega\}.$$

than in

$$\sigma(\mathfrak{E}_2) = \bigcap\{\mathfrak{G} : \mathfrak{G} \supseteq \mathfrak{E}_2 \text{ and } \mathfrak{G} \text{ is a } \sigma\text{–algebra for } \Omega\}.$$

It follows that $\sigma(\mathfrak{E}_1) \subseteq \sigma(\mathfrak{E}_2)$. ■

Proposition 4.4.

Let Ω be a nonempty set. Assume $\mathfrak{E}_1, \mathfrak{E}_2$ are subsets of 2^Ω such that

$$\sigma(\mathfrak{E}_1) \supseteq \mathfrak{E}_2 \quad \text{and} \quad \sigma(\mathfrak{E}_2) \supseteq \mathfrak{E}_1.$$

Then $\sigma(\mathfrak{E}_1) = \sigma(\mathfrak{E}_2)$.

PROOF: ⋆ Left as an exercise. ■

Example 4.2. Consider the following sets of intervals of real numbers.

$$\mathfrak{J}_1 := \{]a,b] : a < b\}, \quad \mathfrak{J}_2 := \{[a,b] : a < b\},$$
$$\mathfrak{J}_3 := \{]a,b[: a < b\}, \quad \mathfrak{J}_4 := \{[a,b[: a < b\}.$$

Then $\sigma(\mathfrak{I}_1) = \sigma(\mathfrak{I}_2) = \sigma(\mathfrak{I}_3) = \sigma(\mathfrak{I}_4)$.
For example, to prove that $\mathfrak{I}_2 = \mathfrak{I}_3$, it suffices according to Proposition 4.4 to show that

any closed interval $[a,b]$ belongs to $\mathfrak{I}_3$, any open interval $]a,b[$ belongs to $\mathfrak{I}_2$.

This follows from

$$[a,b] = \bigcap_n \left]a - \frac{1}{n}, b + \frac{1}{n}\right[\quad \text{and} \quad]a,b[= \bigcup_n \left[a + \frac{1}{n}, b - \frac{1}{n}\right].$$

The above generalizes to n–dimensional space: Let

$\mathfrak{I}_5 := \{]a_1,b_1]\times]a_2,b_2] \times \cdots \times]a_n,b_n] : a_1 < b_1, a_2 < b_2, \dots, a_n < b_n\}$,
$\mathfrak{I}_6 := \{[a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_n,b_n] : a_1 < b_1, a_2 < b_2, \dots, a_n < b_n\}$,
$\mathfrak{I}_7 := \{]a_1,b_1[\times]a_2,b_2[\times \cdots \times]a_n,b_n[: a_1 < b_1, a_2 < b_2, \dots, a_n < b_n\}$,
$\mathfrak{I}_8 := \{[a_1,b_1[\times[a_2,b_2[\times \cdots \times [a_n,b_n[: a_1 < b_1, a_2 < b_2, \dots, a_n < b_n\}$,

Then $\sigma(\mathfrak{I}_5) = \sigma(\mathfrak{I}_6) = \sigma(\mathfrak{I}_7) = \sigma(\mathfrak{I}_8)$. □

For the following see SCF2 Definition 1.1.2.

Definition 4.5 (Borel sets)**.**

- The σ–algebra generated by either all open or all closed or all half–open intervals in $\mathbb{R}^n$ is called the **Borel σ–algebra** of subsets of $\mathbb{R}^n$ and is denoted $\mathfrak{B}(\mathbb{R}^n)$.
- The sets in this σ–algebra are called **Borel sets**.
- Abbreviations: We also write $\mathfrak{B}^n$ for $\mathfrak{B}(\mathbb{R}^n)$. In the case of the real numbers ($n = 1$) we also write $\mathfrak{B}^1$ or $\mathfrak{B}(\mathbb{R})$ for $\mathfrak{B}(\mathbb{R}^1)$. □

- We do not consider what corresponds to the Borel sets when we deal with the extended real numbers $\bar{\mathbb{R}}$, i.e., we add $\pm\infty$. Such "extended Borel sets", denoted by $\mathfrak{B}(\bar{\mathbb{R}})$, can be defined. Again, extended Borel sets will not be dealt with in this course. □

Remark 4.5. We can express Example 4.2 as follows. Each one of the interval sets $\mathfrak{I}_5, \mathfrak{I}_6, \mathfrak{I}_7, \mathfrak{I}_8$ generates the Borel σ–algebra. □

For the following see SCF2 Definition 1.1.2.

Definition 4.6 (Abstract measures)**.** Let $(\Omega, \mathfrak{F})$ be a measurable space.

A **measure** on $\mathfrak{F}$ is an extended real–valued function

$$\mu : \mathfrak{F} \to \overline{\mathbb{R}}_+; \quad A \mapsto \mu(A) \qquad \text{such that}$$

$$\mu(\emptyset) \;=\; 0\,, \qquad \text{(positivity)} \tag{4.26}$$

$$A, B \in \mathfrak{F} \text{ and } A \subseteq B \quad \Rightarrow \quad \mu(A) \leq \mu(B)\,, \qquad \text{(monotony)} \tag{4.27}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \text{ disjoint} \quad \Rightarrow \quad \mu\Big(\biguplus_{n\in\mathbb{N}} A_n\Big) \;=\; \sum_{n\in\mathbb{N}} \mu(A_n)\,. \qquad (\sigma\text{–additivity}) \tag{4.28}$$

- The triplet $(\Omega, \mathfrak{F}, \mu)$ is called a **measure space**
- We call μ a **finite measure** on $\mathfrak{F}$ if $\mu(\Omega) < \infty$.
- We call any subset N of a set with measure zero a μ**–null set**. Note that N need not be measurable.
- If $\mu(\Omega) = 1$ then μ is called a **probability measure** or simply a **probability** and $(\Omega, \mathfrak{F}, \mu)$ is then called a **probability space**. □

Disjointness in (4.28) means that $A_i \cap A_j = 0$ for any $i, j \in \mathbb{N}$ such that $i \neq j$ (see def.2.4 on p.10).

Do not confuse measurable spaces $(\Omega, \mathfrak{F})$ and measure spaces $(\Omega, \mathfrak{F}, \mu)$!

Remark 4.6 (σ–algebras are appropriate domains for measures)**.** The σ–additivity of measures is what makes working with them such a pleasure in many ways. It can be stated as follows:

For a disjoint sequence of measurable sets the measure of its disjoint union is the sum of the measures. Property (4.21c) in the definition of σ–algebras is required for exactly that reason.

you cannot take advantage of the σ–additivity of a measure μ if its domain does not contain countable unions and intersections of all its constituents.

Here are two not very useful measures which are easy to understand.

Example 4.3. You can easily verify that the following set functions μ_1 and μ_2 define measures on an arbitrary nonempty set Ω with an arbitrary σ-field $\mathfrak{F}$.

$$\mu_1(A) \;:= 0 \;\text{ for all } A \in \mathfrak{F}, \qquad \textbf{zero measure or null measure}$$
$$\mu_2(\emptyset) \;:= 0; \qquad \mu(A) \;:=\; \infty \;\text{ if } A \neq \emptyset.$$

Keep the second example in mind when you work with non–finite measures. □

Remark 4.7.

(1) We emphasize that the only difference between (general) measures and probability measures is that the latter must assign a measure of one to the entire space Ω.

(2) Many things that apply to probabilities can be extended to general measures, and this will matter to us even if we are only interested in probability spaces, since will see in the context of the expectation $E[X]$ of a random variable X that assignments of the form

$$A \mapsto E[X \cdot \mathbf{1}_A] \text{ where } A \in \mathfrak{F} \text{ and } \mathbf{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

define a measure on $(\Omega, \mathfrak{F})$.

(3) Traditionally, mathematicians write $P(A)$ and $(\Omega, \mathfrak{F}, P)$ rather than $\mu(A)$ and $(\Omega, \mathfrak{F}, \mu)$ for probability measures and probability spaces. The elements of $\mathfrak{F}$ (the measurable subsets) are then thought of as **events** for which $P(A)$ is interpreted as the probability with which the event A might happen.

(4) A measure space can support many different measures: If μ is a measure on $\mathfrak{F}$ and $\alpha \geq 0$ then $\alpha\mu : A \mapsto \alpha\mu(A)$ also is a measure on $\mathfrak{F}$. $\square$

Fact 4.1. *Assume that the real–valued function*

$$\mu_0 : \mathfrak{I}_5 \longrightarrow \mathbb{R}, \quad B \mapsto \mu_0(B),$$

is defined on the set of half–open n–dimensional intervals

$$\mathfrak{I}_5 = \{]a_1, b_1] \times]a_2, b_2] \times \cdots \times]a_n, b_n] : a_1 < b_1, a_2 < b_2, \ldots, a_n < b_n\}$$

of Example 4.2 on p.51 and satisfies the measure defining properties of positivity, monotony, and σ–additivity. Then μ_0 can be uniquely extended to a measure μ on the measurable space $\big(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n)\big)$

In other words, there exists a uniquely defined measure μ on the Borel sets $\mathfrak{B}(\mathbb{R}^n)$ (see Definition 4.5 (Borel sets) on p.52) such that

$$\mu(]a_1, b_1] \times]a_2, b_2] \times \cdots \times]a_n, b_n]) = \mu_0(]a_1, b_1] \times]a_2, b_2] \times \cdots \times]a_n, b_n])$$

for any half–open interval $]a_1, b_1] \times]a_2, b_2] \times \cdots \times]a_n, b_n]$, $a_1 < b_1, a_2 < b_2, \ldots, a_n < b_n$. $\square$

For the following see SCF2 Example 1.1.3 - Uniform (Lebesgue) measure on [0, 1]

The most important measures we encounter in real life are those that measure the length of sets in one dimension, the area of sets in two dimensions and the volume of sets in three dimensions.

Definition 4.7 (Lebesgue measure)**.** For

- intervals $[a, b] \in \mathbb{R}$,
- rectangles $[a_1, b_1] \times [a_2, b_2] \in \mathbb{R}^2$,
- boxes or quads $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \in \mathbb{R}^3$,
- in general, n**-dimensional parallelepipeds** $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \in \mathbb{R}^n$,

we define

$$\begin{aligned} \lambda_0^1(]a, b]) &:= b - a, \\ \lambda_0^2(]a_1, b_1] \times]a_2, b_2]) &:= (b_1 - a_1)(b_2 - a_2), \\ \lambda_0^3(]a_1, b_1] \times]a_2, b_2] \times]a_3, b_3]) &:= (b_1 - a_1)(b_2 - a_2)(b_3 - a_3), \\ \lambda_0^n(]a_1, b_1] \times \cdots \times]a_n, b_n]) &:= (b_1 - a_1)(b_2 - a_2) \ldots (b_n - a_n). \end{aligned} \tag{4.29}$$

One can show that those set functions satisfies the conditions stated in Fact 4.1. [15] Thus λ_0^n uniquely extends from the parallelepipeds to a measure λ^n on the Borel sets of $\mathbb{R}^n$.

The measure λ_0^n is called (n-dimensional) **Lebesgue measure**.

Note that Lebesgue measure is not finite: $\lambda^n(\mathbb{R}^n) = \infty$! $\square$

Fact 4.2. *It is not possible to extend the set functions μ_0^n which define Lebesgue measure to a measure on the entire power set $2^{\mathbb{R}^n}$ of $\mathbb{R}^n$.*

This (very hard to prove) fact makes it a mathematical necessity to introduce σ–algebras as small enough subsets of the power set 2^Ω which are suitable as domains for a measure.

We will see later that σ–algebras also have a practical importance: they reflect the information that is associated with certain random phenomena, for example, the evolution of the price of a financial asset. $\square$

Remark 4.8 (Finite disjoint unions)**.** If we have only finitely many sets then "σ–additivity" which stands for "additivity of countably many" becomes simple additivity. We obtain the following by setting $A_{N+1} = A_{N+2} = \ldots = 0$:

(4.30) $$\begin{aligned}&A_1, A_2, \ldots, A_N \in \mathfrak{F} \text{ mutually disjoint}\\ &\quad\Rightarrow\ \mu(A_1 \uplus A_2 \uplus \ldots \uplus A_N) \ = \ \mu(A_1) + \mu(A_2) + \ldots + \mu(A_N) \quad \text{(additivity)}.\end{aligned}$$

In the case of only two disjoint measurable sets A and B the above simply becomes

$$\mu(A \uplus B) \ = \ \mu(A) + \mu(B). \ \square$$

Proposition 4.5 (Simple properties of measures)**.** *Let $A, B, \in \mathfrak{F}$ and let μ be a measure on $\mathfrak{F}$. Then*

(4.31a) $$\mu(A) \geq 0 \quad \textit{for all } A \in \mathfrak{F},$$

(4.31b) $$A \subseteq B \ \Rightarrow \ \mu(B) \ = \ \mu(A) + \mu(B \setminus A),$$

(4.31c) $$\mu(A \cup B) + \mu(A \cap B) \ = \ \mu(A) + \mu(B).$$

If μ is finite then also

(4.32a) $$A \subseteq B \ \Rightarrow \ \mu(B \setminus A) \ = \ \mu(B) - \mu(A),$$

(4.32b) $$\mu(A \cup B) \ = \ \mu(A) + \mu(B) - \mu(A \cap B).$$

PROOF: The first property follows from the fact that $\mu(\emptyset) = 0$, $\emptyset \subseteq A$ for all $A \in \mathfrak{F}$ and (4.27.
To prove the second property, observe that $B = A \uplus (B \setminus A)$.
Proving (4.31c) is more complicated because neither A nor B may be a subset of the other. We have

(4.33a) $$A \cup B \ = (A \cap B) \uplus (B \setminus A) \uplus (A \setminus B)$$

(4.33b) $$A \cup B \ = A \uplus (B \setminus A) \ = \ B \uplus (A \setminus B)$$

[15] Positivity and monotony are easy, but σ–additivity is hard.

It follows from (4.33a) that

$$\mu(A \cup B) \;=\; \mu(A \cap B) + \mu(B \setminus A) + \mu(A \setminus B) \tag{4.34}$$

Since $A \cap B \subseteq A,\; B \setminus A \subseteq B,\; A \setminus B \subseteq A,$ formula (4.34) shows that $\mu(A \cup B) = \infty$ can only be true if $\mu(A) = \infty$ or $\mu(B) = \infty$. In this case (4.31c) is obviously true. Hence we may assume that $\mu(A \cup B) < \infty$.

It follows from (4.33b) that

$$2 \cdot \mu(A \cup B) \;=\; \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B) \tag{4.35}$$

We subtract the left and right sides of (4.34) from those of (4.35) and obtain

$$\begin{aligned} \mu(A \cup B) &= \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B) - \mu(A \cap B) - \mu(B \setminus A) - \mu(A \setminus B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B) \end{aligned}$$

and the third property is proved. ■

We stated as a fact without proof (Fact 4.1 on 54), that one can extend any set function which acts like a measure on the half–open parallelepipeds of $\mathbb{R}^n$ to a measure on $\mathfrak{B}(\mathbb{R}^n)$, the Borel σ–algebra of $\mathbb{R}^n$. The situation is much simpler for countable measurable spaces.

Proposition 4.6.

Let Ω be a countable, nonempty set, i.e., the elements of Ω can be written as a finite or infinite sequence $\Omega = \omega_1, \omega_2, \omega_3, \dots$ Let

$$\mathfrak{E} \;:=\; \{\,\{\omega\} : \omega \in \Omega\,\} \;=\; \{\,\textit{all singleton sets of}\ \Omega\,\}.$$

Then any nonnegative and extended real–valued function μ_0 which is defined on $\mathfrak{E}$ extends uniquely to a measure μ on the entire power set of Ω by means of the formula

$$\mu(A) \;=\; \sum_{\omega \in A} \mu_0\{\omega\}, \quad (A \subseteq \Omega). \tag{4.36}$$

PROOF: ⋆ This is immediate from the fact that $A = \biguplus_{\omega \in A} \{a\}$. ■

Example 4.4 (Binomial distribution)**.** You are very familiar with the last proposition in the context of discrete probability measures. It is then customarily written $p_n = P\{\omega_n\}$ and called a **probability mass function** (or just a **probability function** in [16] Wackerly, Mendenhall and Scheaffer: Mathematical Statistics with Applications).

For example, if we define $\Omega := \{0, 1, 2, \dots, n\}$ and $\mathfrak{F} := 2^\Omega$ then the $\text{Bin}(n,p)$ distribution is the (probability) measure P on the measurable space $(\Omega, \mathfrak{F})$ defined on the singleton events $\{0\}, \{1\}, \dots, \{n\}$ by its probability mass function

$$p_j \;:=\; P\{j\} \;:=\; \text{Bin}(n,p)\{j\} \;:=\; \binom{n}{j} p^j\,(1-p)^{n-j}. \;\square$$

We next examine the analogue of Lebesgue measure (see Definition 4.7, p.54) on the space $\mathbb{Z}$ of the integers.

Let

$$\mathfrak{E} := \{\{k\} : k \in \mathbb{Z}\} = \{\text{all singleton sets of the integers}\}.$$

According to Proposition 4.6, the function

$$\Sigma_0 : \mathfrak{E} \longrightarrow [0, \infty[; \quad \Sigma_0\{k\} := 1$$

has a unique extension

$$\Sigma : 2^{\mathbb{Z}} \longrightarrow [0, \infty], \quad \text{given by} \quad \Sigma(A) = \sum_{k \in A} 1 = |A| \text{ for all } A \subseteq \mathbb{Z}.$$

In other words, $\Sigma(A)$ is the size of A, i.e., the number of elements of$A$.

We generalize this to the d–dimensional case as follows. Recall that a symbol with an arrow on top denotes a vector. So we write, e.g.,

$$\vec{x} = (x_1, x_2, \dots, x_d)$$

for elements of $\mathbb{R}^d$. Recall that $\mathbb{Z}^d = \mathbb{Z} \times \cdots \times \mathbb{Z}$ (d factors), i.e.,

$$\mathbb{Z}^d = \{\vec{k} = (k_1, \dots, k_d) : k_1, \dots, k_d \in \mathbb{Z}\}.$$

We define the counting measure in multiple dimensions as follows. Let $d \in \mathbb{N}$ and

$$\mathfrak{E} := \{\{\vec{k}\} : \vec{k} \in \mathbb{Z}^d\} = \{\text{all singleton sets of } d\text{–dim. vectors with integer coordinates}\}.$$

Then the function

$$\Sigma_0^d : \mathfrak{E} \longrightarrow [0, \infty[; \quad \Sigma_0\{\vec{k}\} := 1$$

has according to Proposition 4.6 a unique extension

$$\Sigma^d : 2^{(\mathbb{Z}^d)} \longrightarrow [0, \infty], \quad \text{given by} \quad \Sigma^d(A) = \sum_{k \in A} 1 = |A| \text{ for all } A \subseteq \mathbb{Z}^d.$$

Here is the formal definition of counting measure.

Definition 4.8.

A. We call the measure defined by

(4.37) $$\Sigma : 2^{\mathbb{Z}} \longrightarrow [0, \infty]; \qquad A \mapsto \Sigma(A) := |A| \text{ for all } A \subseteq \mathbb{Z},$$

the **summation measure** or the **counting measure** on the integers.

B. We call the measure defined by

(4.38) $$\Sigma^d : 2^{(\mathbb{Z}^d)} \longrightarrow [0, \infty]; \qquad A \mapsto |A| \text{ for all } A \subseteq \mathbb{Z}^d.$$

the d–**dimensional summation measure** or the d–**dimensional counting measure**. □

NOTATION ALERT: The name "summation measure" is not at all common in the mathematical literature!

Proposition 4.7 (Continuity properties of measures). *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space.*

$$\text{If } B_n \uparrow B \quad \text{then} \quad \lim_{n\infty} \mu(B_n) = \mu(B) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right), \tag{4.39a}$$

$$\text{If } A_n \downarrow A \text{ in } \mathfrak{F} \text{ and } \mu(A_1) < \infty \quad \text{then} \quad \lim_{n\infty} \mu(A_n) = \mu(A) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right). \tag{4.39b}$$

PROOF: To prove formula (4.39a), we replace the sequence B_n with a disjoint sequence C_n such that $A = \biguplus_n C_n$ [16] and use the σ–additivity of μ.

To prove (4.39b), apply the already proven formula (4.39a) to

$$B_n := A_n^{\complement}, \quad B := A^{\complement}$$

(thus $B_n \uparrow B$), and note that

$$\mu(B_n) = \mu(\Omega) - \mu(A_n), \quad \mu(B) = \mu(\Omega) - \mu(A).$$

This last step requires the assumption that $\mu(A_1) < \infty$ (and thus $0 \leq \mu(A_n) \leq \mu(A_1) < \infty$). ■

Remark 4.9. The finiteness condition of formula (4.39b) is never an issue with probability measures P since $P(A) \leq 1$ for all $A \in \mathfrak{F}$. But the unexpected can happen for nonfinite measures such as the one dimensional summation measure Σ of Definition 4.8, which is characterized by

$$\Sigma(A) = |A|, \quad (A \subseteq \mathbb{Z}).$$

Here is an example of a sequence of sets $A_k \in \mathbb{Z}$ which does not satisfy the condition $\Sigma(A_1) < \infty$ (matter of fact, $\Sigma(A_k) = \infty$ for all k), and for which formula (4.39b) is not valid.

Let $A_k := \{j \in \mathbb{N} : j \geq k\}$. Then $A_k \downarrow \emptyset$ as you can see as follows.

Let $A := \bigcap_{j \in \mathbb{N}} A_j$ and assume to the contrary that A is not empty, i.e., it contains some $n \in \mathbb{N}$. This is impossible since

$$n \notin A_{n+1}, \quad \text{thus} \quad n \notin \bigcap_{n \in \mathbb{N}} A_n = A,$$

contrary to our assumption $n \in A$.

$$\text{Hence} \quad A = \emptyset, \quad \text{hence} \quad \Sigma\left(\bigcap_n A_n\right) = \Sigma(\emptyset) = 0.$$

On the other hand, $\Sigma(A_n) = \infty$ for each n, thus $\lim_{n\to\infty} \Sigma(A_n) = \infty$ since A_n contains infinitely many elements. We have found a case in which formula(4.39b) does not hold. □

[16] see Proposition 3.2 (Rewrite unions as disjoint unions) on p.40

Proposition 4.8. ⋆

Let $(\Omega, \mathfrak{F}, \mu)$ *be a measure space and* $A \in \mathfrak{F}$. *Then the set function*

$$\mu_A : \mathfrak{F} \longrightarrow [0, \infty], \qquad A' \mapsto \mu_A(A') := \mu(A \cap A')$$

defines a measure on $(\Omega, \mathfrak{F})$.

PROOF:
Only σ–additivity needs a little effort, and it follows easily from Proposition 3.1 (Distributivity of unions and intersections) on p.40. ■

Proposition 4.9. ⋆

Let $(\Omega, \mathfrak{F}, \mu)$ *be a measure space with a sequence of measures* μ_n *that satisfy*

$$\mu_n \uparrow \mu, \qquad \text{or} \qquad \mu_1(\Omega) < \infty \text{ and } \mu_n \downarrow \mu.$$

Then $\lim\limits_{n\to\infty} \mu_n : A \mapsto \lim\limits_{n\to\infty} \mu_n(A)$ *is a measure.*

PROOF: Not given here. We only mention that Proposition 4.7 (Continuity properties of measures) on p.58 is essential to show that μ is σ–additive once it has been shown to be (finitely) additive. ■

4.2 Measurable Functions and Random Elements

Introduction 4.1. We all know what a random variable X is: X has a real number as an outcome, and that outcome is random. We also know that such a random variable comes with a probability distribution.

- For example, if X is a standard normal random variable, then the probability that X attains a value $a \leq X \leq b$ can be computed as

  $$P\{a \leq X \leq b\} = \int_a^b f_X(x)\,dx, \quad \text{where } f_X(x) = \frac{1}{\sqrt{2\pi}} e^{x^2/2} \text{ is the probability density.}$$

  This is an example of a continuous random variable.
- Or X might be a discrete random variable which only attains countably many distinct outcomes $x_1, x_2, \dots$, i.e., $P\{X = x_1\} + P\{X = x_2\} + \dots = 1$. Such random variables are defined by their probability mass function

  $$p_j = P\{X = x_j\}, \quad (j = 1, 2, \dots).$$

  An example would be a Bin(n, p)–distributed random variable (see Example 4.4 (Binomial distribution) on p.56) for which $p_j = \binom{n}{j} p^j (1-p)^{n-j}$.

These settings are not general enough for our needs, and we must make some amendments.

- "... that outcome is random": Let us rephrase that as follows. The outcome of X depends on randomness. Might as well say that X is **a function** of randomness:

$$X \;=\; f(\text{randomness}).$$

  That is a great improvement but "randomness" is to wordy.
- We agree that ω means randomness: $X \;=\; f(\omega)$.
- Mathematical symbols are in short supply and it is common practice to use the same symbol for outcome (X) and assignment symbol (f). We write

$$X \;=\; X(\omega).$$

- A function needs domain and codomain. Since arguments are called ω it is natural to call the domain Ω. Since we say that random variables are real-valued functions the codomain must be $\mathbb{R}$ or a subset thereof.
- So a random variable X is a function

$$X : \Omega \longrightarrow \mathbb{R}; \qquad \omega \mapsto X(\omega).$$

- It is important to have a probability measure P defined on the domain Ω of the random variable X rather than the real numbers (the codomain of X). We have seen in Fact 4.2 on p.55 that not all measures can assign values to all subsets of Ω.
- So the domain of P might just be a σ–algebra of subsets of Ω! So Ω must be a probability space $(\Omega, \mathfrak{F}, P)$, and a random variable is a function

$$X : (\Omega, \mathfrak{F}, P) \longrightarrow \mathbb{R}; \qquad \omega \mapsto X(\omega).$$

- What good is it if there are some important events like, e.g.,

$$\{-1 \leq X \leq 1\} \;=\; \{\omega \in \Omega : -1 \leq X(\omega) \leq 1\} \;=\; X^{-1}\big([-1,1]\big)\,,$$

  for which $P\{-1 \leq X \leq 1\}$ is not available, because $\{-1 \leq X \leq 1\} \notin \mathfrak{F}$?
- What events are important, i.e., what are the sets $B \in \mathbb{R}$ such that the preimage $X^{-1}(B)$ (also written $\{X \in B\}$) [17] must belong to $\mathfrak{F}$?
- The answer to that question will generally be that the preimages $\{X \in B\}$ of Borel sets B need probabilities:

$$\text{If}\quad B \in \mathfrak{B}(\mathbb{R}) \quad \text{then we need that} \quad X^{-1}(B) \in \mathfrak{F}.$$

We have collected enough material to define random variables, but we must proceed in reverse and start with the concept of measurability which requires that the preimages of certains sets belong to the σ–algebra $\mathfrak{F}$ defined on the domain of the given random variable. □

Definition 4.9 (Measurable function)**.** Let

[17] see the **Notational conveniences II** box that follows Proposition 3.3 on p.43)

$$f \,:\, (\Omega, \mathfrak{F}) \longrightarrow (\Omega', \mathfrak{F}')$$

be a function which has the measurable space $(\Omega, \mathfrak{F})$ as its domain and the measurable space $(\Omega', \mathfrak{F}')$ as its codomain.

We say that f is $(\mathfrak{F}, \mathfrak{F}')$–**measurable**, if

(4.40) $$f^{-1}(A') \in \mathfrak{F}, \quad \text{for all } A' \in \mathfrak{F}'.$$

If $\Omega' = \mathbb{R}^n$ or $\Omega' = \mathbb{R}$ and $\mathfrak{F}'$ is the Borel σ–algebra we also say that f is $\mathfrak{F}$–**measurable**
If both $\Omega' = \mathbb{R}^n$ or $\Omega' = \mathbb{R}$ and also $\Omega = \mathbb{R}^n$ or $\Omega = \mathbb{R}$ with the Borel σ–algebras then we also say that f is Borel measurable.

We write $m(\mathfrak{F}, \mathfrak{F}')$ for the set of all $(\mathfrak{F}, \mathfrak{F}')$–measurable functions, and we write $m(\mathfrak{F})$ for the set of all $(\mathfrak{F}, \mathfrak{B})$–measurable functions (i.e., the codomain is the measure space $(\mathbb{R}, \mathfrak{B})$). Thus,

$$\begin{aligned} f \text{ is } (\mathfrak{F}, \mathfrak{F}')\text{–}\textbf{measurable} \ &\Leftrightarrow f \in m(\mathfrak{F}, \mathfrak{F}')\,, \\ f \text{ is } \mathfrak{F}\text{–}\textbf{measurable} \ &\Leftrightarrow f \in m(\mathfrak{F})\,. \ \square \end{aligned}$$

Example 4.5. **(a)** Consider the set $\mathbb{Q}_+ = \{q \in \mathbb{Q} : q \geq 0\}$ with σ–algebras $\mathfrak{F} := \{\emptyset, \mathbb{Q}_+\}$, $\mathfrak{F}' := 2^{\mathbb{Q}_+} = \{\text{ all subsets of } \mathbb{Q}_+ \}$.

Let $f : (\mathbb{Q}_+, \mathfrak{F}) \to (\mathbb{Q}_+, \mathfrak{F}')$ be defined as $f(q) = 4q$. Then f is not $(\mathfrak{F}, \mathfrak{F}')$–measurable.

For example $\{4\} \in \mathfrak{F}'$, but its preimage $\{f = 4\} = \{1\} \notin \mathfrak{F}$. Matter of fact, only constant functions with domain Ω are guaranteed to be measurable if the domain σ–algebra is $\{\emptyset, \Omega\}$. (Here, $\Omega = \mathbb{Q}_+$.)

(b) Consider the set $\mathbb{Q}_+ = \{q \in \mathbb{Q} : q \geq 0\}$ with the σ–algebra $\mathfrak{F} := 2^{\mathbb{Q}_+}$, the set $[0, \infty[$ with the σ–algebra $\mathfrak{F}' := \mathfrak{B}([0, \infty[)$ (the Borel sets of $[0, \infty[$), and the function $g : (\mathbb{Q}_+, \mathfrak{F}) \to ([0, \infty[, \mathfrak{F}')$, defined as $f(q) = \sin(\sqrt{4q})$. Then g is $(\mathfrak{F}, \mathfrak{F}')$–measurable, since any preimage of any function belongs to the power set of the domain.

(c) Consider the set $\mathbb{N}$ with σ–algebras $\mathfrak{F} := 2^{\mathbb{N}}$, $\mathfrak{F}' := \{\emptyset, \mathbb{N}\}$. Let $h : (\mathbb{N}, \mathfrak{F}) \to (\mathbb{N}, \mathfrak{F}')$ be an arbitrary function. Then b is $(\mathfrak{F}, \mathfrak{F}')$–measurable for the reason given in **(b)**. $\square$

See SCF2 Definition 1.2.1 for the next definition.

Definition 4.10 (Random Variable)**.** Let

$$X \,:\, (\Omega, \mathfrak{F}, P) \longrightarrow (\mathbb{R}, \mathfrak{B})$$

be a function which has a probability space $(\Omega, \mathfrak{F}, P)$ as its domain and the real numbers with the Borel σ–algebra as its codomain.
If X is $\mathfrak{F}$–measurable, i.e.,

(4.41) $$\{X \in B\} \text{ belongs to } \mathfrak{F} \text{ for all Borel sets } B,$$

then we call X a **random variable**. on $(\Omega, \mathfrak{F}, P)$.
If there is a countable subset A of $\mathbb{R}$ such that the random variable X "lives" on A, i.e.,

$$X(\Omega) \;=\; \{X(\omega) \,:\, \omega \in \Omega\} \;\subseteq\; A$$

then we call X a **discrete random variable**. □

Remark 4.10. ⋆

(1) If X is a discrete random variable and $A = \{x_1, x_2, \dots\}$ is countable set which contains the range $X(\Omega)$ of X then we can shrink the codomain of X to the measurable space $(A, 2^A)$ and talk about the random variable

$$X \,:\, (\Omega, \mathfrak{F}, P) \longrightarrow (A, 2^A).$$

Here is the reason that we can and often will take the entire power set 2^A as the σ–algebra of the codomain of X:

- All singletons $\{a\} \subseteq A$ are Borel sets, thus each $B \subseteq A$ is Borel since it is the countable union $B = \bigcup_{a \in B} \{a\}$ of Borel sets.

(2) Occasionally we allow X to assume the values ∞, and $-\infty$, i.e., X can be an extended real–valued, $\mathfrak{F}$–measurable, function. □

It seems awkward not to call a measurable function $\Omega \to \Omega'$ from a probability space $(\Omega, \mathfrak{F}, P)$ to a measurable space $(\Omega', \mathfrak{F}')$ a random variable only because its function values are not numbers. We give a name to such measurable functions of randomness.

Definition 4.11 (Random element)**.** ⋆

- Let $(\Omega, \mathfrak{F}, P)$ be a probability space, $(\Omega', \mathfrak{F}')$ a measurable space. A **random element** is an $(\mathfrak{F}, \mathfrak{F}')$–measurable function $X : \Omega \to \Omega'$. □

Note that all random variables are random elements.

For the following see also SCF2 Definition 1.3.9 and SCF2 Definition 1.1.5.

Definition 4.12 (Almost everywhere and almost surely)**.** Let $(\Omega, \mathfrak{F})$ be a measurable space and let A be the set of all $\omega \in \Omega$ such that a certain property is true. For example,

- $A = \{\omega \in \Omega : f(\omega) \leq g(\omega)\}$,
- $A = \{\omega \in \Omega :$ the function $t \mapsto Y_t(\omega)$ is continuous $\}$,
- $A = \{\omega \in \Omega : \; |X(\omega)| \leq 1\}$.

(1) In the context of a measure space $(\Omega, \mathfrak{F}, \mu)$ we say that the property is satisfied, or holds, or is true μ–**almost everywhere** if $\mu(A^\complement) = 0$. We also write μ–**a.e.**

(2) In the context of a probability space $(\Omega, \mathfrak{F}, P)$ we say that the property is satisfied, or holds, or is true P–**almost surely** if $P(A^\complement) = 0$ or, equivalently, if $P(A) = 1$. We also write P–**a.s.**

(3) In either case we will drop the μ– and P– prefixes if there is no confusion about which measure or probability this refers to. □

Remark 4.11. ⋆ The set A might not be measurable. To be precise we would have had to formulate the above as follows. The property holds μ–a.e. if there is a measurable set B such that $\mu(B) = 0$ and B contains the set $A^\complement$ on which this property is not satisfied. We will not worry about such fine points concerning measurability. □

Remark 4.12. We follow the lead of SCF2 and often will not explicitly mention that a certain property is assumed to be true or can be proven to be true only almost everywhere/almost surely. □

Remark 4.13.

Since random variables are special cases of measurable functions, it follows that
All statements that are true for measurable functions are true for random variables! □

Theorem 4.1.

Let $(\Omega, \mathfrak{F})$ and $(\Omega', \mathfrak{F}')$ be measurable spaces and $f : \Omega \to \Omega'$. Let $\mathfrak{E}' \subseteq \mathfrak{F}'$ be a generator of $\mathfrak{F}'$, i.e.,

$$\sigma(\mathfrak{E}') = \mathfrak{F}'.$$

to prove that f is $(\mathfrak{F}, \mathfrak{F}')$–measurable it suffices to show that

$$f^{-1}(A') \subseteq \mathfrak{F} \text{ for all } A' \in \mathfrak{E}'. \tag{4.42}$$

PROOF: ⋆

Step 1. We show that

$$\mathscr{H}' := \{ H' \subseteq \Omega' : f^{-1}(H') \in \mathfrak{F} \} \quad \text{is a } \sigma\text{–algebra.}$$

Clearly, $\emptyset \in \mathscr{H}'$. We will show that countable unions of sets in $\mathscr{H}'$ also belong to $\mathscr{H}'$. The proof that $H' \in \mathscr{H}'$ implies $H' \in \mathscr{H}'$ is similar.

Let $H'_n \in \mathscr{H}'$ for $n \in \mathbb{N}$. Then $f^{-1}(H'_n) \in \mathfrak{F}$ by definition of $\mathscr{H}'$. Since $\mathfrak{F}$ is a σ–algebra, $\bigcup_n f^{-1}(H'_n) \in \mathfrak{F}$. Since $\bigcup_n f^{-1}(H'_n) = f^{-1}\big(\bigcup_n H'_n\big)$ by Theorem 3.4 (f^{-1} is compatible with all basic set ops) on p.43, it follows that $f^{-1}\big(\bigcup_n H'_n\big) \in \mathfrak{F}$, i.e., $\bigcup_n H'_n \in \mathscr{H}'$.

Step 2. By assumption, $f^{-1}(E') \in \mathfrak{F}$ for all $E' \in \mathfrak{E}'$. Thus, $\mathfrak{E}' \subseteq \mathscr{H}'$, thus,

$$\sigma(\mathfrak{E}') \subseteq \sigma(\mathscr{H}') \tag{$\star$}$$

Since $\sigma(\mathfrak{E}') = \mathfrak{F}'$ by assumption, and $\mathscr{H}' = \sigma(\mathscr{H}')$ by **Step 1**, it follows from $(\star)$ that $\mathfrak{F}' \subseteq \mathscr{H}'$, i.e., $f^{-1}(A') \in \mathfrak{F}$ for all $A' \in \mathfrak{F}'$. Thus, $f \in m(\mathfrak{F}, \mathfrak{F}')$. ■

Corollary 4.1. *Let $(\Omega, \mathfrak{F})$ be a measurable space and $f : (\Omega, \mathfrak{F}) \to (\mathbb{R}, \mathfrak{B}^1)$. to prove that f is $\mathfrak{F}$–measurable it suffices to show that one of the following four conditions is met:*

(1) $\{f < c\} \in \mathfrak{F}$ *for all* $c \in \mathbb{R}$,
(2) $\{f \leq c\} \in \mathfrak{F}$ *for all* $c \in \mathbb{R}$,
(3) $\{f > c\} \in \mathfrak{F}$ *for all* $c \in \mathbb{R}$,
(4) $\{f \geq c\} \in \mathfrak{F}$ *for all* $c \in \mathbb{R}$. □

Note that this implies the following. If the domain of f actually is a probability space $(\Omega, \mathfrak{F}, P)$ then f is a random variable if one of the above four conditions is satisfied.

PROOF: ⋆ Essentially follows from Theorem 4.1 above and Remark 4.5 on p.52. ■

Proposition 4.10.

- *Any continuous function $f : \mathbb{R}^m \to \mathbb{R}^n$ is Borel–measurable, i.e., $(\mathfrak{B}^m, \mathfrak{B}^n)$–measurable.*
- *In particular, any continuous, real–valued function $f(x)$ of real values x is Borel–measurable.* □

PROOF: ⋆ A triviality if you recall that the open n–dimensional parallelepipeds generate $\mathfrak{B}^n$ and if you know the following:

f is continuous (at each $\vec{x} \in \mathbb{R}^m$) $\Leftrightarrow$ the preimages of all open sets in $\mathbb{R}^n$ are open in $\mathbb{R}^m$. ■

Proposition 4.11. ⋆

Let $(\Omega, \mathfrak{F})$ be a measurable space and f, g extended real–valued Borel measurable functions. Then each one of the sets

$$\{f < g\}, \quad \{f \leq g\}, \quad \{f > g\}, \quad \{f \geq g\},$$

is $\mathfrak{F}$–measurable.

PROOF:

For the set $\{f < g\}$ we proceed as follows. For $q \in \mathbb{Q}$ let $A_q := \{f < q < g\}$. Then $A_q = \{f < q\} \cap \{q < g\}$ is measurable as the intersection of two measurable sets. Note that

$$f(\omega) < g(\omega) \quad \Leftrightarrow \quad \text{there is (at least one) } q \in \mathbb{Q} \text{ such that } f(\omega) < q < g(\omega)\,,$$

and thus

$$\{f < g\} = \bigcup_{q \in \mathbb{Q}} A_q\,.$$

It follows that $\{f < g\}$ is measurable as the countable union of the measurable sets A_q.

From this we obtain measurabilty of the set $\{f \leq g\}$ since

$$\{f \leq g\} = \bigcap_{n \in \mathbb{N}} \left\{ f < g + \frac{1}{n} \right\}.$$

Lastly, $\{f > g\}$ and $\{f \geq g\}$ are measurable as complements of the measurable sets $\{f \leq g\}$ and $\{f < g\}$ ■

For the following see Definitions 2.17 and 2.18 on p.20.

Theorem 4.2.

Let $(\Omega, \mathfrak{F})$ be a measurable space and $f, g : \Omega \to \mathbb{R}$. Let $c \in \mathbb{R}$.
If f, g in $m(\mathfrak{F})$ then each of the following also is $(\mathfrak{F}, \mathfrak{B})$–measurable:

$$c, \quad cf, \quad f \pm g, \quad fg; \quad f/g \text{ (on } \{g \neq 0\}), |f|, \quad f^+, \quad f^-, \quad f \vee g, \quad f \wedge g.$$

Here c denotes the constant function $\omega \mapsto c$ and cf denotes the function $\omega \mapsto cf(\omega)$.

- *Moreover, all statements above which involve two functions f and g generalize to finitely many measurable functions $f_1, f_2, \dots, f_n$.*
- *Moreover, the statements about $f \vee g$ and $f \wedge g$ generalize to sequences $(f_n)_n$ of functions as follows: If each f_n is measurable then so are the functions*

$$\sup_n f_n : \omega \mapsto \sup\{f_n(\omega) : n \in \mathbb{N}\}, \qquad \inf_n f_n : \omega \mapsto \inf\{f_n(\omega) : n \in \mathbb{N}\}.$$

PROOF: Omitted except for this one:
We prove that $f(\omega) := \sup_n f_n(\omega)$ is measurable as follows. Observe that for any $c \in \mathbb{R}$ it is true that

$$f(\omega) \leq c \quad \Leftrightarrow \quad f_n(\omega) \leq c \text{ for all } n,$$

thus

$$\{ f \leq c \} = \bigcap_{n \in \mathbb{N}} \{ f_n \leq c \},$$

and this set is $\mathfrak{F}$–measurable as the intersection of the $\mathfrak{F}$–measurable sets $\{f_n \leq c\}$. The assertion now follows from Corollary 4.1. ■

Example 4.6 (Binomial random variable v.s. binomial distribution)**.** This example continues Example 4.4 (Binomial distribution) on p.56 which was about the binomial distribution $\text{Bin}(n, p)$ defined by its probability mass function

$$p_j = P\{j\} = \binom{n}{j} p^j (1-p)^{n-j}. \tag{4.43}$$

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let $X \in m(\mathfrak{F})$, i.e., X is a random variable on $(\Omega, \mathfrak{F}, P)$. We all are familiar with what it means that X is a $\text{Bin}(n, p)$–distributed random variable. It satisfies formula (4.43), right?

Not exactly! There is a problem with the probability P. In formula (4.43) it occurs as a measure on the measurable space

$$\big(\{0,1,\dots,n\}, 2^{\{0,1,\dots,n\}}\big)$$

and NOT on our abstract measurable space $(\Omega, \mathfrak{F})$ which may not have numbers $0, 1, 2, \dots$ as elements ω.

Here is the explanation. These numbers j are not the argument ω of the random variable $\omega \mapsto X(\omega)$; they are the function values $j = X(\omega)$. If, by chance, randomness occurs as ω_1, then the associated outcome for X might be, e.g., $X(\omega_1) = 7$. On the other hand, if ω_2 happens instead, then we observe $X(\omega_2)$, and that outcome might be $X(\omega_2) = 4$. And if ω_3 happens instead, then we observe the outcome $X(\omega_3)$, which might again be 7, and so on.

So the answer is that $\text{Bin}(n,p)\{j\} = \binom{n}{j} p^j (1-p)^{n-j}$ refers to events on the codomain $(\mathbb{R}, \mathfrak{B}^1)$ of X, and this leads to the following question.

- There must be a relationship between the measure P on $(\Omega, \mathfrak{F})$, the random variable X, and the measure $\text{Bin}(n,p)$ on $(\mathbb{R}, \mathfrak{B}^1)$. What is it?

The answer to the first question was given in Introduction 4.1 to this chapter 4.2 (Measurable Functions and Random Elements). See p.59. We will use X and P to build a measure P_X on $(\mathbb{R}, \mathfrak{B}^1)$ as follows:

$$P_X(B) \;:=\; P\{X \in B\} \;=\; P\{\omega \in \Omega \,:\, X(\omega) \in B\}, \;\; (B \in \mathfrak{B}^1).$$

That will work for any random variable. Matter of fact, that will work for any measurable function $f : (\Omega, \mathfrak{F}, \mu) \to (\Omega', \mathfrak{F}')$, since we can define a measure μ_f on $\mathfrak{F}'$ from the measure μ on $\mathfrak{F}$ via

$$\mu_f(A') \;:=\; \mu\{f \in A'\} \;=\; \mu\{\omega \in \Omega \,:\, f(\omega) \in A'\}, \;\; (A' \in \mathfrak{F}'). \;\; \square$$

Proposition 4.12. *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and $(\Omega', \mathfrak{F}')$ a measurable space.*

> *Let $f : \Omega \to \Omega'$ be $(\mathfrak{F}, \mathfrak{F}')$ measurable. Then the set function*
>
> (4.44) $$\mu_f : \mathfrak{F}' \to [0, \infty];\, A' \mapsto \mu\{f \in A'\} \;=\; \mu\{\omega \in \Omega \,:\, f(\omega) \in A'\}$$
>
> *defines a measure on $(\Omega', \mathfrak{F}')$. Moreover, if μ is a probability measure on $\mathfrak{F}$, i.e., $\mu(\Omega = 1)$, then μ_f is a probability measure on $\mathfrak{F}'$.*

PROOF: [$\star$] $\mu_f(\emptyset) = 0$, since $f^{-1}(\emptyset) = \emptyset$, and μ is a measure.

We show here in detail that μ_f is monotone: $A \subseteq B \Rightarrow \mu_f(A) \le \mu_f(B)$, for all $A, B \in \mathfrak{F}'$. According to Proposition 3.3 on p.43, $A \subseteq B$ implies $f^{-1}(A) \subseteq f^{-1}(B)$. Since μ is a measure, this implies $\mu\big(f^{-1}(A)\big) \le \mu\big(f^{-1}(B)\big)$, i.e., by definition of μ_f, $\mu_f(A) \le \mu_f(B)$

The proof that $\mu_f\left(\biguplus_n B_n\right) = \sum_n \mu_f(B_n)$ for any disjoint sequence $B_n \in \mathfrak{F}'$, is just as simple, since the order of taking preimages and unions can be switched. See Proposition 3.4 (f^{-1} is compatible with all basic set ops) on p.43. ■

For the following see SCF2 Definition 1.2.3.

Definition 4.13 (Image measure)**.**

(1) We call the measure μ_f of Proposition 4.12 the **image measure** of μ under f or the **measure induced** by μ and f.

(2) We now switch notation from f and μ to the more customary X and P for the sake of clarity. In the case of a random variable X on a probability space $(\Omega, \mathfrak{F}, P)$ we call the image measure P_X of P under X which is, according to (4.44), given by

$$P_X(B) := P\{X \in B\} = P\{\omega \in \Omega : X(\omega) \in B\}, \ (B \in \mathfrak{B}^1) \tag{4.45}$$

the **probability distribution** or simply the **distribution** of X. SCF2 also calls P_X the **distribution measure** of X. □

Proposition 4.13.

Let Ω be a nonempty set, $(\Omega', \mathfrak{F}')$ a measurable space, and $f : \Omega \to \Omega'$ an arbitrary function. Then

(1) *the collection $\sigma(f) := \{f^{-1}(A') : A' \in \mathfrak{F}'\}$ of all preimages of $\mathfrak{F}'$–measurable sets is a σ–algebra.*

(2) *The function f is $(\sigma(f), \mathfrak{F}')$–measurable.*

(3) *$\sigma(f)$ is the smallest σ–algebra $\mathfrak{F}$ on Ω which makes f $(\mathfrak{F}, \mathfrak{F}')$–measurable in the following sense: If $\mathfrak{F}$ is a σ–algebra on Ω and there are sets $A \in \sigma(f)$ which do not belong to $\mathfrak{F}$, then f is not $(\mathfrak{F}, \mathfrak{F}')$–measurable.*

PROOF: ⋆

(1) follows from Proposition3.4 (f^{-1} is compatible with all basic set ops) on p.43.

(2) is easy to see from the definition of measurability of a function. ■

Definition 4.14. Let Ω, Ω' be nonempty, $\mathfrak{F}'$ a σ–algebra on Ω', and $f : \Omega \to \Omega'$.

We call the σ–algebra from Proposition 4.13

$$\sigma(f) := \{f^{-1}(A') : A' \in \mathfrak{F}'\} \tag{4.46}$$

the **σ–algebra generated by** f. □

Remark 4.14. Assume that $f : (\Omega, \mathfrak{F}) \to (\Omega', \mathfrak{F}')$ with measurable spaces for both domain and codomain.

(1) The minimality of $\sigma(f)$ stated in Proposition 4.13.**(3)** implies that
f is $(\mathfrak{F}, \mathfrak{F}')$–measurable $\Leftrightarrow \sigma(f) \subseteq \mathfrak{F}$.

(2) In particular, if X is a random variable defined on a probability space $(\Omega, \mathfrak{F}, P)$, then $\sigma(X) \subseteq \mathfrak{F}$, since X is $\mathfrak{F}$ measurable by the very definition of a random variable.

(3) In a sense we can think of $\sigma(X)$ as the information one associates with a random element X. This is discussed at length in Chapter 5 (Conditional Expectations) and in SCF2, ch.2. □

4.3 Convergence of Function Sequences

This subchapter is very skeletal in nature. It contains excerpts of [10] Fochler, Michael: Lecture Notes for Math 447 - Probability.

Much of this material is optional.

Definition 4.15 (Convergence of Random Variables)**.**

Let Y_n ($n \in \mathbb{N}$) and Y be random variables on a probability space (Ω, P). We define

(4.47) $Y_n \overset{\text{pw}}{\to} Y$ or $\text{pw} - \lim\limits_{n\to\infty} Y_n = Y$, if $\lim\limits_{n\to\infty} Y_n(\omega) = Y(\omega)$, for all $\omega \in \Omega$,

(4.48) $Y_n \overset{\text{a.s.}}{\to} Y$ or $\text{a.s.} - \lim\limits_{n\to\infty} Y_n = Y$, if $P\{\omega \in \Omega : \lim\limits_{n\to\infty} Y_n(\omega) = Y(\omega)\} = 1$,

(4.49) $Y_n \overset{\text{P}}{\to} Y$ or $\text{P} - \lim\limits_{n\to\infty} Y_n = Y$, if $\forall\, \varepsilon > 0 \ \lim\limits_{n\to\infty} P\{\omega \in \Omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\} = 0$,

(4.50) $Y_n \overset{\text{D}}{\to} Y$, if $\lim\limits_{n\to\infty} F_{Y_n}(y) = F_Y(y)$, $\forall\, y \in \mathbb{R}$ where the CDF F_Y of Y is continuous.

We also say:
If $Y_n \overset{\text{pw}}{\to} Y$, Y is the **pointwise limit** of the Y_n, or: Y_n **converges pointwise** to Y.
If $Y_n \overset{\text{a.s.}}{\to} Y$, Y is the **almost sure limit** of the Y_n, or: Y_n **converges almost surely** to Y.
If $Y_n \overset{\text{P}}{\to} Y$, Y is the **limit in probability**; of the Y_n, or: Y_n **converges in probability** to Y.
If $Y_n \overset{\text{D}}{\to} Y$, Y is the **limit in distribution** of the Y_n, or: Y_n **converges in distribution** to Y.

Example 4.7. $\star$

Consider $\Omega := [0, 1]$ as a probability space (Ω, P) by defining

$$P(]a, b]) := b - a, \text{ for } 0 \le a < b \le 1 .$$

In other words, P is the uniform distribution on $[0, 1]$.

We rename the functions f_n, f, g, h of (**??**) in the introduction to Y_n, Y, U, V, since doing so will make it less confusing to examine the convergence behavior of the sequence. This particularly applies to converges in probability and in distribution. Accordingly, we define

$$Y_n(\omega) := \omega^n, \ \ U(\omega) = 0, \ \ V(\omega) := \omega, \ \ (\text{for } 0 \le \omega \le 1) \quad Y(\omega) := \begin{cases} 0, & \text{if } 0 \le \omega < 1, \\ 1, & \text{if } \omega = 1 . \end{cases}$$

Part I: Pointwise and a.s convergence

Pointwise convergence behavior of the Y_n corrresponds to that of (**??**):

- Y is the pointwise limit of the sequence Y_n,
- U is the pointwise limit of the Y_n on $[0, 1[$ only, but not on Ω,
- V is not the pointwise limit of the Y_n (except for $\omega = 0)$ or $\omega = 1)$.

With respect to almost sure convergence, we see that

- $Y_n \overset{\text{a.s.}}{\to} Y$, since $\{\lim_{n\to\infty} Y_n = Y\} = [0,1] = \Omega$, and $P(\Omega) = 1$.
- $Y_n \overset{\text{a.s.}}{\to} U$, since $\{\lim_{n\to\infty} Y_n \neq U\} = \{1\}$, and $P(\{1\}) = 0$.
- $(Y_n)_n$ does not converge to V a.s., since $P\{\lim_{n\to\infty} Y_n = V\} = P\{0,1\} = 0 \neq 1$.

Part II: Convergence in probability

Next, we examine convergence in probability. We will see that a sequence of random variables can have more than one P–limit by showing the following: The sequence $\omega \mapsto Y_n(\omega) = \omega^n$ has both $\omega \mapsto U(\omega) = 0$ and $\omega \mapsto Y(\omega) = 1$ if $\omega = 1$ and 0 else as P–limits.

By definition of $P\text{–}\lim_{n\to\infty} Y_n = \widetilde{Y}$, we must prove that, for any fixed, but arbitrary $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{\,|Y_n - \widetilde{Y}| > \varepsilon\,\} = 0. \qquad \text{See (4.49).}$$

Since this probability decreases as ε increases and we must show that it approaches 0 as $n \to \infty$, we only need to worry about the very small ε. Thus, we may assume that $0 < \varepsilon < 1$.

We observe that, for $Y_n(\omega) = \omega^n$ and $0 < \varepsilon < 1$,

(A)
$$\begin{aligned} &\big[\,|Y_n(\omega)| \geq \varepsilon \;\Leftrightarrow\; \omega^n \geq \varepsilon \;\Leftrightarrow\; \omega \geq \varepsilon^{1/n}\,\big] \\ \Rightarrow\; &\big[\,P\{\,|Y_n| \geq \varepsilon\,\} = P\big([\varepsilon^{1/n}, 1]\big) = 1 - \varepsilon^{1/n}\,\big]. \end{aligned}$$

(B)
$$0 < \varepsilon < 1 \;\Rightarrow\; \lim_{n\to\infty} \varepsilon^{1/n} = 1 \;\Rightarrow\; \lim_{n\to\infty} \big(1 - \varepsilon^{1/n}\big) = 0.$$

Part II (1): We now prove that $P\text{–}\lim_{n\to\infty} Y_n = Y$:

(a)
$$\begin{aligned} &\big[\,|Y_n(\omega) - Y(\omega)| \geq \varepsilon \;\Leftrightarrow\; |Y_n(\omega)| \geq \varepsilon \text{ and } \omega \neq 1\,\big] \\ \Rightarrow\; &\big[\,P\{|Y_n - Y| \geq \varepsilon\} \leq P\{|Y_n| \geq \varepsilon\} \overset{\textbf{(A)}}{=} 1 - \varepsilon^{1/n} \overset{\textbf{(B)}}{\to} 0, \text{ as } n \to \infty.\,\big]. \end{aligned}$$

Thus, $\lim_{n\to\infty} P\{|Y_n - Y| \geq \varepsilon\} = 0$.

Part II (2): We now prove that $P\text{–}\lim_{n\to\infty} Y_n = U$:

- We could repeat the proof for the P–convergence of Y_n to Y with very minor modifications and the reader is encouraged to do so. Instead, we will use that result to show that $P\text{–}\lim_{n\to\infty} Y_n = U$
- Since the outcome $\{1\}$ has probability zero and $Y(\omega) = U(\omega)$ for $\omega \neq 1$,

$$\begin{aligned} P\{|Y_n - Y| \geq \varepsilon\} &= P\{|Y_n - Y| \geq \varepsilon \text{ and } \omega \neq 1\} \\ &= P\{|Y_n - U| \geq \varepsilon \text{ and } \omega \neq 1\} = P\{|Y_n - U| \geq \varepsilon\}. \end{aligned}$$

- Since $\lim_{n\to\infty} P\{|Y_n - Y| \geq \varepsilon\} = 0$,

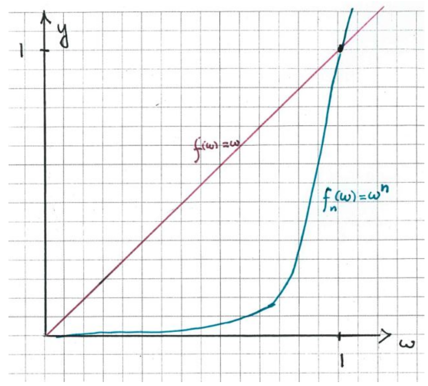
$$\lim_{n\to\infty} P\{|Y_n - U| \geq \varepsilon\} = \lim_{n\to\infty} P\{|Y_n - Y| \geq \varepsilon\} = 0.$$

Thus, $P\text{–}\lim_{n\to\infty} Y_n = U$.

Part II (3): Next, we show that it is not true that $(Y_n)_n$ converges in probability to V.

We argue by picture rather than giving an exact proof, since that would require some very tedious of terms containing $\ln(k)$.

- The picture makes it very clear that $\varepsilon = 1/10 \;\Rightarrow\; \omega - \omega^n > \varepsilon$ for $\frac{49}{100} \le \omega \le \frac{51}{100}$ and $n \ge 100$.
  Thus, $P\{|Y_n - V| \ge \varepsilon\} \ge \varepsilon \cdot \left(\frac{51}{100} - \frac{49}{100}\right) = \frac{2}{1000}$.
  Thus, $\lim\limits_{n\to\infty} P\{|Y_n - V| \ge \varepsilon\} = 0$ is not true.
- Since $\lim\limits_{n\to\infty} P\{|Y_n - V| \ge \varepsilon\} = 0$ must hold for **ALL** ε and we showed that this is not so for $\varepsilon = \dfrac{1}{10}$, it follows that $(Y_n)_n$ does not converge in probability to V.



Part III: Convergence in distribution

We will show that Y_n does not converge to V in distribution as follows.

- Let $0 < y < 1$. We recall that $P]a,b] = b - a$, for all $0 \le a < b \le 1$.
- From $V(\omega) = \omega$, we get $F_V(y) = P\{V \le y\} = P\{\omega \in \Omega : V(\omega) \le y\} = P]0,y] = y$.
- Since $Y_n(\omega) = \omega^n$, $F_{Y_n}(y) = P\{Y_n \le y\} = P\{\omega \in \Omega : \omega^n \le y\} = P]0, y^{1/n}] = y^{1/n}$.
- We note that $0 < y < 1 \;\Rightarrow\; \lim\limits_{n\to\infty} y^{1/n} = 1$.
  Thus, $F_V(y) = y$, whereas, $\lim\limits_{n\to\infty} F_{Y_n}(y) = 1$ for $0 < y < 1$.
  Thus, $\lim\limits_{n\to\infty} F_{Y_n}(y) \ne F_V(y)$ for $0 < y < 1$.
- Since all those y are points of continuity for F_V, it follows that $(Y_n)_n$ does not converge in distribution to V.

On the other hand, the theorem that follows now shows that $(Y_n)_n$ converges in distribution to Y and U, since we have shown convergence in probability to those random variables. ☐

Theorem 4.3 (Relationship between the modes of convergence). ⋆

Let Y and $Y_1, Y_2, \dots$ be random variables on a probability space (Ω, P). Then,

(4.51) $$Y_n \overset{pw}{\to} Y \;\Rightarrow\; Y_n \overset{a.s.}{\to} Y \;\Rightarrow\; Y_n \overset{P}{\to} Y \;\Rightarrow\; Y_n \overset{D}{\to} Y\,.$$

Moreover, if Y, the prospective limit, is constant a.s. (so that $P\{Y = E[Y]\} = 1$), then

(4.52) $$Y_n \overset{P}{\to} Y \;\Leftrightarrow\; Y_n \overset{D}{\to} Y\,.$$

PROOF:

I: It is obvious that $Y_n \overset{\mathbf{pw}}{\to} Y \;\Rightarrow\; Y_n \overset{\mathbf{a.s.}}{\to} Y$ for the following reason:

- Let $A := \{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) \neq Y(\omega)\}$.
- Then, $Y_n \overset{\text{pw}}{\to} Y \Rightarrow A = \emptyset \Rightarrow P(A) = 0 \Rightarrow Y_n \overset{\text{a.s.}}{\to} Y$.

II: The proofs that $Y_n \overset{\text{a.s.}}{\to} Y \Rightarrow Y_n \overset{\text{P}}{\to} Y$ and $Y_n \overset{\text{P}}{\to} Y \Rightarrow Y_n \overset{\text{D}}{\to} Y$ are outside the scope of this course. Fairly accessible proofs for those who can work with sets like

$$\bigcap_{n\geq 1}\left(\bigcup_{j\geq n}\{\omega \in \Omega : |Y_j(\omega) - Y(\omega)| \geq \varepsilon\}\right)$$

and are familiar with the exact definition of convergence of sequences [18] can be found at this Wikipedia link. ■

There are many theorems concerning the convergence of random variables. We only mention here the following two which will be used later in this chapter.

Theorem 4.4 (Slutsky's Theorem). ⋆

Let $Y_1, Y_2, \dots)$ and $U_1, U_2, \dots$ be two sequences of random variables. Let Y be another random variable and c a constant such that

- $Y_n \xrightarrow{D} Y$ *(convergence in distribution)*
- $U_n \xrightarrow{P} c$ *(convergence in probability)*

Then,

$$Y_n + U_n \xrightarrow{D} Y + c\,, \tag{4.53}$$

$$Y_n \cdot U_n \xrightarrow{D} cY\,, \tag{4.54}$$

$$\frac{Y_n}{U_n} \xrightarrow{D} \frac{Y_n}{c}\,, \quad \textit{assuming that } c \neq 0. \tag{4.55}$$

PROOF: Omitted. See, e.g., [6] Bickel and Doksum: Mathematical Statistics.

Theorem 4.5 (Convergence is maintained under continuous transformations). ⋆

Let $Y_1, Y_2, \dots)$ and Y be random variables on some probability space (Ω, P). Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Then,

$$Y_n \xrightarrow{a.s.} Y \Rightarrow f \circ Y_n \xrightarrow{a.s.} f \circ Y\,.$$

$$Y_n \xrightarrow{P} Y \Rightarrow f \circ Y_n \xrightarrow{P} f \circ Y\,.$$

$$Y_n \xrightarrow{D} Y \Rightarrow f \circ Y_n \xrightarrow{D} f \circ Y\,.$$

PROOF: Omitted. [19] ■

[18] x_n converges to $x \Leftrightarrow$ for all $\varepsilon > 0$ one can find $N \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ whenever $n \geq N$.

[19] A proof can be found at this Convergence of random variables (Mann–Wald theorem, general transformation theorem) Wikipedia link.

Example 4.8 (Convergence in probability but not a.s.). ⋆

Consider the "sliding hump" example. [20] As our probability space we choose $\Omega := [0,1]$, the unit interval in $\mathbb{R}$, with the probability measure defined by $P\big(]a,b]\big) := b-a$.

(a) We partition Ω into the two intervals $I_1 = [0,1/2]$ and $I_2 =]1/2,1]$.

- For $n = 1,2$, let $Y_n(\omega) := \begin{cases} 1\,, & \text{if } \omega \in I_n, \\ 0\,, & \text{else}\,. \end{cases}$

(b) We partition Ω into the three intervals $I_3 = [0,1/3]$, $I_4 =]1/3,2/3]$, and $I_5 =]2/3,1]$, then into $I_6 = [0,1/4]$, $I_7 =]1/4,2/4]$, $I_8 =]2/4,3/4]$, and $I_9 =]3/4,1]$, and so on .....

- We define random variables Y_n as in **(a)**: For $n \in \mathbb{N}$, let $Y_n(\omega) := \begin{cases} 1\,, & \text{if } \omega \in I_n, \\ 0\,, & \text{else}\,. \end{cases}$

(c) Then the sequence Y_n converges in probability to the (deterministic) random variable $\omega \mapsto Y(\omega) := 0$. A proof is given directly after this example.

(d) But this sequence of random variables does not converge almost surely. In fact, there is no $0 \le \omega \le 1$ for which $\lim\limits_{n\to\infty} Y_n(\omega)$ exist:

- Fix $\omega \in [0,1]$. By construction, there are indices $n_1 = n_1(\omega) < n_2 = n_2(\omega) < n_3 = n_3(\omega) < \cdots$, such that $\omega \in I_{n_k}$ and I_{n_k} has length $1/k$. (Thus, $P(I_{n_k}) = 1/k$.)

(e) Let $\omega' \in [0,1]; \omega' \neq \omega$. The subsequences $n_k(\omega)$ and $n_k(\omega')$ will differ for all k so large that $\dfrac{1}{k} < \dfrac{|\omega-\omega'|}{2}$, i.e., $\dfrac{2}{k} < |\omega-\omega'|$, since $\omega \in I_{n_k(\omega)}$ and $\omega' \in I_{n_k(\omega')} \Rightarrow I_{n_k(\omega)} \cap I_{n_k(\omega')} = \emptyset$. (Draw a picture!)

(f) It follows for such big k, that $Y_{n_k(\omega)}(\omega) = 1$ and $Y_{n_k(\omega)}(\omega') = 0$.
On the other hand, $Y_{n_k(\omega')}(\omega) = 0$ and $Y_{n_k(\omega')}(\omega') = 1$.
Thus, the full sequences $Y_n(\omega)$ does not have a limit, since it would have to be 1 along the subsequence $n_k(\omega)$ and 0 along the subsequence $n_k(\omega')$.

(g) ω is arbitrary in $\Omega = [0,1]$. This shows that there is no $\omega \in \Omega$ for which $\lim\limits_{n\to\infty} Y_n(\omega)$ exists. ☐

PROOF that (Y_n) converges in probability:
If we write $|I_n|$ for the length of the interval I_n, then

(h) ⊡ $|I_n| = 1 \Leftrightarrow n = 1$ ⊡ $|I_n| = 1/2 \Leftrightarrow n = 2,3$ ⊡ $|I_n| = 1/3 \Leftrightarrow n = 4,5,6$.
Thus, if $s_1 = 1,\ s_2 = s_1 + 2,\ s_3 = s_2 + 3, \dots, s_k = s_{k-1} + k = \sum\limits_{j=1}^{k} j = \dfrac{k\cdot(k+2)}{2}, \dots,$

(i) then $I_n = 1/k \Leftrightarrow n = s_{k-1}+1,\ s_{k-1}+2, \dots, s_{k-1}+k \Leftrightarrow s_{k-1} < n \le s_k$.

(j) It should be clear that $\big[\, n \to \infty \,\big]\big[\, k \to \infty \,\big]$ For a proof: ⊡ "$\Leftarrow$" follows from $n \ge k$.
⊡For the other direction, we observe that $n \overset{\text{(i)}}{\le} 2s_k = 2k(k+1) < 2(k+1)^2$, i.e., $\sqrt{n/2} - 1 < k$. Thus, $\big[\, n \to \infty \,\big] \Rightarrow \big[\, k \to \infty \,\big]$ and "$\Rightarrow$" follows.

[20] See this StackExchange link.

(k) Since $Y_n(\omega) := \begin{cases} 1\,, & \text{if } \omega \in I_n, \\ 0\,, & \text{else} \end{cases}$ for $n \in \mathbb{N}$, we obtain $P\{|Y_n| \geq \varepsilon\} = 0$ for $\varepsilon \leq 1$ and, with n_k as defined in **(k)**, $P\{|Y_{n_k}| \geq \varepsilon\} = \dfrac{1}{k}$ for $0 < \varepsilon \geq 1$. Thus, $P\{|Y_{n_k}| \geq \varepsilon\} \leq \dfrac{1}{k}$ for $\varepsilon > 0$.

(l) Fix $\varepsilon > 0$ and $k \in \mathbb{N}$. $|I_n|$ and hence, $P\{|Y_n| > \varepsilon\}$ is nonincreasing with n. Thus, $n \geq n_k \;\Rightarrow\; P\{|Y_n| > \varepsilon\} \;\leq\; P\{|Y_{n_k}| > \varepsilon\} \;=\; \dfrac{1}{k}$. Since $\big[\, n \to \infty \,\big] \overset{\textbf{(j)}}{\Rightarrow} \big[\, k \to \infty \,\big]$, it follows that $\lim\limits_{n\to\infty} P\{|Y_n| > \varepsilon\} \,=\, 0$ and this shows that $Y_n \overset{P}{\to} 0$. ■

□

4.4 Stochastic Processes and Filtrations

In finance and other disciplines we are interested in undertanding random evolutions in time, i.e., trajectories $t \mapsto X(t,\omega)$ which are thought of be random and thus are a function of randomness ω. Time may be discrete if we we observe $X(t,\omega)$ only at countably many discrete times $t = t_0 < t_1 < t_2 < \cdots$ or it may be continuous if we observe $X(t,\omega)$ for $t_0 \leq t \leq T$ or $t_0 \leq t < T$, where $0 \leq t_0 < T < \infty$. For example, $X(t,\omega)$ can the price of a stock at some future time t which is unknown to us, and ω captures that uncertainty.

Definition 4.16 (Stochastic Process)**.**

> A **stochastic process** X on a probability space $(\Omega, \mathfrak{F}, P)$, often just called a **process**, is a collection of random elements X_t which take values $X_t(\omega)$ in a measurable space $(\Omega', \mathfrak{F}')$, the **state space**, of the process.

Being a random element, each X_t is $\mathfrak{F}$–$\mathfrak{F}'$ measurable.

The argument t takes values in an interval of the form $[\,t_0, T\,]$ or $[t_0, T[$ or $[\,t_0, \infty[$ or in a discrete collection $\{t_0 < t_1 < t_2 < \dots\}$, finite or infinite, of real numbers. We interpret t as time. Usually the start time t_0 will be zero and the end time T, if it is given, will be the time of expiration of one or several financial instruments.

> Unless something different is specified, the symbol I will denote the index set of the stochastic process X.

Depending on what is convenient we will include or omit the randomness argument ω, and the same applies to the index t. Here is an incomplete list of the notation you will encounter for a stochastic process.

$$X \;=\; X_t \;=\; X(t) \;=\; (X_t)_t \;=\; \big(X(t)\big)_{t_0 \leq t \leq T} \;=\; X_t(\omega) \;=\; X(t,\omega) \;=\; \dots$$

Unless stated otherwise, we assume that X is numeric, i.e., $X_t(\omega)$ is an extended real number for each randomness argument ω and time t. Thus each random element X_t actually is a (extended real–valued) random variable. Note that we also deal with vector valued stochastic processes

$$\vec{X} \;=\; \vec{X}_t \;=\; [X_t^{(1)}, X_t^{(2)}, \dots, X_t^{(m)}].$$

We sometimes use the notation $X(\cdot, \omega)$ if we want to emphasize that we consider the randomness ω as fixed and only t varies. We call this function $X(\cdot, \omega) : t \mapsto X(t, \omega)$ the ω–**trajectory** or ω–**path** or, in short, the **trajectory** or **path** of X.

At other times we write $X(t, \cdot)$ or $X_t(\cdot)$ if we want to emphasize X as the random variable which results when we look at the potential outcomes at a fixed time t. $\square$

We introduce some more terminology for random elements indexed by time which do not qualify as stochastic processes in the sense of Definition 4.16 (Stochastic Process) on p.73 because the time index does not live in a contiguous interval.

Definition 4.17. Given are a probability space $(\Omega, \mathfrak{F}, P)$, a measurable space $(\Omega', \mathfrak{F}')$, an index set $I \subset [0, \infty[$, and a family $X = (X_t, t \in I)$, of Ω'–valued random elements with index set I. We further assume that the indices $t \in I$ are to be interpreted as points in time.

- **(a)** If I is a contiguous interval of the form $[\,t_0, T\,]$ or $[t_0, T[$ or $[\,t_0, \infty[$ $(t_0 \geq 0)$, then we refer to X as a **continuous time stochastic process** with start time t_0 and, in the first case, with end time or expiration time T.
- **(b)** If I is an infinite sequence of real numbers $0 \leq t_0 < t_1 < \cdots$ or a finite sequence of real numbers $0 \leq t_0 < t_1 \leq t_n = T$, we call X a **discrete time stochastic process** with start time t_0 and, in the second case, with end time or **expiration time** T.
- **(c)** If I is an infinite, contiguous sequence of integers $0 \leq k_0, k_0 + 1, k_0 + 2, \dots$ then we call X a **stochastic sequence**. with start time k_0. This is a special case of a discrete time stochastic process.

- **(d)** If the index set of the form $I = 1, 2, \dots d$ and we interpret $X_1, \dots X_d$ as the coordinate values of a d–tuple rather than the values of a real–valued process observed at the times $1, 2, \dots d$, then we prefer to write

$$\vec{X} = (X^{(1)}, \dots X^{(d)}) \quad \text{or} \quad \vec{X}(\omega) = \big(X^{(1)}(\omega), \dots X^{(d)}(\omega)\big)$$

  and call this expression a (d–dimensional) **random vector**.
- **(e)** If the set Ω' of **(a)** – **(c)** satiksfies $\Omega' \subseteq \mathbb{R}$, we speak of a **real–valued** stochastic process or stochastic sequence. If the $\Omega' \subseteq \mathbb{R}^d$, we speak of a **vector–valued** stochastic process or stochastic sequence. $\square$

Remark 4.15. Any nonnegative finite or infinite sequence of real numbers $t_0 < t_1 < \cdots$ is a suitable index set for a discrete time stochastic process. Thus stochastic sequences and random vectors are special cases of such processes.

We will almost exclusively deal with stochastic processes which are either of
- continuous time stochastic processes,
- discrete time stochastic processes. $\square$

Before we can proceed we must discuss the information associated with a stochastic process. We briefly touched upon a σ–algebra as the information belonging to a random variable in Remark 4.14**(3)** on p.67. We recall Proposition 4.13 in which we defined $\sigma(f) := \{f^{-1}(A') : A' \in \mathfrak{F}'\}$, the σ–algebra generated by f, for any function $f : \Omega \to \Omega'$ from an arbitrary, nonempty set Ω to a measurable space $(\Omega', \mathfrak{F}')$.

We can generalize this notion to more than one function as long as they all have the same domain Ω. So let $g : \Omega \to \Omega''$ also have a codomain which is a measurable space $(\Omega'', \mathfrak{F}'')$. we then define

$$\sigma(f,g) := \sigma\{A \subseteq \Omega : A = f^{-1}(A') \text{ for some } A' \in \mathfrak{F}' \text{ or } A = g^{-1}(A'') \text{ for some } A'' \in \mathfrak{F}''\},$$

i.e., $\sigma(f,g)$ is the smallest σ–algebra that contains all preimages of measurable events for both f and g. This definition easily scales for any finite or infinite, even uncountable, collection of functions $f_i : \Omega \to (\Omega_i, \mathfrak{F}_i)$ which have measurable spaces as codomains.

Definition 4.18.

Let Ω be an arbitrary, nonempty set and let $f_i : \Omega \to \Omega_i,\ i \in I$ be a family of functions which have measurable spaces $(\Omega_i, \mathfrak{F}_i)$ as codomains and are indexed by an arbitrary, nonempty, index set I. No assumptions are made about I so do not think of those functions f_i as being indexed by "time"! We call the σ–algebra

(4.56) $$\sigma(f_i : i \in I) := \sigma\{\, A \subseteq \Omega \,:\, A = f_i^{-1}(A_i) \text{ for some } i \in I \text{ and } A_i \in \mathfrak{F}_i \,\}$$

the **σ–Algebra generated by the family of functions** f_i □

Remark 4.16. This last definition can be applied to the special case of a collection of random elements $X_i, i \in I$ on a probability space $(\Omega, \mathfrak{F}, P)$, indexed again by an arbitrary index set I. Thus each $X_i(\omega)$ is an element of a measurable spaces $(\Omega_i, \mathfrak{F}_i)$. We then have

(4.57) $$\sigma(X_i : i \in I) = \sigma\big\{\, A \subseteq \Omega \,:\, A = \{X_i \in A_i\} \text{ for some } i \in I \text{ and } A_i \in \mathfrak{F}_i \,\big\}.$$

Note that since each X_i is a random element, each preimage $\{X_i \in A_i\}$ belongs to $\mathfrak{F}$, thus

$$\sigma(X_i : i \in I) \subseteq \mathfrak{F}. \ \square$$

We are now back to stochastic processes and index sets I which can be interpreted as time intervals.

As we just have seen in Remark 4.16 we can associate with each random element X_t of a stochastic process $X = \big(X_u\big)_{u \in I}$ the σ–algebra $\sigma(X_t)$, which we interpret as the stochastically relevant information of X_t. See Remark 4.14 on p.67. However, we are not only interested in the stochastically relevant information of X_t, but in that of the entire past of the process X up to time t. Since this information is stored in $\sigma\{X_s : s \le t\}$, we are lead to the definition of a filtration.

Definition 4.19 (Filtration for a process X_t)**.**

For a continuous time or discrete time stochastic process X with index set I we define, for $t \in I$,

(4.58) $$\mathfrak{F}_t^X := \sigma\{X_s : s \in I, s \leq t\}$$

We call the family $(\mathfrak{F}_t^X)_{t \in I}$ of all those sub–σ–algebras of $\mathfrak{F}$ the **filtration generated by** X. □

Remark 4.17. For the following see also Remark 4.14 on p.67.

The σ–algebra $\mathfrak{F}_t^X$ associated with a stochastic process $(X_s)_{s \in I}$ is, in a sense to be made more precise in Chapter 5.1 (Functional Dependency of Random Variables), the container of all stochastically relevant information of this process up to time t. □

The next example shows you how to interpret the previous remark. It is very important that you understand it intuitively, without trying to apply any mathematical reasoning.

Example 4.9 (Filtrations as seat of the information of the past)**.** In the following we assume that X is real–valued and $I = [0, \infty[$.

(1) Let $A = \{2.78 < X_s \leq 3.14,\ \text{for } 5 \leq s < 7\}$. Then $A \in \mathfrak{F}_7^X$, but not $A \in \mathfrak{F}_{6.999}^X$, since observing the process X_s up to time $t = 6.999$ and seing that $2.78 < X_s \leq 3.14$ for $5 \leq s \leq 6.999$ does not determine whether or not $2.78 < X_7 \leq 3.14$.

(2) For some arbitrary $t, h > 0$. Let $B = \{X_{t+h} < 0\}$. Then $B \in \mathfrak{F}_{t+h}^X$. but not $B \in \mathfrak{F}_t^X$, since one cannot decide whether or not B has occurred just from knowing how X behaved up to and including time t.

(3) Assume that X has continuous trajectories $s \mapsto X_s(\omega)$ Then $Z(\omega) = \int\limits_0^T X_u(\omega) du$ (Riemann integral) is defined for any given $T > 0$ and $\omega \in \Omega$. Z is $\mathfrak{F}_T^X$–measurable since knowing the behavior of the trajectory $X(\cdot, \omega)$ between times 0 and T suffices to understand the behavior of $\int\limits_0^T X_u(\omega) du$. But note that $Z \notin m(\mathfrak{F}_{T-\delta}^X)$ for any $\delta > 0$, no matter how small.

(4) Assume that X has continuous trajectories $s \mapsto X_s(\omega)$. Let

$$\tau(\omega) := \inf\{s \geq 0 : X_s(\omega) \geq 20\},$$

i.e., the random time τ denotes the first time that the trajectory enters the interval $[20, \infty[$. Then the event $\{\tau \leq 8.5\}$ is in $\mathfrak{F}_{8.5}$, since

$$\tau(\omega) \leq 8.5 \Leftrightarrow X_s(\omega) \geq 20 \text{ for some } s \leq 8.5,$$

and this clearly is determined by the behavior of $X_s(\omega)$ for $0 \leq s \leq 8.5$.

(4a) More generally assume again that X has continuous trajectories. Let γ be an arbitrary real number. Let

$$\tau(\omega) := \inf\{s \geq 0 : X_s(\omega) \geq \gamma\}$$

be the time of first entry into $[\gamma, \infty[$. Then $\{\tau \leq t\}$ is in $\mathfrak{F}_t$ for any $t > 0$, since

$$\tau(\omega) \leq t \Leftrightarrow X_s(\omega) \geq \gamma \text{ for some } s \leq t.$$

(5) Assume that X has continuous trajectories $s \mapsto X_s(\omega)$ and let

$$\rho(\omega) := \sup\{s \geq 0 : X_s(\omega) \geq 20\},$$

i.e., the random time ρ denotes the last time that the trajectory is inside the interval $[20, \infty[$. Then the event $\{\rho \leq t\}$ is not in $\mathfrak{F}_t$ for any $t > 0$ since we cannot predict at time t the future behavior of the trajectory. □

Remark 4.18. It is obvious that, for a time t after time s, more info (more measurable preimages) has accrued until time t than just until the time s of the past. In other words,

$$\text{if } s < t \quad \text{then } \mathfrak{F}_s^X \subseteq \mathfrak{F}_t^X. \ \square$$

The property just mentioned by itself is so useful that we encapsulate it in its own definition, without referring to stochastic processes.

Definition 4.20 (Filtration-general)**.**

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and $I \subseteq \mathbb{R}$. Assume that for each $t \in I$ there is a sub–σ–algebra $\mathfrak{F}_t$ of $\mathfrak{F}$ and that this family $(\mathfrak{F}_t)_{t \in I}$ satisfies monotony with respect to t:

$$\text{If } s < t \quad \text{then } \mathfrak{F}_s \subseteq \mathfrak{F}_t$$

for all $s, t \in I$. We call such a family a **filtration** on $(\Omega, \mathfrak{F}, P)$, and we call the quadruple $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t \in I}, P)$, usually denoted by $(\Omega, \mathfrak{F}, (\mathfrak{F}_t), P)$ or $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ if there is no confusion about I or its particulars are irrelevant for the discussion at hand, a **filtered probability space**. □

We have a special definition for a processes $X = (X_t)_{t \in I}$ if its trajectories $X_s, s \in I, s \leq t$ are determined by the member $\mathfrak{F}_t$ of a filtration $(\mathfrak{F}_t)_{t \in I}$.

Definition 4.21 (Adapted Process)**.**

Let X be a discrete time or continuous time process with index set I on a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t \in I}, P)$. If the trajectory $X(s)\,(s \in I, s \leq t)$, is determined by the information in $\mathfrak{F}_t$ for each time t, i.e., if

$$X_s \text{ is } \mathfrak{F}_t\text{–measurable} \quad \text{for each } s \in I \text{ such that } s \leq t,$$

then we say that X is **adapted to the filtration** $\mathfrak{F}_t$. □

Proposition 4.14.

Every process X_t is adapted to its own filtration $\mathfrak{F}_t^X = \sigma\{X_s : s \in I, s \leq t\}$.

PROOF:
Let $t \in I$. To prove that X_t is $\mathfrak{F}_t^X$–$\mathfrak{F}'$ measurable, we claim that it suffices to show that

(A) $$\{ X_t \in B \} \in \mathfrak{F}_t^X \qquad \text{for all } B \in \mathfrak{F}' .$$

This is why. Let $s < t$ and $B \in \mathfrak{F}'$. Then, by **(A)**, $\{X_s \in B\} \in \mathfrak{F}_s^X$. But $s < t \Rightarrow \mathfrak{F}_s^X \subseteq \mathfrak{F}_t^X$, thus $\{X_s \in B\} \in \mathfrak{F}_t^X$, thus X is $\left(\mathfrak{F}_t^X\right)_t$–adapted.
Let
$$\mathfrak{E}_1 := \{X_t^{-1}(B) : B \in \mathfrak{F}'\}.$$
and
$$\mathfrak{E}_2 := \{A \subseteq \Omega : A = X_u^{-1}(B) \text{ for some } B \in \mathfrak{F}' \text{ and some } u \leq t\}.$$

Then $\mathfrak{E}_1 \subseteq \mathfrak{E}_2$, $\sigma(\mathfrak{E}_1) = \sigma(X_t)$, and $\mathfrak{E}_2 = \mathfrak{F}_t^X$. It follows from Proposition 4.3 (Monotonicity of generated σ–algebras) on p.51 that $\sigma(\mathfrak{E}_1) \subseteq \sigma(\mathfrak{E}_2)$, i.e., $\sigma(X_t) \subseteq \mathfrak{F}_t^X$. Thus, **(A)** holds. ■

If a random variable $\omega \mapsto \tau(\omega)$ is nonnegative then one can interpret τ as a **random time** It can be used, e.g., as the time argument of a stochastic process $\left(X_t\right)_{t \geq 0}$. The resulting random variable

$$\omega \mapsto X_{\tau(\omega)}(\omega)$$

then denotes the value of the ω–trajectory $X(\cdot, \omega)$ at time $\tau(\omega)$.
We will now use special random times, called stopping times, to create adapted processes.

Definition 4.22 (Stopping time). ⋆

We call a random time tau on a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_t)$ a **stopping time** if

(4.59) $$\{\tau \leq t\} = \{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathfrak{F}_t \qquad \text{for all } t \in [0, \infty[.$$

Proposition 4.15. ⋆

If τ is a random time on a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_t)$ then
$$\tau \text{ is a stopping time} \quad \Leftrightarrow \quad \text{the process } (t, \omega) \mapsto X(t, \omega) := \mathbf{1}_{[0,\tau(\omega)[}(t) \text{ is } \mathfrak{F}_t\text{–adapted}.$$

PROOF: We note that

(A) $$X_t(\omega) := \begin{cases} 1 & \text{if } \tau(\omega) > t, \\ 0 & \text{if } \tau(\omega) \leq t. \end{cases}$$

Let $c \in \mathbb{R}$. Then the value of the set $\{X_t < c\}$ only depends on whether either $c \leq 0$ or $0 < c \leq 1$ or $c > 1$. We obtain from **(A)** the following.

Case $c \leq 0$: $\{X_t < c\} = \emptyset$,
Case $c > 1$: $\{X_t < c\} = \Omega$,
Case $0 < c \leq 1$: $\{X_t < c\} = \{X_t = 0\} = \{\tau \leq t\}$.

Since the empty set and Ω belong to any σ–algebra of Ω the $\mathfrak{F}_t$–adaptedness of X_t is entirely determined by the last case $0 < c \leq 1$ as follows:

$$X_t \text{ is } \mathfrak{F}_t\text{–adapted} \quad \Leftrightarrow \quad \{X_t = 0\} \in \mathfrak{F}_t \text{ for all } t \quad \Leftrightarrow \quad \{\tau \leq t\} \in \mathfrak{F}_t \text{ for all } t \quad \Leftrightarrow \quad \tau \text{ is a stopping time}.$$

This concludes the proof. ■

Remark 4.19. In a financial market filtrations appear, e.g., as follows. Given are one or more "underlying assets", e.g., stocks, whose prices $S^{(1)}, \dots, S^{(n)}$ depend on time t and randomness ω, i.e., each stock price $S^{(j)}$ is a stochastic process $S_t^{(j)}(\omega)$. They will be "driven", i.e., stochastically determined, by one or more processes $W_t^{(1)}, \dots, W_t^{(m)}$. [21] By this we mean that each stock price $S^{(j)}$ is adapted to the filtration defined by

$$\mathfrak{F}_t := \sigma\big(W_s^{(j)} : 1 \leq j \leq m, s \leq t, s \in I\big) \text{ for each } t \in I,$$

i.e., to the filtration generated by those $W_t^{(j)}$. Optimal estimates of future financial data with respect to this fitration will play a key role in determining the price of a financial derivative which is based on the underlying assets. Those optimal estimates are obtained by means of conditional expectations, a tool that will be discussed in Chapter 5. □

4.5 Integration and Expectations

The following should be read in conjunction with SCF2 ch.1.3: Expectations.

Remark 4.20. We recall that **(1)** if $f : \mathbb{R} \to \{0,1\}$ and $g : \mathbb{R}^n \to \{0,1\}$ are Riemann–integrable and **(2)** if also the sets $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}^n$ are Riemann–integrable, i.e., the Riemann integrals

$$\int_{-\infty}^{\infty} \mathbf{1}_A(x)\,dx \qquad \text{and} \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{1}_B(x_1, x_2, \dots, x_n)\,dx_1 dx_2 \cdots dx_n$$

of the indicator functions $\mathbf{1}_A : \mathbb{R} \to \{0,1\}$ and $\mathbf{1}_B : \mathbb{R}^n \to \{0,1\}$ exist, then we write

$$\int_A f(x)\,dx = \int_{-\infty}^{\infty} f(x)\mathbf{1}_A(x)\,dx, \tag{4.60}$$

$$\int_B g(x_1, \dots, x_n)\,dx_1 \cdots dx_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n)\mathbf{1}_B(x_1, \dots, x_n)\,dx_1 \cdots dx_n. \tag{4.61}$$ □

Introduction 4.2. We start out with a few things we know about integration from calculus.

A. If $f : \mathbb{R} \to \mathbb{R}$ is a function of the form

$$f(x) = \sum_{j=1}^{k} c_j \, \mathbf{1}_{]a_j, b_j]}(x),$$

[21] so-called Brownian motions or Wiener processes

then

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx \;&=\; \sum_{j=1}^{k} c_j \int_{-\infty}^{\infty} \mathbf{1}_{]a_j,b_j]}(x) \;=\; \sum_{j=1}^{k} c_j \int_{a_j}^{b_j} dx\\ &=\; \sum_{j=1}^{k} c_j (b_j \,-\, a_j) \;=\; \sum_{j=1}^{k} c_j\, \lambda^1(]a_j,b_j])\,.\end{aligned} \tag{4.62}$$

Here λ_1 denotes Lebesgue measure which was introduced in Definition 4.7 on p.54.

B. Things are similar in the multidimensional case. If $g : \mathbb{R}^n \to \mathbb{R}$ has the form

$$g(\vec{x}) \;=\; \sum_{j=1}^{k} c_j\, \mathbf{1}_{]u_{1j},v_{1j}]\times\cdots\times]u_{nj},v_{nj}]}(\vec{x}),\;\; (u_{ij} < v_{ij} \text{ for } i = 1,\dots,n),$$

where $\vec{x} = (x_1, x_2, \dots, x_n)$, then

$$\begin{aligned}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\dots,x_n)\,dx_1\cdots dx_n \;&=\; \sum_{j=1}^{k} c_j \int_{u_{1j}}^{v_{1j}}\cdots\int_{u_{nj}}^{v_{nj}} dx_1\cdots dx_n\\ &=\; \sum_{j=1}^{k} c_j (v_{1j} \,-\, u_{1j})\cdots(v_{nj} \,-\, u_{nj})\\ &=\; \sum_{j=1}^{k} c_j\, \lambda^n(]u_{1j},v_{1j}]\times\cdots\times]u_{nj},v_{nj}])\,.\end{aligned} \tag{4.63}$$

C. If X is a random variable on the probability space $(\Omega, \mathfrak{F}, P)$ and if $f : \mathbb{R} \to \mathbb{R}$ is of the form

$$f(x) \;=\; \sum_{j=1}^{k} c_j\, \mathbf{1}_{]a_j,b_j]}(x),\;\; (k \in \mathbb{N}),$$

then the expected value $E[f\circ X]$ of the composite function $f\circ X : \omega \mapsto f\big(X(\omega)\big)$ is

$$E[f\circ X] \;=\; \sum_{j=1}^{k} c_j E\big[\mathbf{1}_{]a_j,b_j]}(X)\big] \;=\; \sum_{j=1}^{k} c_j P\{X \in]a_j,b_j]\} \;=\; \sum_{j=1}^{k} c_j P_X(]a_j,b_j]). \tag{4.64}$$

Here P_X is the distribution of X, i.e., the image of P under X.

In each of those three cases we have a function of the form $f \;=\; \sum\limits_{j=1}^{k} c_j \mathbf{1}_{A_j}$ which takes finitely many values c_j, and we have computed in each case an integral or an expected value of the form $\sum\limits_{j=1}^{k} c_j \mu(A_j)$ for a suitable measure μ. We will now establish a common thread. □

Definition 4.23 (Integral of a simple function)**.**

Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space, $n \in \mathbb{N}$, and $A_1, A_2, \dots, A_n \in \mathfrak{F}$ a finite collection of measurable sets. Let $f : \Omega \to \mathbb{R}$ be defined as

$$f \;:=\; \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}, \quad 0 \le c_j < \infty \text{ for } j = 1, \dots, n. \tag{4.65}$$

We call such a function a **simple function**.

Note that $f \ge 0$ and f is measurable as the sum of the measurable functions $\omega \mapsto c_j \cdot \mathbf{1}_{A_j}(\omega)$. We call

$$\int f \, d\mu \;:=\; \int f(\omega) \, d\mu(\omega) \;:=\; \int f(\omega) \, \mu(d\omega) \;:=\; \sum_{j=1}^{n} c_j \mu(A_j). \tag{4.66}$$

the **integral** aka **abstract integral** of f with respect to μ, also the μ**–integral** of f. □

Remark 4.21. ⋆

A. We made no assumption about finiteness of μ, so some or all of the A_j may have infinite measure. We confined ourselves to non-negative c_j in order to avoid expressions of the form $\infty - \infty$.

B. Note that the choice of k, A_j, and c_j is not unique for a given function f. For example, the constant function

$$f : (\mathbb{R}, \mathfrak{B}^1, \lambda^1) \longrightarrow \mathbb{R}; \quad x \mapsto 3,$$

can be written as

$$\begin{aligned} f \;&= 3 \cdot \mathbf{1}_{\mathbb{R}} \;=\; 3 \cdot \mathbf{1}_{]-\infty,0[} \;+\; 3 \cdot \mathbf{1}_{[0,\infty[} \\ &= 1 \cdot \mathbf{1}_{]-\infty,-1[} \;+\; 2 \cdot \mathbf{1}_{]-\infty,1[} \;+\; 1 \cdot \mathbf{1}_{]-1,\infty[} \;+\; 2 \cdot \mathbf{1}_{[1,\infty[}. \end{aligned}$$

Thus the following is important since it ensures that the definition of $\int f d\mu$ is consistent:

C. Let the simple, nonnegative, function f have representations

$$f \;:=\; \sum_{j=1}^{k} c_j \mathbf{1}_{A_j} \;=\; \sum_{j=1}^{k'} c'_j \mathbf{1}_{A'_j}.$$

Then $\sum\limits_{j=1}^{k} c_j \mu(A_j) \;=\; \sum\limits_{j=1}^{k'} c'_j \mu(A'_j)$, thus $\int f d\mu$ does not depend on the choice of the sets A_j and the coefficients c_j. □

We extend the definition of $\int f d\mu$ to more general measurable functions, in particular all $f \in m(\mathfrak{F})$ which are nonnegative or nonpositive.

For the following review the decomposition $f = f^+ - f^-$ given in Definition 2.17 (Absolute value, positive and negative part) on p.20.

Definition 4.24 (Integral of a measurable function)**.** Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and f an extended real–valued, $\mathfrak{F}$–measurable, function.

(1) If $f \geq 0$, we define

$$\int f\,d\mu \;:=\; \sup\left\{\int h\,d\mu \;:\; h \text{ is simple and } 0 \leq h \leq f\right\}. \tag{4.67}$$

If not both $\int f^+ d\mu = \infty$ and $\int f^- d\mu = \infty$, we define

$$\int f\,d\mu \;:=\; \int f^+\,d\mu \;-\; \int f^-\,d\mu. \tag{4.68}$$

Again, we call $\int f d\mu$ the **integral** aka **abstract integral** of f with respect to μ.

(2) If $\int |f|\,d\mu \;<\; \infty$ we call f **integrable** with respect to μ or just μ**–integrable**.

As in (4.66) on p.81, we have the following alternate notation.

$$\int f\,d\mu \;=\; \int f(\omega)\,d\mu(\omega) \;=\; \int f(\omega)\,\mu(d\omega). \;\square$$

Remark 4.22. ⋆ Note that there are measurable functions f which are not μ–integrable even though $\int f d\mu$ exists. For example, let

$$f : (\mathbb{R}, \mathfrak{B}^1, \lambda^1) \longrightarrow (\mathbb{R}, \mathfrak{B}^1); \qquad f(x) \;:=\; x^+ \;=\; x\,\mathbf{1}_{[0,\infty[}.$$

Here is a formal proof that $\int x^+ d\lambda^1(x) = \infty$. For each $n \in \mathbb{N}$, let $h_n := n \cdot \mathbf{1}_{[n,2n]}$. Then $h_n \leq f$ and this simple function has integral $\int h_n d\lambda = n \cdot \lambda^1([n, 2n]) = n^2$. Thus

$$\int x^+\,d\lambda^1 \;=\; \sup\left\{\int h\,d\lambda^1 \;:\; h \text{ is simple and } 0 \leq h \leq x^+\right\} \;\geq\; \sup_{n\in\mathbb{N}}\left\{\int h_n\,d\lambda^1\right\} \;=\; \infty.$$

In particular the integral $\int x^+ d\lambda^1$ exists but is infinite. Since $|f(x)| \;=\; f(x)$ for all x we see that $\int |f| d\lambda^1 = \infty$, thus f is not λ^1-integrable. $\square$

We next define expected values of random variables as abstract integrals $\int \cdots dP$.

Definition 4.25 (Expected value of a random variable)**.** Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X a random variable on that space, possibly extended real–valued.

If $\int X dP$ exists, we define the **expectation** or **expected value** $E[X]$ of X, with respect to P, also simply written as EX, as

$$E[X] \;:=\; \int X\,dP \;=\; \int X(\omega)\,dP(\omega) \;=\; \int X(\omega)\,P(d\omega). \;\square \tag{4.69}$$

Definition 4.26. (p–integrable functions and random variables)

(1) Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and f an extended real–valued, $\mathfrak{F}$–measurable, function. Let $p \geq 1$. If $\int |f|^p \, d\mu < \infty$ we call f p**–integrable** with respect to μ.

(2) Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X a random variable on that space, possibly extended real–valued. Let $p \geq 1$. If $E\big[|X|^p\big] < \infty$ we call X a p**–integrable random variable**

(3) If $p = 2$ we also refer to **square–integrable** functions and random variables

Note that X is a p–integrable random variable if and only if X is a p–integrable function with respect to the (probability) measure P.

Proposition 4.16. $\star$

Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and $A \in \mathfrak{F}$. Let μ_A be the measure defined in Proposition 4.8 on p.59:

$$\mu_A(A') = \mu(A \cap A')$$

If $f \in m(\mathfrak{F})$ is μ–integrable then $f\mathbf{1}_A$ is integrable with respect to both μ and μ_A, and then

$$\int f\mathbf{1}_A \, d\mu = \int f\mathbf{1}_A \, d\mu_A = \int f \, d\mu_A.$$

PROOF: Not entirely trivial. You first prove this for simple functions h and then use

$$0 \leq h \leq f \Leftrightarrow 0 \leq h\mathbf{1}_A \leq f\mathbf{1}_A$$

to prove the general case. ■

The last proposition shows that if f is μ–integrable and $A \in \mathfrak{F}$ then $\int f\mathbf{1}_A d\mu$ exists. We are in a position to define the following.

Definition 4.27. Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space, $A \in \mathfrak{F}$.

If f is a measurable function and $\int f\mathbf{1}_A d\mu$ exists (is not of the form $\infty - \infty$) then we call

$$\int_A f \, d\mu := \int f \cdot \mathbf{1}_A \, d\mu \tag{4.70}$$

the **integral** or **abstract integral**, of f over A with respect to μ. We also write

$$\int_A f \, d\mu = \int_A f(\omega) \, d\mu(\omega) = \int_A f(\omega) \, \mu(d\omega).$$

Observe that $\int_\Omega f du = \int f du$. □

For the following see SCF2 Theorem 1.3.4. We formulate it twice, once for general measures and then again for probability spaces.

Theorem 4.6 (Fundamental properties of the abstract integral)**.**

Let f be a measurable function on a emasure space $(\Omega, \mathfrak{F}, \mu)$.

a. *If f takes only finitely many distinct function values $x_0, x_1, \dots, x_n$, then*

$$\int f\,d\mu \;=\; \sum_{k=0}^{n} x_k\,\mu\left(f^{-1}\{x_k\}\right).$$

Also, if Ω is finite and $\mathfrak{F} = 2^\Omega$, then

$$\int f\,d\mu \;=\; \sum_{\omega\in\Omega} f(\omega)\,\mu\{\omega\}.$$

b. (***Integrability***) *The measurable function f is integrable if and only if*

$$\int f^+\,d\mu \;<\; \infty \quad \textit{and} \quad \int f^-\,d\mu \;<\; \infty.$$

Let g be another measurable function on $(\Omega, \mathfrak{F}, \mu)$.

c. (***Comparison***) *If $f = g$ a.e. and f and g are integrable or nonnegative a.e., then*

$$\int f\,d\mu \;=\; \int g\,d\mu.$$

d. (***Linearity***) *If α and β are real constants and f and g are integrable or if α and β are nonnegative constants and f and g are nonnegative, then*

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

PROOF: See SCF2, proof of Theorems 1.3.1 and 1.3.4. ■

And this is the version for probability spaces which you will find as SCF2 Theorem 1.3.4.

Theorem 4.7.

Let X be a random variable on a probability space $(\Omega, \mathfrak{F}, P)$.

a. *If X takes only finitely many distinct values $x_0, x_1, \dots, x_n$, then*

$$E(X) \;=\; \sum_{k=0}^{n} x_k\,P\{X = x_k\}.$$

Also, if Ω is finite and $\Omega = 2^\Omega$, then

$$E(X) \;=\; \sum_{\omega\in\Omega} X(\omega)\,P\{\omega\}.$$

b. (***Integrability***) *The random variable X is integrable if and only if*

$$E[X^+] \;<\; \infty \quad \textit{and} \quad E[X^-] \;<\; \infty$$

Now, let Y be another random variable on $(\Omega, \mathfrak{F}, P)$.

c. *(**Comparison**) If $X = Y$ a.s. and X and Y are integrable or a.s. nonnegative, then*

$$E\,X \;=\; E\,Y.$$

d. *(**Linearity**) If α and β are real constants and X and Y are integrable or if α and β are nonnegative constants and X and Y are nonnegative, then*

$$E(\alpha X + \beta Y) \;=\; \alpha E(X) + \beta E(Y).$$

e. *(**Jensen's inequality:**) The following **need NOT be true** for measures which are not probability measures. If φ is a convex, real–valued function defined on $\mathbb{R}$ and if $E(X) < \infty$, then*

$$\varphi\big(E(X)\big) \;\le\; E\big(\varphi(X)\big).$$

PROOF: See SCF2. ■

Theorem 4.8. *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and assume that the extended real–valued functions $f, g \in m(\mathfrak{F}, \mathfrak{B})$ both are μ–integrable. We have the following.*

$$\text{(4.71)} \qquad \textit{If} \int_\Gamma f\,d\mu \;\le\; \int_\Gamma g\,d\mu \;\; \textit{for all } \Gamma \in \mathfrak{F} \qquad \textit{then} \quad f \;\le\; g \;\; \mu\text{–a.e.}$$

$$\text{(4.72)} \qquad \textit{If} \int_\Gamma f\,d\mu \;=\; \int_\Gamma g\,d\mu \;\; \textit{for all } \Gamma \in \mathfrak{F} \qquad \textit{then} \quad f \;=\; g \;\; \mu\text{–a.e.}$$

PROOF: ⋆

Proof of (4.71): We assume that $\int_\Gamma f d\mu \le \int_\Gamma g\,d\mu$ for all $\Gamma \in \mathfrak{F}$, and $f \le g$ μ–a.e.. Let $A := \{f > g\}$.

Let $A := \{f > g\}$ and assume that $\mu(A) > 0$. It suffices to show that

$$\textbf{(A)} \qquad \text{there exists } \Gamma \in \mathfrak{F} \text{ such that } \int_\Gamma f\,d\mu \;>\; \int_\Gamma g\,d\mu\,,$$

since this contradicts the assumptions made in (4.71). This allows us to conclude that the assumption $\mu(A) > 0$ is wrong, since it lead to that contradiction. Thus, $\mu\big(\{f > g\}\big) = 0$. This proves that $f \;\le\; g$, μ–a.e., and we are done.

It remains to prove **(A)** by finding $\Gamma \in \mathfrak{F}$ such that $\int_\Gamma f d\mu > \int_\Gamma g d\mu$.

For $n \in \mathbb{N}$ let $A_n := \{f > g + \frac{1}{n}\}$. Then $A_n \uparrow A$, hence $\mu(A_n) \uparrow \mu(A)$. See Proposition 4.7 (Continuity properties of measures) on p.58.

Assume to the contrary that $\mu(A) > 0$. Then there exists $\gamma > 0$ such that $\mu(A) = 2\gamma$ and hence some

$n \in \mathbb{N}$ such that $\mu(A_n) \geq \gamma$. Since $f > g + \frac{1}{n}$ on all of A_n,

$$\int_{A_n} f\,d\mu \geq \int_{A_n} \left(g + \frac{1}{n}\right) d\mu = \int_{A_n} g\,d\mu + \frac{1}{n}\mu(A_n) \geq \int_{A_n} g\,d\mu + \frac{\gamma}{n} > \int_{A_n} g\,d\mu.$$

In other words, $\Gamma := A_n$ satisfies **(A)**. This concludes the proof of (4.71).

Proof of (4.72): Note that, according to the already proven validity of (4.71), the assumption

$$\int_A f\,d\mu = \int_A g\,d\mu \text{ for all } A \in \mathfrak{F} \quad \text{implies} \quad f \leq g \ \mu\text{–a.e., and } g \leq f \ \mu\text{–a.e.}$$

This proves $f = g$ μ–a.e. ■

The following theorem, [SCF2 Theorem 1.3.8, is specific to Lebesque measure. It is true in multiple dimensions, but we only state it for the one dimensional case.

Theorem 4.9. *Connection between Riemann and Lebesgue integrals] Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function and let $a < b$.*

(1) *The Riemann integral $\int\limits_a^b f(x)\,dx$ exists (i.e., the lower and upper Riemann sums converge to the same limit) $\Leftrightarrow$ the set of points x in $[a,b]$ where $f(x)$ is not continuous has Lebesgue measure zero.*

(2) *If the Riemann integral $\int\limits_a^b f(x)\,dx$ exists, then f is Borel–measurable (so the Lebesgue integral $\int\limits_{[a,b]} f(x)\,d\lambda^1(x)$ also exists), and both integrals agree.*

PROOF: ⋆ Beyond the scope of this course. ■

Remark 4.23.

(1) Theorem 4.9**(1)** can be expressed as follows: The Riemann integral $\int_a^b f(x)\,dx$ exists $\Leftrightarrow$ $f(x)$ is almost everywhere continuous on $[a,b]$.

(2) All singleton sets $\{x\}$ in $\mathbb{R}$ have Lebesgue measure zero, hence any finite set of points has Lebesgue measure zero. Thus **(1)** above guarantees that if we have a real–valued function f on $\mathbb{R}$ that is continuous except at finitely many points, then there will be no difference between Riemann and Lebesgue integrals of this function.

(3) Lebesgue integrals are the appropriate vehicle to develop and prove mathematical theory. But to actually evaluate integrals we use the formulas for computing Riemann integrals.

(4) Because the Riemann and Lebesgue integrals agree whenever the Riemann integral is defined, we often use the familiar notation $\int_a^b f(x)\,dx$ instead of $\int_{[a,b]} f(x)\,d\lambda^1(x)$, even if we do Lebesgue integration.

(5) If the set B over which we integrate is Borel but not necessarily an interval, we also write $\int_B f(x)\,dx$ instead of $\int_B f(x)\,d\lambda^1(x)$. □

4.6 Convergence of Measurable Functions and Integrals

The following corresponds to SCF2 Chapter 1.4, but note that what is formulated in these lecture notes for arbitrary measure spaces $(\Omega, \mathfrak{F}, \mu)$ is developed there only for the measurable space $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \lambda^1)$.

We start by applying the definition of a.e. and a.s (almost everywhere and almost surely, see Definition 4.12 on p.62), to the convergence of functions. In this case the property of interest for a given $\omega \in \Omega$ is whether the sequence of numbers or extended real numbers $f_1(\omega), f_2(\omega), \dots$ has a limit.

For the next two definitions see SCF2 Definitions 1.4.1 and 1.4.3.

Definition 4.28 (Convergence almost everywhere)**.**

Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space, and $f_n, f : \Omega \to \mathbb{R}$ Borel–measurable functions ($n \in \mathbb{N}$). Let

$$A \;:=\; \{\omega \in \Omega : \lim_{n\to\infty} f_n(\omega) = f(\omega)\}.$$

If $\mu(A^{\complement}) = 0$, we say that **the sequence f_n has limit f μ–almost everywhere** , and we write

$$\lim_{n\to\infty} f_n \;= f \;\mu\text{–a.e.}\,, \qquad \text{or} \qquad f_n \to f \;\mu\text{–a.e. as } n \to \infty\,. \;\square$$

Definition 4.29 (Convergence almost surely)**.**

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X_n, X a sequence of random variables with domain Ω such that $\lim_{n\to\infty} X_n \;= X \; P$–a.e. as defined above. In the context of a probability space we prefer to say that **the sequence X_n has limit X P–almost surely** , and we write

$$\lim_{n\to\infty} X_n \;= X \; P\text{–a.s.} \qquad \text{or} \qquad X_n \to X \; P\text{–a.s. as } n \to \infty\,. \;\square$$

Definition 4.30 (iid random variables)**.**

A sequence of random variables $X_1, X_2, \dots$ is called **indedependent and identically distributed** aka **iid**, if it is a sequence of independent random variables and if all X_n have the same distribution, i.e.,

$$P\{X_1 \in B\} \;=\; P\{X_2 \in B\} \;=\; P\{X_3 \in B\} \;=\; \cdots \quad \text{holds true for all Borel sets } B.\; \square$$

The next theorem gives one of the most important examples of almost sure convergence.

Theorem 4.10 (Strong Law of Large Numbers)**.**

Let X_n be an iid sequence of integrable random variables, i.e., $E\big[|X_n|\big] < \infty$ for all n. Then,

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} \;=\; E\big[X_1\big] \quad a.s.$$

PROOF: See, e.g., [9] Dudley, Real Analysis and Probability. ■

There also is a less powerful version of the Law of Large Numbers which only asserts convergence in distribution

There also is a less powerful version of the Law of Large Numbers. It can be stated in two different ways. The version given in SCF2 replaces convergence a.s. with convergence in probability. In most other sources convergence a.s. is replaced with convergence in distribution. The next theorem combines both versions.

Theorem 4.11 (Weak Law of Large Numbers)**.**

Let X_n be an iid sequence of integrable random variables, i.e., $E\big[|X_n|\big] < \infty$ for all n. Then both,

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = E\big[X_1\big] \quad \textit{in probability,}$$
$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = E\big[X_1\big] \quad \textit{in distribution.}$$

PROOF: Can be found, for the version which states convergence in probability, in most undergraduate texts on probability theory. ■

Remark 4.24. Note that convergence in probability and convergence in distribution are equivalent under the assumptions of Theorem 4.11, because the limit is $E[X_1]$, a (deterministic) constant. See Theorem 4.3 (Relationship between the modes of convergence) on p.70. □

In the laws of large numbers the limit is deterministic because division by zero causes the standard deviations of the arithmetic averages $(X_1 + \cdots + X_n)/n$ to go to zero. To see this, note that the variance of a sum of independent random variables is the sum of the variances.

Thus, if $Var[X_j] = \sigma^2$, [22] and if $S_n = \sum\limits_{j=1}^{n} X_j$, then

$$Var\left[\frac{S_n}{n}\right] \;=\; \frac{1}{n^2}\sum_{j=1}^{n}\sigma^2 \;=\; \frac{\sigma^2}{n}\,.$$

Thus, the standard deviations $\sqrt{Var\left[S_n/n\right]} \;=\; \sigma/\sqrt{n}$ converge to zero. We have reason to assume that if we keep the standard deviations constant by dividing S_n/n by $\sigma/\sqrt{n}$, then there might be

[22]Since the X_j have identical distribution for each j, it is true that

$$E[X_1] \;=\; E[X_2] \;=\ldots, \qquad \text{and} \qquad Var[X_1] \;=\; Var[X_2] \;=\ldots\;=\;\sigma\,.$$

a non–deterministic limit. In addition, we center the expectations of S_n/n at zero by replacing X_j with $X_j - E[X_j]$, we obtain the well–known Central Limit Theorem.

Theorem 4.12 (Central Limit Theorem)**.**

Let X_n be an iid sequence of square–integrable random variables, i.e., $E\left[X_n^2\right] < \infty$ for all n. Let $\mathcal{N}(\alpha, \sigma^2)$ denote the normal distribution with mean α and variance σ^2. Then,

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}\sigma} \sum_{j=1}^{n} \left(X_j - E[X_j]\right) \quad \text{exists in distribution} \quad \text{and has a } \mathcal{N}(0,1) \text{ distribution.}$$

PROOF: Can be found in most undergraduate texts on Probability. ■

The following is SCF2 Example 1.4.4.

Example 4.10. Let $(\Omega, \mathfrak{F}, \mu) := (\mathbb{R}, \mathfrak{B}^1, \lambda^1)$ the real numbers with Lebesgue measure. Let $f_n : \mathbb{R} \to \mathbb{R}$ be the continuous and hence $(\mathfrak{B}^1, \mathfrak{B}^1)$–measurable functions

$$f_n(x) := \sqrt{\frac{n}{\sqrt{2\pi}}}\, e^{-\frac{nx^2}{2}} \qquad \text{(the density function of the } N(0,n)\text{–distribution)}, \tag{4.73}$$

$$f(x) := \begin{cases} 0 & \text{if } x \neq 0, \\ \infty & \text{if } x = 0. \end{cases} \tag{4.74}$$

Then $f_n(\omega) \to f(\omega)$ as $n \to \infty$ for all ω, thus $f_n \to 0$ λ^1–a.e., since $\lambda^1\{0\} = 0$. But observe $\int_{\mathbb{R}} f_n(x) d\lambda^1(x) = 1$ for all x whereas $\int_{\mathbb{R}} f(x) d\lambda^1(x) = 0$. What conditions are needed so this does not happen, in other words, what guarantees that we can switch $\int$ and $\lim_n$? □

Here is another example that shows that switching the order of integration and taking a limit may yield different results.

Example 4.11. Let $(\Omega, \mathfrak{F}, \mu) := (\mathbb{R}, \mathfrak{B}^1, \lambda^1)$ the real numbers with Lebesgue measure. Let $f_n : \mathbb{R} \to \mathbb{R}$ be defined as

$$f_n := \mathbf{1}_{[n,\infty[}, \quad n = 1, 2, 3, \dots, \qquad \text{i.e., } fn(x) = 1 \text{ for } x \geq n \text{ and zero else.} \tag{4.75}$$

Then each f_n is Borel measurable (why?) and $f_n(\omega) \to 0$ as $n \to \infty$. But the integrals $\int_{\mathbb{R}} f_n \, d\lambda^1$ do not converge to $\int_{\mathbb{R}} 0 \, d\lambda^1 = 0$ since each $\int_{\mathbb{R}} f_n \, d\lambda^1$ equals infinity. □

We have had two examples where a sequence of functions converges a.e., but the integrals do not converge to the integral of that limit function. We are now formulating conditions under which this cannot happen.

The following corresponds to SCF2 Theorem 1.4.5.

Theorem 4.13 (Monotone Convergence Theorem)**.**

(1) *Let* $(\Omega, \mathfrak{F}, \mu)$ *be a measure space and let* $f, f_1, f_2, \dots : \Omega \to \mathbb{R}$ *be* $m(\mathfrak{F}, \mathfrak{B})$*.*

$$\text{If } 0 \leq f_1 \leq f_2 \leq \dots \text{ a.e. and } \lim_{n\to\infty} f_n = f \text{ a.e., then } \lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

(2) *Let* X *and* $X_1, X_2, X_3, \dots$ *be random variables on a probability space* $(\Omega, \mathfrak{F}, P)$*.*

$$\text{If } 0 \leq X_1 \leq X_2 \leq \dots \text{ a.s. and } \lim_{n\to\infty} X_n = X \text{ a.s., then } \lim_{n\to\infty} E[X_n] = E[X].$$

PROOF ⋆ : Will not be given. Observe though that **(2)** matches **(1)** in the special case that $\mu(\Omega) = 1$. ■

Remark 4.25. ⋆

Observe that neither Example 4.10 nor Example 4.11 satisfy the condition of the theorem. (The functions in example 4.11 are nonnegative and monotone, but there they are decreasingrather than increasing.) □

Here is another example where the Monotone Convergence Theorem does not apply.

Example 4.12. Is it possible to find Borel measurable functions $f, f_n : \mathbb{R} \to \mathbb{R}$ as follows?

(1) f_n is a bounded sequence, i.e., there is a constant α such that $|f_n(x)| \leq \alpha$ for all x

(2) $f_n \downarrow f$, but $\lim\limits_{n\to\infty} \int f_n d\lambda \neq \int f d\lambda$.

The answer: Yes, this is possible.

Let $\alpha_n \in \mathbb{R}$ such that $\alpha_n \downarrow 0$. Let $f_n(x) := \alpha_n$. Clearly, this sequence of constant functions satisfies $\int f_n d\lambda = \alpha_n \lambda(\mathbb{R}) = \infty$ for all n, thus $\lim_n \int f_n d\lambda = \infty$.

On the other hand, $\int \big(\lim_n f_n\big) d\lambda = \int 0 d\lambda = 0$.

Any sequence $f_n \downarrow 0$ such that $\int f_n d\lambda = \infty$ for all n will do the trick. Thus, $f_n := \mathbf{1}_{[n,\infty[}$, i.e., $f_n(x) = 1$ if $x \geq n$, and 0 otherwise, is another example that satisfies **(1)** and **(2)**.

Note that the Monotone Convergence theorem does not apply since $f_n \uparrow f$ is not satisfied. The Dominated convergence theorem does not apply either, since f_1 is not integrable, thus no integrable g such that $|f_n| \leq g$ for all n can be found. □

Just as useful as the Monotone Convergence Theorem is the following one (SCF2 Theorem 1.4.9.)

Theorem 4.14 (Dominated convergence Theorem)**.**

(1) *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and let $f, g, f_1, f_2, \dots : \Omega \to \mathbb{R}$ be $m(\mathfrak{F}, \mathfrak{B})$. Further assume that $g \geq 0$ and g is integrable, i.e., $\int g d\mu < \infty$.*

$$\text{If } |f_j| \leq g \text{ a.s. for each } j \text{ and } \lim_{n\to\infty} f_n = f \text{ a.s.,} \quad \text{then } \lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

(2) *Let X, Y and $X_1, X_2, X_3, \dots$ be random variables.*

$$\text{If } |X_j| \leq Y \text{ a.s. for each } j \text{ and } \lim_{n\to\infty} X_n = X \text{ a.s.,} \quad \text{then } \lim_{n\to\infty} E[X_n] = E[X].$$

PROOF ⋆ : Will not be given. Observe again that **(2)** matches **(1)** in the special case that $\mu(\Omega) = 1$. ■

You should appreciate how useful the above two theorems are for your other Math classes where integration or summation or probability plays a role. Here is an example which you can find, e.g., in [4] Bauer, Heinz: Measure and Integration Theory.

Proposition 4.17. ⋆

Let $(\Omega, \mathfrak{F}, \mu)$ be a probability space and $a < b$ two real numbers. Assume that the function $f :]a, b[\times \Omega \to \mathbb{R}$ satifies the following.

(1) *For any fixed $a < t < b$, the function $\omega \mapsto f(t, \omega)$ is μ–integrable (and thus $\mathfrak{F}$–measurable).*

(2) *For any fixed $\omega \in \Omega$, the function $t \mapsto f(t, \omega)$ has a partial derivative*

$$f_t : s \mapsto f_t(s, \omega) = \frac{\partial f}{\partial t}(s, \omega).$$

Note that t is not a variable in this context since its only purpose is to indicate differentiation with respect to the first argument of $f(\cdot, \cdot)$.

(3) *There exists a non-negative and μ–integrable function $g : \Omega \to \mathbb{R}$ which dominates $|f_t|$:*

$$|f_t(s, \omega)| \leq g(\omega) \text{ for all } a < s < b, \, \omega \in \Omega.$$

Then we can differentiate under the integral. More specifically,

$$s \mapsto \int_\Omega f(s, \omega) \, d\mu(\omega) \quad \text{is differentiable for each } \omega,$$

Further,

$$\omega \mapsto f_t(s, \omega) \quad \text{is } \mu\text{–integrable for each } a < s < b, \text{ and}$$

$$\int_\Omega f_t(s_0, \omega) \, d\mu(\omega) = \frac{d}{dt} \int_\Omega f(s_0, \omega) \, d\mu(\omega).$$

PROOF: Fix $a < s_0 < b$ and an arbitrary sequence $a < s_n < b$ of real numbers such that $s_n \neq s_0$ for all n and $\lim_n s_n = s_0$. Define $h_n : \Omega \to \mathbb{R}$ as

$$h_n(\omega) := \frac{f(s_n, \omega) - f(s_0, \omega)}{s_n - s_0}.$$

Then h_n is μ–integrable for each n by assumption **(1)** and, by assumption **(2)**,

$$\lim_{n\to\infty} h_n(\omega) \;=\; f_t(s_0,\omega) \;\text{ for all } \omega\in\Omega. \tag{4.76}$$

In particular, the function $\omega \mapsto f_t(s_0,\omega$ is measurable as limit of the measurable h_n.

We next show that $|h_n| \leq g$ so we will be able to apply dominated convergence. According to the mean–value theorem of differential calculus we can find for each s_n a value α_n in the open interval with endpoints s_n and s_0 such that

$$h_n(\omega) \;=\; \frac{f(s_n,\omega) \,-\, f(s_0,\omega)}{s_n \,-\, s_0} \;=\; f_t(\alpha_n,\omega)\,.$$

From assumption **(3)**, we thus obtain $|h_n(\omega)| \leq g(\omega)$. It follows that the function $\omega \mapsto f_t(s_0,\omega)$ is μ–integrable. We apply dominated convergence to formula (4.76) and obtain

$$\lim_{n\to\infty}\int_\Omega h_n(\omega)\,d\mu(\omega) \;=\; \int_\Omega f_t(s_0,\omega)\,d\mu(\omega). \tag{4.77}$$

From the definition of h_n and linearity of the integral we obtain

$$\int_\Omega h_n(\omega)\,d\mu(\omega) \;=\; \frac{\int_\Omega f(s_n,\omega)\,d\mu(\omega) \,-\, \int_\Omega f(s_0,\omega)\,d\mu(\omega)}{s_n \,-\, s_0} \;\;\text{for all } n\,,$$

and this sequence of difference quotients has limit

$$\lim_{n\to\infty}\int_\Omega h_n(\omega)\,d\mu(\omega) \;=\;=\; \frac{d}{dt}\int_\Omega f(s_0,\omega)\,d\mu(\omega).$$

We apply formula (4.77) and obtain

$$\int_\Omega f_t(s_0,\omega)\,d\mu(\omega) \;= \frac{d}{dt}\int_\Omega f(s_0,\omega)\,d\mu(\omega).\;\; \blacksquare$$

Here is a simple consequence of monotone convergence.

Theorem 4.15.

(1). *Let $(\Omega,\mathfrak{F},\mu)$ be a measure space and let $f\geq 0$ be an extended real–valued, Borel–measurable function on Ω. Then the set function*

$$\nu:\mathfrak{F}\longrightarrow[0,\infty],\qquad \nu(A)\;:=\int_A f\,d\mu \tag{4.78}$$

defines a measure on $\mathfrak{F}$.

PROOF:

A. To show that $\nu(\emptyset)=0$ we observe that $\mathbf{1}_\emptyset = 0$, thus $f\cdot\mathbf{1}_\emptyset=0$, thus

$$\nu(\emptyset) = \int 0\,d\mu \;=\; \mu(\Omega)\cdot 0 \;=\; 0.$$

(We have had to use the rule $\infty\cdot 0 = 0$ once or twice!)

B. ν is monotone since $A \subseteq A'$ for measurable A and A' implies $f \cdot \mathbf{1}_A \leq f \cdot \mathbf{1}_{A'}$, thus

$$\nu(A) = \int f \cdot \mathbf{1}_A \, d\mu \;\leq\; \int f \cdot \mathbf{1}_{A'} \, d\mu \;=\; \nu(A').$$

C. ν is σ–additive: Let $A_n \in \mathfrak{F}$ be disjoint and $A := \biguplus_{n \in \mathbb{N}} A_n$. For $k \in \mathbb{N}$ let $B_k := \biguplus_{j \leq k} A_j$. Then

$$0 \;\leq\; \sum_{j=1}^{n} f \cdot \mathbf{1}_{A_j} \;=\; f \cdot \mathbf{1}_{B_n} \uparrow f \cdot \mathbf{1}_A \,.$$

Thus, by monotone convergence,

$$\nu(A) \;=\; \int f \cdot \mathbf{1}_A \, d\mu = \lim_{n \to \infty} \int f \cdot \mathbf{1}_{B_n}. \;=\; \lim_{n \to \infty} \sum_{j=1}^{k} \int f \cdot \mathbf{1}_{A_j} \;=\; \lim_{n \to \infty} \sum_{j=1}^{k} \nu(A_j) \;=\; \sum_{j=1}^{\infty} \nu(A_j) \;\blacksquare$$

4.7 The ILMD Mehod

Introduction 4.3. The abstract integral was defined or computed in the following stages:

(2) For simple functions $f(\omega) \;=\; \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}(\omega),$ we defined $\int f d\,\mu \;=\; \sum_{j=1}^{n} c_j \mu(A_j) \,.$

(3) For any nonnegative (measurable) function f, choose simple functions $0 \leq f_n \uparrow f$. By monotone convergence, $\int f d\,\mu \;=\; \lim_{n \to \infty} \int f_n d\,\mu.$

(4) For arbitrary (measurable) $f = f^+ - f^-$ such that $\int f^+ d\,\mu \;< \infty$ or $\int f^- d\,\mu \;< \infty$, we defined $\int f d\,\mu \;=\; \int f^+ d\,\mu \;-\; \int f^- d\,\mu \,.$

Note that replacing f and f_n with $f\mathbf{1}_A$ and $f_n\mathbf{1}_A, \;\; A \in \mathfrak{F}$, also covers $\int_A \cdots d\mu$.

Why is **(1)** missing? We reserve that case for particularly simple simple functions, the indicator functions. We could have preceded Definition 4.23 (Integral of a simple function) on p.81, which handles **(2)**, by the following.

(1) For $A \in \mathfrak{F}$, define $\int \mathbf{1}_A d\,\mu \;=\; \mu(A) \,.$

This section describes a general method for proving statements that are about integrals. □

Remark 4.26 (The ILMD Mehod). [⋆] If one wants to prove a theorem in which integration plays a central role, the following procedure, which we call the **ILMD method**, [23] often is successful.

[23] When googling the phrase "ILMD Mehod", the author found the following result:

- The Improved Local Mean Decomposition (ILMD) is employed to decompose remanufacturing cost time series data into several components with smooth, periodic fluctuation and use this as input.

So be sure to explain the term when you use it in discussions with others! Other authors use different terms. For example, [14] Shreve, Steve: Stochastic Calculus for Finance II: Continuous-Time Models refers to the ILMD method as the **"Standard Machine"**. .

I	Prove the statement for integrands which are **I**ndicator functions $\mathbf{1}_A(\omega)$.
L	**L**inearity of the integral (Theorem **??(b)** on p.**??** often extends the result to simple functions at little or no cost.
M	**M**onotone convergence (Theorem **??(c)** on p.**??** often extends the result to nonnegative integrands at little or no cost.
D	Writing f as the **D**ifference of two nonnegative functions, e.g., $f = f^+ - f^-$, extends the result to general integrands. This might prove more difficult than the preceding two steps, since expressions of the form $\infty - \infty$ must be avoided. $\square$

The proof of the next theorem demonstrates the usefulness of ILMD.

Theorem 4.16 (Integrals under Transforms)**.** *$(\Omega, \mathfrak{F}, \mu)$ be a measure space and let $(\Omega', \mathfrak{F}')$ be a measurable space. Assume that $f : \Omega \to \Omega'$ is $m(\mathfrak{F}, \mathfrak{F}')$. and $g : \Omega' \to \mathbb{R}$ is $m(\mathfrak{F}', \mathfrak{B}^1)$. We denote again by μ_f the image measure of μ under f on $\mathfrak{F}'$, defined in Definition 4.13 on p.66 and given by*

$$\mu_f(A') \ = \ \mu\{f \in A'\} \ = \ \mu\{\omega \in \Omega \, : \, f(\omega) \in A'\}.$$

If $g \geq 0$ or $g \circ f$ is integrable then

$$\int g \circ f \, d\mu \ = \ \int g \, d\mu_f, \quad \text{i.e.,} \quad \int g\big(f(\omega)\big) \, d\mu(\omega) \ = \ \int g(\omega') \, d\mu_f(\omega'). \tag{4.79}$$

PROOF:

Step 1. Assume that $g = \mathbf{1}_{A'}$ for some $A' \in \mathfrak{F}'$. Note that

$$\mathbf{1}_{A'}\big(f(\omega)\big) \ = \ 1 \ \Leftrightarrow \ f(\omega) \in A' \ \Leftrightarrow \ \omega \in f^{-1}(A'),$$

thus,

$$\int_\Omega \mathbf{1}_{A'}\big(f(\omega)\big) \, d\mu(\omega) \ = \ \int_\Omega \mathbf{1}_{f^{-1}(A')}(\omega) \, d\mu(\omega) \ = \ \mu\big(f^{-1}(A')\big) \ = \ \mu_f(A') \ = \ \int_{\Omega'} \mathbf{1}_{A'}(\omega') \, d\mu_f(\omega').$$

We have shown the validity of formula (4.79) for $g = \mathbf{1}_{A'}$.

Step 2. Let $g \geq 0$ be a simple function $g = \sum\limits_{j=1}^{n} c_j \mathbf{1}_{A'_j}$ $(n \in \mathbb{N}, c_j \geq 0, A_j \in \mathfrak{F})$. It then follows from the linearity of the integral and what we already haven proven in step 1 that

$$\int_\Omega g \circ f \, d\mu \ = \ \sum_{j=1}^{n} c_j \int_\Omega \mathbf{1}_{A'_j} \circ f \, d\mu \ = \ \sum_{j=1}^{n} c_j \int_{\Omega'} \mathbf{1}_{A'_j} \, d\mu_f \ = \ \int_{\Omega'} g \, d\mu_f.$$

Step 3. Assume that g is a nonnegative, $\mathfrak{F}'$–$\mathfrak{B}^1$ measurable function. For each nonnegative integer n let

$$B_{j,n} \ := \left\{ \frac{j}{2^n} \ \leq \ g \ < \ \frac{j+1}{2^n} \right\} \quad (j = 0, 1, \dots, 4^n - 1),$$

$$g_n(\omega') \ := \ \sum_{j=0}^{4^n - 1} \frac{j}{2^n} \cdot \mathbf{1}_{B_{j,n}}(\omega').$$

Then g_n is a simple function which is constant on the preimages $g^{-1}\big([\frac{j}{2^n},\frac{j+1}{2^n}[\big)$ of the partition

$$0 < \frac{1}{2^n} < \frac{2}{2^n} < \cdots \frac{4^n}{2^n} = 2^n.$$

We have $g_n \le g_{n+1}$ for all n since each partition is a refinement of the previous one.
Moreover $g_n(\omega') \uparrow g(\omega')$ for each ω since, if j is the index such that $\frac{j}{2^n} \le g(\omega') < \frac{j+1}{2^n}$, then

$$\omega' \in B_{j,n}, \text{ thus } g_n(x) = \frac{j}{2^n} \le g(\omega') < \frac{j+1}{2^n}, \text{ thus } |g_n(\omega') - g(\omega')| < \frac{j+1}{2^n} - \frac{j}{2^n} = \frac{1}{2^n}.$$

It now follows from **Step 2** and the monotone convergence theorem that

$$\int_\Omega g\circ f\,d\mu = \lim_{n\to\infty}\int_\Omega g_n\circ f\,d\mu = \lim_{n\to\infty}\int_{\Omega'} g_n\,d\mu_f = \int_{\Omega'} g\,d\mu_f.$$

If $f \ge 0$ then we are done.

Step 4. From now on we may assume that $g\circ f$ is μ–integrable, i.e., both $\int(g\circ f)^+d\mu < \infty$ and $\int(g\circ f)^-d\mu < \infty$. We have shown in step 3 that the nonnegative functions $g^+\circ f$ and $g^-\circ f$ satisfy

$$\int_\Omega g^+\circ f\,d\mu = \int_{\Omega'} g^+\,d\mu_f, \qquad \int_\Omega g^-\circ f\,d\mu = \int_{\Omega'} g^-\,d\mu_f, \tag{4.80}$$

We also have

$$\begin{aligned}(g^+\circ f)(\omega) &= g^+\big(f(\omega)\big) = \big[g\big(f(\omega)\big)\big]^+ = (g\circ f)^+(\omega),\\ (g^-\circ f)(\omega) &= g^-\big(f(\omega)\big) = \big[g\big(f(\omega)\big)\big]^- = (g\circ f)^-(\omega).\end{aligned} \tag{4.81}$$

It follows that

$$\begin{aligned}\int_\Omega |g\circ f|\,d\mu &= \int_\Omega (g\circ f)^+\,d\mu + \int_\Omega (g\circ f)^-\,d\mu\\ &\overset{(4.81)}{=} \int_\Omega (g^+\circ f)\,d\mu + \int_\Omega (g^-\circ f)\,d\mu\\ &\overset{(4.80)}{=} \int_{\Omega'} g^+\,d\mu_f + \int_{\Omega'} g^-\,d\mu_f.\end{aligned}$$

All quantities here are finite since $\int(g\circ f)^+d\mu < \infty$ and $\int(g\circ f)^-d\mu < \infty$. We thus may subtract and obtain

$$\int_\Omega g\circ f\,d\mu = \int_{\Omega'} g^+\,d\mu_f - \int_{\Omega'} g^-\,d\mu_f. \quad \blacksquare$$

Here is another application of the ILMD Mehod.

Proposition 4.18.

Let $(\Omega,\mathfrak{F},\mu)$ be a measure space and let $f \ge 0$ be an extended real–valued, Borel–measurable function on Ω. Let ν be the measure defined by

$$\nu(A) := \int_A f\,d\mu$$

(see Theorem 4.15 on p.92). Moreover, let φ be an extended real–valued, Borel–measurable function on Ω such that $\varphi \ge 0$ or φ is ν–integrable. Then,

$$\int_A \varphi\,d\nu = \int_A \varphi\cdot f\,d\mu, \quad \text{for all } A \in \mathfrak{F}. \tag{4.82}$$

PROOF:
Step 1. We prove formula (4.82) for indicator functions. Assume that $\varphi = \mathbf{1}_B$ for some $B \in \mathfrak{F}$. Then

$$\int_A \varphi \, d\nu = \int \mathbf{1}_A \mathbf{1}_B \, d\nu = \int \mathbf{1}_{A\cap B} \, d\nu = \nu(A \cap B)$$
$$= \int_{A\cap B} f \, d\mu = \int \mathbf{1}_A \mathbf{1}_B f \, d\mu = \int_A \mathbf{1}_B f \, d\mu = \int_A \varphi f \, d\mu \, .$$

We have shown the validity of formula (4.82) for $\varphi = \mathbf{1}_B$.
We only give an outline of the remainder of the proof. It closely follows the corresponding steps in the proof of Theorem 4.16 on p.94.
Step 2. linearity of the integral allows to extend the formula from indicator functions to simple functions $\varphi = \sum\limits_{j=1}^{n} c_j \mathbf{1}_{A_j}$ $(n \in \mathbb{N}, c_j \geq 0, A_j \in \mathfrak{F})$.
Step 3. Assume that φ is a nonnegative, $\mathfrak{F} - -\mathfrak{B}^1$ measurable function. We construct a increasingsequence φ_n of simple functions such that $\varphi_n \uparrow \varphi$ in a fashion similar to the proof of Theorem4.16. It easily follows from the monotone convergence theorem that (4.82) is true for φ.
Step 4. To prove the proposition for ν–integrable φ we decompose $\varphi = \varphi^+ - \varphi^-$. Then

$$\int_A \varphi \, d\nu = \int_A \varphi^+ \, d\nu - \int_A \varphi^- \, d\nu = \int_A \varphi^+ \cdot f \, d\mu - \int_A \varphi^- \cdot f \, d\mu$$
$$= \int_A (\varphi^+ - \varphi^-) \cdot f \, d\mu = \int_A \varphi \cdot f \, d\mu \, .$$

Here we repeatedly used linearity of the integral and we applied what we proved in **Step 3** to obtain the second equation. ■

4.8 Equivalent Measures and the Radon–Nikodým Theorem

It is not necessary for you to remember the next definition. It is of a technical nature to ensure that certain important theorems are valid.

Definition 4.31 (σ–finite measure). ⋆

- Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space. We call μ a **σ–finite measure** if there exists a sequence $A_n \in \mathfrak{F}$ such that

$$\mu(A_n) < \infty \text{ for all } n, \qquad \text{and} \quad \bigcup_{n\in\mathbb{N}} A_n = \Omega. \; \square$$

Example 4.13. ⋆

- All finite measures are σ–finite. In particular, all probability measures are σ–finite
- Lebesgue measure λ^n is σ–finite: For $k \in \mathbb{N}$ let $A_k := [-k, k]^n$.
  Then $\lambda^n(A_k) = (2k)^n < \infty$, and $A_k \uparrow \Omega$.
- Counting measure Σ (Definition 4.8 on p.57) is σ–finite:
  For $k \in \mathbb{N}$ let $A_k := \{j \in \mathbb{Z} : |j| \leq k\}$. Then $\Sigma(A_k) = 2k + 1 < \infty$, and $A_k \uparrow \mathbb{Z}$. □

The next definition is an important one to remember.

Definition 4.32 (Radon–Nikodým derivative)**.** Let μ and ν be measures on a given measurable space $(\Omega, \mathfrak{F})$, assume that μ is σ–finite (see Definition 4.31 (σ–finite measure) on p.96), and let $f \geq 0$ be in $m(\mathfrak{F}, \mathfrak{B}^1)$.

If μ, ν, and f satisfy formula (4.78) of Theorem 4.15 on p.92, i.e.,

$$\nu(A) \;=\; \int_A f(\omega)\, d\mu(\omega), \quad \text{for all } A \in \mathfrak{F}, \tag{4.83}$$

then we call f the **density of ν with respect to μ** on $\mathfrak{F}$ or also the **Radon–Nikodým derivative of ν with respect to μ** on $\mathfrak{F}$. We write either of

$$\bullet\, f \;=\; \frac{d\nu}{d\mu}, \quad \bullet\, d\nu \;=\; f\, d\mu, \quad \bullet\, d\nu(\omega) \;=\; f(\omega)\, d\mu(\omega), \quad \bullet\, \nu(d\omega) \;=\; f(\omega)\, \mu(d\omega). \tag{4.84}$$ □

Remark 4.27. We assume again that μ is a σ–finite measure on $(\Omega, \mathfrak{F})$. If $\tilde{f}$ is a second function that satisfies $\nu(A) = \int_A \tilde{f} d\mu$ for all $A \in \mathfrak{F}$ and if f and $\tilde{f}$ are μ–integrable, then $\tilde{f} = f$ μ–a.e. This follows from Theorem 4.8 on p.85. A straightforward application of monotone convergence shows that this almost everywhere uniqueness of the Radon–Nikodým derivative also holds if μ–integrability of f and $\tilde{f}$ is replaced with nonnegativity of f and $\tilde{f}$.

These uniqueness results allow us to refer to "the" Radon–Nikodým derivative. □

Proposition 4.19. *Let $(\Omega, \mathfrak{F}, \mu)$ be a σ–finite measure space. Let $f, g \geq 0$ be in $m(\mathfrak{F}, \mathfrak{B}^1)$. Assume that the measures ν and ρ, defined by*

$$\nu(A) \;:=\; \int_A f\, d\mu\,, \quad \rho(A) \;:=\; \int_A g\, d\nu\,, \qquad (A \in \mathfrak{F})$$

are σ–finite so that uniqueness of the Radon–Nikodým derivative allows us to write

$$f \;=\; \frac{d\nu}{d\mu} \qquad \text{and} \qquad g \;=\; \frac{d\rho}{d\nu}\,.$$

Then $\dfrac{d\rho}{d\mu} \;=\; fg$. In other words, there is a

Chain rule for Radon–Nikodým derivatives:

$$\frac{d\rho}{d\mu} \;=\; \frac{d\rho}{d\nu} \cdot \frac{d\nu}{d\mu}\,. \tag{4.85}$$

PROOF: Let $A \in \mathfrak{F}$. We must prove that $\rho(A) = \int_A (gf) d\mu$. It follows from Proposition 4.18 on p.95 that $\int_A \varphi d\nu \;=\; \int_A (\varphi f) d\mu$ for all measurable and nonnegative φ. Thus, for $\varphi = g$,

$$\rho(A) \;=\; \int_A g\, d\nu \;=\; \int_A (gf)\, d\mu\,,$$

and this is what had to be shown. ■

Remark 4.28. ⋆ There are reasons besides the chain rule (4.85) to call the function f in formula (4.83) a derivative. Consider the normal distribution with mean μ and variance σ^2, i.e., the measure ν on $\mathfrak{B}^1$ defined by

$$\nu(]a,b]) \;=\; \int_a^b f(x)\,dx \;=\; \int_{]a,b]} f\,d\lambda^1\,, \quad a,b \in \mathbb{R},\; a<b, \tag{4.86}$$

where f is the normal density

$$f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{\frac{(x-\mu)^2}{2\sigma^2}}\,.$$

Observe that formula (4.86) extends to arbitrary Borel sets (see Fact 4.1 on p.54). In other words, if we write μ for λ^1, then λ^1, ν, and f satisfy formula (4.83), thus

$$f \;=\; \frac{d\nu}{d\lambda^1}\,. \tag{4.87}$$

Actually ν is completely determined by its values on intervals of the form $]-\infty, x]$ since

$$\nu(]a,b]) \;=\; \nu(]-\infty,b]) \;-\; \nu(]-\infty,a]).$$

This should not come as a surprise, since we only stated that the $N(\mu,\sigma^2)$ distribution is defined by its cumulative distribution function

$$F(x) \;=\; \int_{-\infty}^{x} f(u)\,du \;=\; \int_{]-\infty,x]} f(u)\,d\lambda^1(u).$$

By the Fundamental Theorem of Calculus, $f(x) = \frac{dF(x)}{dx}$. Since (4.87) holds true, we have both

$$f(x) \;=\; \frac{dF(x)}{dx}, \qquad f(x) \;=\; \frac{d\nu(x)}{d\lambda^1}$$

(the second equation follows from (4.87)). This is the reason why a function f that satisfies formula (4.83) is called a (Radon–Nikodým) derivative.

A last comment: This example has nothing to do with normal distributions. All we needed was that the function f in formula (4.86) is nonnegative, in $m(\mathfrak{B}^1, \mathfrak{B}^1)$, and such that the function $x \to F(x) = \nu(]-\infty,x])$ is differentiable so that we can apply the Fundamental Theorem of Calculus. Continuity of f at all points suffices for that. □

Definition 4.33 (μ–continuous measure)**.** ⋆

Let μ and ν be measures on a measurable space $(\Omega, \mathfrak{F})$.

- We call ν a **continuous measure with respect to** μ on $\mathfrak{F}$ or a μ**–continuous measure** on $\mathfrak{F}$, and we write $\nu \ll \mu$, if

$$\mu(A) \;=\; 0 \;\Rightarrow\; \nu(A) \;=0, \quad \text{for all } A \in \mathfrak{F}.$$

- We call μ and ν **equivalent measures**, and we write $\mu \sim \nu$, if both

$$\mu \;\ll\; \nu \quad \text{and} \quad \nu \;\ll\; \mu. \;\square$$

Remark 4.29.

(1) Two measures μ and ν on $(\Omega, \mathfrak{F})$ are equivalent if and only if

$$\mu(A) \;=\; 0 \;\Leftrightarrow\; \nu(A) \;= 0, \qquad \text{for all } A \in \mathfrak{F}.$$

Thus the relation $\mu \sim \nu$ above is an equivalence relation on the set of all measures for $(\Omega, \mathfrak{F})$.

(2) Two probabilities P and $\widetilde{P}$ on $(\Omega, \mathfrak{F})$ are equivalent if and only if the P–almost sure events coincide with the $\widetilde{P}$–almost sure events. □

Proposition 4.20.

Let μ and ν be measures on a given measurable space $(\Omega, \mathfrak{F})$ and assume moreover that the measure ν has a Radon–Nikodým derivative with respect to μ on $\mathfrak{F}$. Then $\nu \ll \mu$.

PROOF: ⋆ For convenience we write f rather than $\frac{d\mu}{d\nu}$ for the Radon–Nikodým derivative. Thus f satisfies $\nu(A) = \int_A f d\mu$ for all $A \in \mathfrak{F}$.

We must show that

$$\mu(A) \;=\; 0 \;\Rightarrow\; \int f \mathbf{1}_A \, d\mu \;=\; 0.$$

It suffices to show that $\int h d\mu = 0$ for all simple functions h that satisfy $0 \le h \le f\mathbf{1}_A$, since $\int f\mathbf{1}_A d\mu$ is the supremum of all such integrals.

Since $f\mathbf{1}_A = 0$ on $A^{\complement}$ and thus $0 \le h \le f\mathbf{1}_A = 0$ on $A^{\complement}$, we obtain $h = h\mathbf{1}_A$.

Also, h has the form $h \;=\; \sum\limits_{j=1}^{n} c_j \mathbf{1}_{A_j}$ for suitable $n \in \mathbb{N}$, $c_j \in \mathbb{R}$, and $A_j \in \mathfrak{F}$. Thus,

$$\int h \, d\mu \;=\; \int h\mathbf{1}_A \, d\mu \;=\; \sum_j c_j \int_A \mathbf{1}_{A_j} \, d\mu \;=\; \sum_j c_j \mu(A \cap A_j) \le \sum_j c_j \mu(A) \;=\; 0.$$

The last equation follows from the assumption $\mu(A) = 0$. ■

Theorem 4.17 (Radon–Nikodým Theorem)**.** *Let μ and ν be measures on a measurable space $(\Omega, \mathfrak{F})$.*

If the measure μ is σ–finite then

$$\nu \text{ possesses a Radon–Nikodým derivative } \frac{d\nu}{d\mu} \text{ with respect to } \mu \text{ on } \mathfrak{F} \quad \Leftrightarrow \quad \nu \ll \mu.$$

PROOF: ⋆ The "⇒" direction was proven in Proposition 4.20. The proof of the reverse direction is beyond the scope of these lecture notes. ■

Corollary 4.2. *Let μ and $\widetilde{\mu}$ be equivalent and σ–finite measures on a given measurable space $(\Omega, \mathfrak{F})$. Then both Radon–Nikodým derivatives $\frac{d\widetilde{\mu}}{d\mu}$ and $\frac{d\mu}{d\widetilde{\mu}}$ exist, and they satisfy the relation*

$$\frac{d\widetilde{\mu}}{d\mu} \cdot \frac{d\mu}{d\widetilde{\mu}} = 1 \;\; a.e. \tag{4.88}$$

PROOF: ⋆ The Radon–Nikodým Theorem guarantees the existence of both $\frac{d\widetilde{\mu}}{d\mu}$ and $\frac{d\mu}{d\widetilde{\mu}}$, and (4.88) follows from

$$1 = \frac{d\widetilde{\mu}}{d\widetilde{\mu}} = \frac{d\widetilde{\mu}}{d\mu} \cdot \frac{d\mu}{d\widetilde{\mu}}.$$

The second equation is immediate from Proposition 4.19 (the chain rule for Radon–Nikodým derivatives), and the first one follows from $\widetilde{\mu}(A) = \int_A 1\, d\widetilde{\mu}$ and the a.e. uniqueness of the Radon–Nikodým derivative. ■

Remark 4.30. Assume as in Corollary 4.2 that μ and $\widetilde{\mu}$ are equivalent measures. We write $Z := \frac{d\widetilde{\mu}}{d\mu}$ for convenience. Let $B_0 := \{Z = 0\}$. Then $\widetilde{\mu}(B_0) = 0$ because

$$\widetilde{\mu}(B_0) = \int_{B_0} Z\, d\mu = \int_{B_0} 0\, d\mu = 0.$$

Since $\mu \sim \widetilde{\mu}$ we also have $\mu(B_0) = 0$.

Let X be an arbitrary, nonnegative, random variable. Then

$$\int XZ\, d\mu = \int_{B_0} XZ\, d\mu + \int_{B_0^{\complement}} XZ\, d\mu = 0 + \int_{B_0^{\complement}} XZ\, d\mu. = \int_{B_0^{\complement}} X\mathbf{1}_{\{Z\neq 0\}} Z\, d\mu.$$

The above holds in particular for indicator functions $X = \mathbf{1}_A$ of any $A \in \mathfrak{F}$ and tells us that we may replace Z with $Z\mathbf{1}_{\{Z\neq 0\}}$. This should have been expected since a Radon–Nikodým derivative is a conditional expectation and thus determined only almost everywhere.

We thus may assume that

$$\frac{d\widetilde{\mu}}{d\mu} = 1 \Big/ \frac{d\mu}{d\widetilde{\mu}}. \ \square$$

We finish the discussion of the Radon–Nikodým Theorem with a concrete example.

Example 4.14. Let $p, q : \mathbb{N} \to]0,1[$ be strictly positive. Assume that $\sum_j p(j) = \sum_j q(j) = 1$. Thus $P(\{k\}) := p(k)$ and $Q(\{k\}) := q(k)$ defines two probability measures P and Q on the measurable space $(\mathbb{N}, 2^{\mathbb{N}})$.

Since the empty set is the only set $A \subseteq \mathbb{N}$ such that $P(A) = 0$ and $Q(A) = 0$, those two measures are equivalent. Thus both Radon–Nikodým derivatives $\frac{dQ}{dP}$ and $\frac{dP}{dQ}$ exist. We claim that

$$\frac{dQ}{dP}(k) = \frac{q(k)}{p(k)}, \quad \text{and} \quad \frac{dP}{dQ}(k) = \frac{p(k)}{q(k)}, (k \in \mathbb{N}).$$

For the proof, let $A \in \mathbb{N}$. Since $A = \biguplus[\{k\}; k \in A]$,

$$\begin{aligned}\int_A \frac{q(k)}{p(k)} P(dk) &= \sum_{k\in A} \int_{\{k\}} \frac{q(k)}{p(k)} P(dk) = \sum_{k\in A} \frac{q(k)}{p(k)} P(\{k\}) \\ &= \sum_{k\in A} \frac{Q(\{k\}}{P(\{k\}} P(\{k\}) = \sum_{k\in A} Q(\{k\} = Q(A).\end{aligned}$$

This proves that $\frac{dQ}{dP}(k) = \frac{q(k)}{p(k)}$ for all $k \in \mathbb{N}$ To show that $\frac{dP}{dQ}(k) = \frac{p(k)}{q(k)}$, you can either repeat the proof above with the roles of p, q switched and those of of P, Q switched, or you can use the relation $\frac{dQ}{dP} \cdot \frac{dP}{dQ} = 1$. $\square$

4.9 Digression: Product Measures ⋆

We know from calculus that under certain conditions the order of integration in an integral of the form $\iint f(x,y)dxdy$ can be switched. For example, if $f(x,y)$ is a continuous function of x and y in a bounded rectangle $[a,b]\times[c,d]$, then

$$\int_a^b \left(\int_c^d f(x,y)\,dy\right)dx \;=\; \int_c^d \left(\int_a^b f(x,y)\,dx\right)dy\,.$$

This skeletal chapter gives an outline of how the above generalizes to integration in abstract measure spaces.

Definition 4.34 (Product spaces and product measures of two factors)**.**
Let $(\Omega_1,\mathfrak{F}_1,\mu)$ and $(\Omega_2,\mathfrak{F}_2,\nu)$ be two measure spaces with σ–finite measures μ and ν.

We call the σ–algebra

(4.89) $$\mathfrak{F}_1\otimes\mathfrak{F}_2 \;:=\; \sigma\{\,A_1\times A_2\;:\;\;A_1\in\mathfrak{F}_1, A_2\in\mathfrak{F}_2\,\},$$

which is generated by all "rectangles" of measurable factors A_1 and A_2, the **product σ–algebra** of $\mathfrak{F}_1$ and $\mathfrak{F}_2$. One can show that the set function

(4.90) $$A_1\times A_2\mapsto\mu(A_1)\,\nu(A_2)$$

can be uniquely extended to a measure $\mu\times\nu$ on all of $\mathfrak{F}_1\otimes\mathfrak{F}_2$. We call $\mu\times\nu$ the **product measure**, also just the **product**, of μ and ν, and we call

$$(\Omega_1\times\Omega_2,\mathfrak{F}_1\otimes\mathfrak{F}_2,\mu\times\nu)$$

the **product space** of $(\Omega_1,\mathfrak{F}_1,\mu)$ and $(\Omega_2,\mathfrak{F}_2,\nu)$. □

Example 4.15. We examine the case of two Euclidean spaces $(\mathbb{R}^m,\mathfrak{B}^m,\lambda^m)$ and $(\mathbb{R}^n,\mathfrak{B}^n,\lambda^n)$ with their Borel sets and Lebesgue measures. It can be shown that

$$\mathfrak{B}^m\otimes\mathfrak{B}^n \;=\; \mathfrak{B}^{m+n},$$

and it is obvious from the formula

$$\lambda^m\times\lambda^n(B_1\times B_2) \;=\; \lambda^m(B_1)\,\lambda^n(B_2) \;=\; \lambda^{m+n}(B_1\times B_2)$$

and the uniqueness of the product measure, that $\lambda^m\times\lambda^n=\lambda^{m+n}$. In particular, $\lambda^2=\lambda\times\lambda$. □

Theorem 4.18 (Fubini-Tonelli)**.** *Let $(\Omega_1,\mathfrak{F}_1,\mu)$ and $(\Omega_2,\mathfrak{F}_2,\nu)$ be two measure spaces with σ–finite measures μ and ν. Assume that the extended real–valued function*

$$f:(\Omega_1\times\Omega_2,\mathfrak{F}_1\otimes\mathfrak{F}_2,\mu\times\nu)\to(\bar{\mathbb{R}},\mathfrak{B}^1)$$

is $(\mathfrak{F}_1\otimes\mathfrak{F}_2$–$\mathfrak{B}^1$–measurable. Then $\omega_1\mapsto f(\omega_1,\omega_2)$ is $\mathfrak{F}_1$–measurable for each fixed ω_2 (and thus can be integrated with respect to μ_1), and $\omega_2\mapsto f(\omega_1,\omega_2)$ is $\mathfrak{F}_2$–measurable for each fixed ω_1.

If $f \geq 0$ or f is $\mu \times \nu$–integrable then

$$\int_{A_1 \times A_2} f \, d\mu \times \nu \;=\; \int_{A_1} \left(\int_{A_2} f(\omega_1, \omega_2) \, d\nu(\omega_2) \right) d\mu(\omega_1) \;=\; \int_{A_2} \left(\int_{A_1} f(\omega_1, \omega_2) \, d\mu(\omega_1) \right) d\nu(\omega_2). \tag{4.91}$$

In particular, switching the order of integration yields the same result.

Remark 4.31. ⋆

- We have omitted some technical details concerning μ_1–a.e. and μ_2–a.e. properties in the case of integrable f.
- The case for integrable f was proved first by Guido Fubini in 1907, the case for nonnegative f two years later by Leonida Tonelli, both Italian mathematicians. Since Fubini was first, Theorem 4.18 is often just referred to as Fubini's theorem.
- For general $A \in \mathfrak{F}_1 \otimes \mathfrak{F}_2$ one defines "ω_1–slices" $A_{\omega_1} := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}$ and "ω_2–slices" $A_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$ and evaluates integrals over A as iterated integrals involving those slices. We omit the arguments:

$$\int_A f \, d\mu \times \nu \;=\; \int_{\Omega_1} \left(\int_{A_{\omega_1}} f \, d\nu \right) d\mu \;=\; \int_{\Omega_2} \left(\int_{A_{\omega_2}} f \, d\mu \right) d\nu. \;\square$$

- Of particular interest will be the case of an extended real–valued continuous time stochastic process $X = X(t, \omega), t \in I$ which we assume to be $(\mathfrak{B}(I) \otimes \mathfrak{F})$–measurable. Recall that expectations are integrals dP. Thus Fubini-Tonelli asserts that for $[a, b[\subseteq I$,

$$\int_{[a,b[\times \Omega} X \, d\lambda^1 \times P \;=\; \int_a^b E[X_t] \, dt \;=\; E\left[\int_a^b X_t \, dt \right]$$

4.10 Independence

All material in this chapter is standard and no effort is made to present the material different from SCF2. Consult SCF2 ch.2.2 (Independence) for examples and more background information.

Introduction 4.4. We proceed in stages. Let $(\Omega, \mathfrak{F}, P)$ be a probability space.

Stage 1.

We say that two sets A and B in $\mathfrak{F}$ are independent if

$$P(A \cap B) \;=\; P(A) \cdot P(B).$$

Stage 2.

The following is SCF2 Definition 2.2.1. Let $(\Omega, \mathfrak{F}, P)$ be a probability space, let $\mathfrak{G}$ and $\mathfrak{H}$ be sub–σ–algebras of $\mathfrak{F}$, and let X and Y be random variables on $(\Omega, \mathfrak{F}, P)$.

(a) We say that the **σ–algebras $\mathfrak{G}$ and $\mathfrak{H}$ are independent** if

$$P(A \cap B) = P(A) \cdot P(B) \text{ for all } A \in \mathfrak{G},\ B \in \mathfrak{H}.$$

(b) We say that the **random variables X and Y are independent** if the σ–algebras they generate, $\sigma(X)$ and $\sigma(Y)$, are independent.

(c) We say that **the random variable X is independent of the σ–algebra $\mathfrak{G}$** if the σ–algebras $\sigma(X)$ and $\mathfrak{G}$, are independent.

Note that independence of the (Borel–measurable) random variables X and Y implies that

$$X \text{ and } Y \text{ are independent} \quad \Leftrightarrow \quad \begin{cases} P\{X \in U \text{ and } Y \in V\} = P\{X \in U\} \cdot P\{Y \in V\} \\ \text{for all Borel subsets } U \text{ and } V \text{ of } \mathbb{R}. \end{cases}$$

Stage 3.

SCF2 Definition 2.2.3 generalizes independence from two sub–σ–algebras or random variables to countably many.

Let $(\Omega, \mathfrak{F}, P)$ be a probability space, let $\mathfrak{G}_1, \mathfrak{G}_2, \mathfrak{G}_2, \dots$ be sub–σ–algebras of $\mathfrak{F}$, and let $X_1, X_2, X_3, \dots$ be a sequence of random variables on $(\Omega, \mathfrak{F}, P)$.

(a) We say that the **σ–algebras $\mathfrak{G}_1, \mathfrak{G}_2, \dots, \mathfrak{G}_n$ are independent** if

$$P(A_1 \cap A_2 \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n) \text{ for all } A_j \in \mathfrak{G}_j, \quad j = 1, \dots n.$$

(b) We say that the **random variables $X_1, X_2, \dots X_n$ are independent** if the σ–algebras they generate, $\sigma(X_1), \sigma(X_1), \dots, \sigma(X_n)$, are independent.

(c) We say that the sequence of σ–algebras $\mathfrak{G}_j,\ j \in \mathbb{N}$ is independent if, for each $n \in \mathbb{N}$, the σ–algebras $\mathfrak{G}_j,\ j = 1, \dots, n$ are independent.

(d) We say that the sequence of random variables $X_j,\ j \in \mathbb{N}$ is independent if, for each $n \in \mathbb{N}$, the random variables $X_j,\ j = 1, \dots, n$ are independent.

It is not hard to see that items **(c)** and **(d)** of that definition are equivalent to

(c′) We say that the sequence of σ–algebras $\mathfrak{G}_j,\ j \in \mathbb{N}$ is independent if, for each finite subsequence $n_1, n_2, \dots, n_k$ of distinct integers n_j, the σ–algebras $\mathfrak{G}_{n_j},\ j = 1, \dots, k$ are independent.

(d′) We say that the sequence of random variables $X_j,\ j \in \mathbb{N}$ is independent if, for each finite subsequence $n_1, n_2, \dots, n_k$ of distinct integers n_j, the random variables $X_{n_j},\ j = 1, \dots, k$ are independent.

We will use this observation to define independence of arbitrary (possibly uncountable) families of sub–σ–algebras and random variables. $\square$

Definition 4.35 (Independence)**.** Let $(\Omega, \mathfrak{F}, P)$ be a probability space, let $\mathfrak{G}_i, i \in I$, be an arbitrary, indexed family of sub–σ–algebras of $\mathfrak{F}$, and let $X_i, i \in I$, be an arbitrary, indexed family of random variables on $(\Omega, \mathfrak{F}, P)$.

(a) We say that the **σ–algebras $\mathfrak{G}_i, i \in I$, are independent** if, for each finite subsequence $i_1, i_2, \dots, i_k$ of distinct indices $i_j \in I$,

$$P(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k}) \text{ for all } A_{i_j} \in \mathfrak{G}_{i_j}, \quad j = 1, \dots k.$$

(b) We say that the **random variables $X_i, i \in I$, are independent** if the σ–algebras they generate, $\sigma(X_i), i \in I$, are independent.

Theorem 4.19 (SCF2 Theorem 2.2.5)**.**

Let X and Y be independent mndom variables, and let f and g be Borel-measurable functions on $\mathbb{R}$.

Then, $f \circ X$ and $g \circ Y$ are independent random variables.

PROOF: A simple consequence of the fact that the measurability of f and g yields $\sigma(f \circ X) \subseteq \sigma(X)$ and $\sigma(g \circ Y) \subseteq \sigma(Y)$, so fewer equations of the form $P(A \cap B) = P(A)P(B)$ need to be verified. ■

You will have to consult SCF2, ch.2.2 if you need a refresher on joint distributions to understand the next theorem.

Theorem 4.20 (SCF2 Theorem 2.2.7)**.**

Let X and Y be random variables. We have equivalence

(1)** $\Leftrightarrow$ **(2)** $\Leftrightarrow$ **(3)** $\Leftrightarrow$ **(4)** $\Leftrightarrow$ **(5)

of the following conditions.

***(1)** X and Y are independent.*

***(2)** The joint distribution measure, i.e., the image measure of P under the measurable function $\omega \mapsto \big(X(\omega), X(\omega)\big)$, factors:*

(4.92) $$P_{X,Y}(A \times B) = P_X(A) \cdot P_Y(B) \text{ for all Borel sets } A, B \subseteq \mathbb{R}.$$

***(3)** The joint cumulative distribution function factors:*

(4.93) $$F_{X,Y}(a,b) = F_X(a) \cdot F_Y(b) \text{ for all } a, b \in \mathbb{R}.$$

***(4)** The joint moment–generating function factors:*

(4.94) $$E\left[e^{uX+vY}\right] = E\left[e^{uX}\right] \cdot E\left[e^{vY}\right] \text{ for all } u, v \in \mathbb{R}$$

assuming that the expectations are finite.

***(5)** If there is a joint density then it factors:*

(4.95) $$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \text{ for all } x, y \in \mathbb{R}.$$

The conditions above imply but are not equivalent to the following.
(6) *The expectation factors:*

$$E[X \cdot Y] \;=\; E[X] \cdot E[Y], \quad \textit{provided} \;\; E[\,|X \cdot Y|\,] \;<\; \infty. \tag{4.96}$$

PROOF (outline): See the SCF2 text. ■

4.11 Exercises for Ch.4

Exercise 4.1. Prove Thm.4.1 on p.63 of this document: Let $(\Omega, \mathfrak{F})$ and $(\Omega', \mathfrak{F}')$ be measurable spaces and $f : \Omega \to \Omega'$. Let $\mathfrak{E}' \subseteq \mathfrak{F}'$ such that $\sigma(\mathfrak{E}') \;=\; \mathfrak{F}'$. Then the following is true:

$$\text{If } f^{-1}(A') \subseteq \mathfrak{F} \text{ for all } A' \in \mathfrak{E}' \quad \text{then} \quad f \text{ is } (\mathfrak{F}, \mathfrak{F}')\text{–measurable.} \;\square$$

Exercise 4.2. Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and let $(\Omega', \mathfrak{F}')$ be a countable, measurable space in which $\{\omega'\} \in \mathfrak{F}'$ for all $\omega' \in \Omega'$. Let $f : \Omega \to \Omega'$ be a random element, i.e., f is $(\mathfrak{F}, \mathfrak{F}')$–measurable.
Prove the following. If $P(A) = 1$ or $P(A) = 0$ for all $A \in \mathfrak{F}$, then f = const P–a.s. In other words, there exists $\omega'_0 \in \Omega'$ such that $P\{f = \omega_0\} = 1$.
Hint: There are counterexamples if Ω' is not countable, so use it! □

4

Exercise 4.3. Prove **(1)** and **(2)** of prop.4.13 on p.67 of this document. □

Exercise 4.4. Prove prop.4.12 on p.66 of this document: If $f \in m(\mathfrak{F}, \mathfrak{F}')$ then

$$\mu'(A') \;:=\; \mu\{f \in A'\} \quad \text{defines a measure on } (\Omega', \mathfrak{F}').$$

If μ is a probability measure then so is μ_f. □

Exercise 4.5. Prove closed book prop.4.14 on p.78 of this document: Every process X_t is $\mathfrak{F}_t^X \;=\; \sigma\{X_s : s \in I, s \leq t\}$–adapted. □

Exercise 4.6. Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space with a sub–σ–algebra $\mathfrak{G}$ and let $\mu' \;:=\; \mu\big|_{\mathfrak{G}}$ be the restriction $\mu'(G) := \mu(G)(G \in \mathfrak{G})$ of μ to $\mathfrak{G}$.
Prove that if f is a nonnegative and $\mathfrak{G}$–measurable function then

$$\int f \, d\mu \;=\; \int f \, d\mu'. \;\square$$

5 Conditional Expectations

We will explore in Section 5.1 (Functional Dependency of Random Variables) in what sense a σ–algebra can be interpreted as holding some or all stochastically relevant information about a random variable before devoting the remainder of this chapter to the subject of conditional expectations.

For a random variable X on a probability space $(\Omega, \mathfrak{F}, P)$, we will not define its conditional expectation $E[X \mid \mathfrak{G}]$ with respect to a sub–σ–algebra $\mathfrak{G}$ of $\mathfrak{F}$ as a number. Instead, $E[X \mid \mathfrak{G}]$ will be a $\mathfrak{G}$–measurable random variable (a function of ω!), which satisfies the

$$\textbf{partial averaging property:} \qquad \int_G E[X \mid \mathfrak{G}]\, dP \;=\; \int_G X\, dP \quad \text{for all } G \in \mathfrak{G}\,.$$

This property has its name from the fact that X and $E[X \mid \mathfrak{G}]$ possess matching "averages"

$$\frac{1}{P(G)} \int_G E[X \mid \mathfrak{G}]\, dP \;=\; \frac{1}{P(G)} \int_G X\, dP \quad \text{for all } G \in \mathfrak{G} \text{ such that } P(G) > 0\,,$$

i.e., for that part of the stochastically relevant information about X that is accessible in $\mathfrak{G}$.

In Section 5.2 (σ–Algebras Generated by Countable Partitions and Partial Averages), we examine this first in the special case where $\mathfrak{G}$ is generated by a countable partition

$$\Omega \;=\; G_1 \biguplus G_2 \biguplus G_3 \biguplus \cdots$$

of events G_j, before treating the general case in Section 5.3 (Conditional Expectations in the General Setting).

5.1 Functional Dependency of Random Variables

All propositions and theorems of this subchapter are marked as optional, since they are quite abstract in nature and not easy to understand. Thus it is OK if you **skip them** if you cannot make sense of what they tell you. Note though, that it is **very important** that you study Remark 5.2 on p.109 (at the end of this subchapter) very carefully, since it gives you a feeling for σ–algebras and filtrations as the stores of information of random variables and stochastic processes, and that is very important knowledge if you want to understand the mathematical models of financial markets to be presented in later chapters.

Proposition 5.1 (Doob Composition Lemma). ⋆

Assume that Ω is a nonempty set, not necessarily a measurable space, that $(\Omega', \mathfrak{F}')$ is a measurable space, and that $f : \Omega \to \Omega'$ is a function about which we assume nothing. Recall that f transforms Ω into a measurable space $\big(\Omega, \sigma(f)\big)$ by means of the σ–algebra

$$\sigma(f) \;=\; \{f^{-1}(A') : A' \in \mathfrak{F}\}.$$ [24]

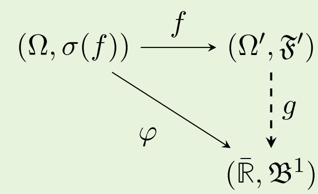
Further, assume that $\varphi : \Omega \to \bar{\mathbb{R}}$ is an extended real–valued function with domain Ω. Then

[24] See Definition 4.14 on p.67 and the proposition preceding it.

(1) *φ is $\big(\sigma(f), \mathfrak{B}^1\big)$–measurable $\Leftrightarrow$ there is $(\mathfrak{F}', \mathfrak{B}^1)$–measurable g such that $\varphi = g \circ f$, i.e., $\varphi(\omega) = g\big(f(\omega)\big)$ for all $\omega \in \Omega$.*

(2) *If $f \geq 0$, then g can be chosen such that $g \geq 0$.*

(3) *If $|f| < \infty$, then g can be chosen such that $|g| < \infty$.*



PROOF (outline):

We will only prove the nontrivial direction "$\Rightarrow$" of **(1)**. The other direction is obvious, since if there is $(\mathfrak{F}', \mathfrak{B}^1)$–measurable g such that $\varphi = g \circ f$ then φ is $\big(\sigma(f), \mathfrak{B}^1\big)$–measurable as the composition of the $\big(\sigma(f), \mathfrak{F}'\big)$–measurable f with the $(\mathfrak{F}', \mathfrak{B}^1)$–measurable g.

The proof of "$\Rightarrow$" is done according to the ILMD Mehod.

Step 1: φ is a $\sigma(f)$ measurable indicator function, i.e., $\varphi = \mathbf{1}_A$ for some $A \in \sigma(f)$. Any such set A must be the preimage $f^{-1}(A')$ of some $A' \in \mathfrak{F}'$. Note that if f is not bijective, then A will generally not uniquely determine A'. We define

$$g := \mathbf{1}_{A'},$$

and it is easily verified that $\mathbf{1}_{A'} \circ f = \mathbf{1}_A$. It follows that $g \circ f = \varphi$.

Step 2: For a nonnegative simple function $\varphi := \sum_{j=1}^{k} c_j \mathbf{1}_{A_j}$ $(c_j \geq 0, A_j \in \sigma(f))$, we define

$$g := \sum_{j=1}^{k} c_j \mathbf{1}_{A'_j},$$

where each $A'_j \in \mathfrak{F}'$ is chosen such that $A_j = f^{-1}(A'_j)$. Then $g \circ f = \varphi$.

Step 3: For general measurable $\varphi \geq 0$ there exists a sequence of simple functions φ_n such that $\varphi_n \uparrow \varphi$. See the proof of step 3 of Theorem 4.16 on p.94. According to **Step 2**, there exist $\mathfrak{F}'$–measurable (simple) functions g_n such that $\varphi_n = g_n \circ f$ for each n. Clearly the sequence g_n is increasing. Therefore, it has an $\mathfrak{F}'$–measurable limit g. This limit function satisfies $\varphi = g \circ f$.

The proof of **(1)** for general g and that of **(3)** will not be given, since it is somewhat tedious to consider the case $\infty - \infty$. But note that we have given a proof of **(2)**. ■

The following corollary to the Doob Composition Lemma is so important, that we give it the status of a theorem.

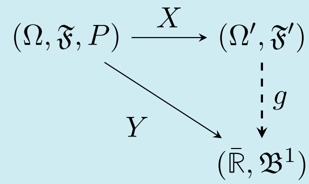
Theorem 5.1 (Functional dependency theorem I). [⋆] *Given are a probability space $(\Omega, \mathfrak{F}, P)$, a measurable space $(\Omega', \mathfrak{F}')$, a random element X, i.e., X is $(\mathfrak{F}, \mathfrak{F}')$–measurable,* [25] *and an extended real–valued random variable Y on $(\Omega, \mathfrak{F}, P)$. Note that our assumptions imply*

$$\sigma(X) \subseteq \mathfrak{F} \quad \text{and} \quad \sigma(Y) \subseteq \mathfrak{F}\,.$$

Thus, the probablities $P\{X \in A'\}$ and $P\{Y \in B\}$ exist for all $A' \in \mathfrak{F}'$ and $B \in \mathfrak{B}$.

[25] See Definition 4.11 on p.62.

Then $\sigma(Y) \subseteq \sigma(X) \;\Leftrightarrow\;$ there is $(\mathfrak{F}', \mathfrak{B}^1)$–measurable g such that $Y = g \circ X$, i.e., $Y(\omega) = g\big(X(\omega)\big)$ for all $\omega \in \Omega$.



PROOF: This is an immediate consequence of the Doob Composition Lemma, Proposition 5.1, since $\sigma(Y) \subseteq \sigma(X) \;\Leftrightarrow\; Y$ is $\big(\sigma(X), \mathfrak{B}^1\big)$–measurable ■

We now apply Doob composition to stochastic processes.

Theorem 5.2 (Functional dependency theorem II). ⋆

Let $X = (X_u)_{0 \le u \le T}$ and $Y = (Y_u)_{0 \le u \le T}$ be stochastic processes on $(\Omega, \mathfrak{F}, P)$ such that Y is adapted to X, i.e., Y_t is $\mathfrak{F}^X_t$–measurable for each $0 \le t \le T$.
Then there is for each $t \in [0, T]$ a $(\mathfrak{F}', \mathfrak{B}^1)$–measurable function $g = g(t, \cdot)$ (which carries t as an additional argument since it depends on t) such that

$$Y_t(\omega) \;=\; g\left(t, \big(X_u(\omega)\big)_{0 \le u \le t}\right) . \tag{5.1}$$

PROOF (outline): We can interpret the process $X = (X_u)_{0 \le u \le T}$ as a random element

$$(X_u)_{0 \le u \le T} : (\Omega, \mathfrak{F}^X_T, P) \to (\Omega', \mathfrak{F}') ; \qquad \omega \mapsto (X_u)_{0 \le u \le T}(\omega)$$

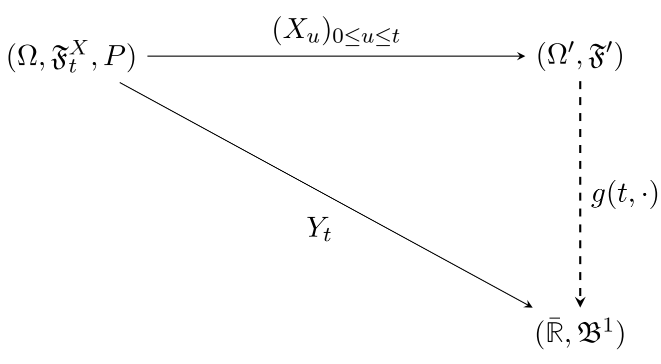
which assigns to $\omega \in \Omega$ its X–trajectory between times 0 and T. So Ω' is the space of all trajectories between times 0 and T and $\mathfrak{F}'$ a suitable σ–algebra on that space.

We can do the above with any $0 \le t \le T$ instead of T and view $(X_u)_{0 \le u \le t}$ as a random element

$$(X_u)_{0 \le u \le t} : (\Omega, \mathfrak{F}^X_t, P) \to (\Omega', \mathfrak{F}') ; \qquad \omega \mapsto (X_u)_{0 \le u \le t}(\omega) \tag{5.2}$$

which assigns to $\omega \in \Omega$ its X–trajectory $\; u \mapsto X_u(\omega) \;$ between times 0 and t.

The Doob composition lemma remains valid in that setting but now the diagram is



This way we obtain for each $t \in [0, T]$ the existence of a $(\mathfrak{F}', \mathfrak{B}^1)$–measurable function $g = g(t, \cdot)$ (which carries t as an additional argument since it depends on t) such that

$$Y_t(\omega) \;=\; g\left(t, \big(X_u(\omega)\big)_{0 \le u \le t}\right) . \;\blacksquare \tag{5.3}$$

Remark 5.1. ⋆ It is easy to see that the condition "Y_t is $\mathfrak{F}_t^X$–measurable for each $0 \le t \le T$" is equivalent to "$\mathfrak{F}_t^Y \subseteq \mathfrak{F}_t^X$ for each $0 \le t \le T$."

Remark 5.2. Given are a probability space $(\Omega, \mathfrak{F}, P)$, and a measurable space $(\Omega', \mathfrak{F}')$. The results of this chapter are not needed to see the following:

(1) For a random element X in $m(\mathfrak{F}, \mathfrak{F}')$, we can interpret the σ–algebra $\sigma(X)$ as the container of all stochastically relevant information of X in the following sense. Knowledge of all events that belong to $\sigma(X)$ means knowledge of the probabilities of all those events $A \subseteq \Omega$ that can be described in terms involving X.

(2) Likewise, the filtration element $\mathfrak{F}_t^X = \sigma\{X_s : s \le t\}$ of Definition 4.19 (Filtration for a process X_t) on p.75 belonging to a stochastic process $(X_t)_t$ of such random elements X_t in $m(\mathfrak{F}, \mathfrak{F}')$ is the container of all stochastically relevant information of this process up to time t (for each time t).

(3) More generally, a process $(X_t)_t$ is adapted to a filtration $(\mathfrak{F}_t)_t \quad \Leftrightarrow \quad \mathfrak{F}_u$ contains all stochastically relevant information of $(X_t)_t$ up to time u (for each u). □

The functional dependency theorems of this subchapter tell us that certain measurabilty conditions for two random elements or two stochastic processes imply an ω–by–ω connection between them.

(4) Assume that the random variable Y is stochastically known to a random element X in the sense that its stochastically relevant information $\sigma(Y)$ is part of that of X. In other words, assume that $\sigma(Y) \subseteq \sigma(X)$. Then that by itself implies that Y is known to X on an ω–by–ω basis: The functional dependency $Y = g \circ X$, i.e., the assignment $\omega \mapsto g(\omega)$, determines $Y(\omega)$ from $X(\omega)$ as $Y(\omega) = g\big(X(\omega)\big)$.

(5) Given are two processes X_t and Y_t. Then $(Y_t)_t$ is $(\mathfrak{F}_t^X)_t$–adapted $\Leftrightarrow$ for each t, the random element $(Y_t(\omega)$ is a (measurable) function of the $X(\cdot, \omega)$ trajectory between times 0 and t. □

5.2 σ–Algebras Generated by Countable Partitions and Partial Averages

Introduction 5.1. We consider σ–algebras as stores of information from a different perspective. In Section 5.1 (Functional Dependency of Random Variables) we were comparing the σ–algebras $\sigma(X)$ and $\sigma(Y)$ of two random variables X and Y and saw that a functional dependency $Y = g \circ X$ exists if $\sigma(Y) \subseteq \sigma(X)$.

Now we relate a random variable X on a probability space $(\Omega, \mathfrak{F}, P)$ to a σ–algebra $\mathfrak{G} \subseteq \mathfrak{F}$ which only contains some but not all of the stochastically relevant information about X, i.e., we examine the relationship of X and $\mathfrak{G}$ in case that

$$\sigma(X) \text{ is } \underline{\text{not}} \text{ contained in } \mathfrak{G}.$$

The following questions arise in this context.

(A) Is there a random variable $X_{\mathfrak{G}} \in m(\mathfrak{G}, \mathfrak{B}^1)$ which is, in some sense, the best possible approximation of X?

(B) Is such an $X_{\mathfrak{G}}$ uniquely determined?

(C) What happens in the extreme case $\mathfrak{G} = \{\emptyset, \Omega\}$? [26]

Since we expect $\mathfrak{G}$ and $X_{\mathfrak{G}}$ to be about stochastically relevant information of X, and since all such information is about probabilities, we immediately have the following partial answer to **(B)**:

$X_{\mathfrak{G}}$ is, at best, only determined almost surely, i.e., up to a set of probability zero.

In other words, if a best approximation $X_{\mathfrak{G}}$ exists, then any random variable $X'_{\mathfrak{G}} \in m(\mathfrak{G}, \mathfrak{B}^1)$ which satisfies $X'_{\mathfrak{G}} = X_{\mathfrak{G}}$ P–a.s. will serve as well.

Consider the special case in which a finite or infinite sequence of events $G_1, G_2, \dots$ is a partition of Ω and generates $\mathfrak{G}$, i.e., if J denotes the finite or infinite index set for this sequence,

$$G_i \cap G_j = \emptyset \text{ for } i \neq j, \quad \biguplus_{j \in J} G_j = \Omega, \quad \mathfrak{G} = \sigma\{G_j : j \in J\}. \tag{5.4}$$

The partitioning events G_j are the "atoms" of $\mathfrak{G}$, since each $G \in \mathfrak{G}$ is a union of some or all of the G_j. See Proposition 4.2 on p.51. Let n be the finite or infinite number of sets G_j.

(1) If $|J| = 1$, then $\Omega = G_1$, i.e., $\mathfrak{G} = \{\emptyset, \Omega\}$. Only constant functions $\Omega \to \mathbb{R}$ are $\mathfrak{G}$–measurable, and the best estimate $\omega \mapsto X_{\mathfrak{G}}(\omega)$ of a random variable X by a number is its expectation $X_{\mathfrak{G}}(\omega) = E[X]$. We have found answers to questions **(A)** and **(C)**.

(2) If $|J| = 2$, then $\Omega = G_1 \uplus G_2$, thus $G_2 = G_1^{\complement}$, and $\mathfrak{G} = \{\emptyset, G_1, G_2, \Omega\}$. We now can separately consider the cases $\omega \in G_1$, $\omega \in G_2$ and take the weighted averages on G_1 and G_2, i.e, we define

$$\begin{aligned} X_{\mathfrak{G}}(\omega) &:= \begin{cases} \frac{1}{P(G_1)} E[X\mathbf{1}_{G_1}] & \text{if } \omega \in G_1, \\ \frac{1}{P(G_2)} E[X\mathbf{1}_{G_2}] & \text{if } \omega \in G_2. \end{cases} \\ &= \frac{1}{P(G_1)} E[X\mathbf{1}_{G_1}] \cdot \mathbf{1}_{G_1}(\omega) + \frac{1}{P(G_2)} E[X\mathbf{1}_{G_2}] \cdot \mathbf{1}_{G_2}(\omega) \\ &= \sum_{j=1,2} \frac{1}{P(G_j)} E\left[X\mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j}(\omega). \end{aligned}$$

(3) For general J we take the weighted averages on each G_j and splice them into a function of ω:

$$X_{\mathfrak{G}}(\omega) := \frac{1}{P(G_j)} E\left[X\mathbf{1}_{G_j}\right] \text{ if } \omega \in G_j, \quad \text{i.e.,} \quad X_{\mathfrak{G}}(\omega) = \sum_{j \in J} \frac{1}{P(G_j)} E\left[X\mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j}(\omega).$$

The equations given in **(2)** and **(3)** only work if $P(G_j) \neq 0$ for all indices j. Otherwise we amend those formulas as follows. We partition our index set J into two index sets

$$J = J_1 \biguplus J_0, \quad \text{defined as } J_1 := \{j \in \mathbb{N} : P(G_j) > 0\}, \quad J_0 := \{j \in \mathbb{N} : P(G_j) = 0\}.$$

We have learned that $X_{\mathfrak{G}}$ can be determined at best up to a P–null set. The set $A_0 := \biguplus_{J_0} G_j$ has probability zero as the countable union of P–null sets. Thus we do not change any stochastically

[26] The other extreme case, $\mathfrak{G} = \mathfrak{F}$, is not up for discussion since we assumed that $\sigma(X) \nsubseteq \mathfrak{G}$.

relevant properties if we set $X_{\mathfrak{G}}$ on A_0 to some arbitrary number, most conveniently zero. In other words, we replace the definition given in **(3)** with

$$X_{\mathfrak{G}}(\omega) := \sum_{j \in J_1} \frac{1}{P(G_j)} E\left[X \mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j}(\omega)\,. \tag{5.5}$$

Now let us reason why $X_{\mathfrak{G}}$ might be a solution to question **(A)**. For this we briefly explore the connection between $X_{\mathfrak{G}}$ and conditional expectations $E[X \mid G]$ with respect to events $G \in \mathfrak{G}$. You have encountered such conditional expectations in your probability course for the special case that X is a discrete random variable. For an event G, they were defined as

$$E[X \mid G] = \sum_x x P\{X = x \mid G\}\,.$$

In particular, if Y is another variable and $G = \{Y = y\}$, then

$$E[X \mid Y = y] = \sum_x x P\{X = x \mid Y = y\}\,.$$

If X is not discrete but possesses a conditional density $f_{X|G}(x)$ instead, then we defined

$$E[X \mid G] = \int_{-\infty}^{\infty} x f_{X|G}(x)\,dx\,, \qquad \text{i.e.,} \quad P(A \mid G) = \int_A f_{X|G}(x)\,dx \ \text{ for all events } A\,.$$

We obtain for indicator functions $X = \mathbf{1}_A (A \in \mathfrak{F})$ the following.

$$\begin{aligned} X_{\mathfrak{G}}(\omega) &= \sum_j \frac{1}{P(G_j)} E\left[\mathbf{1}_{G_j} \mathbf{1}_A\right] \cdot \mathbf{1}_{G_j}(\omega) = \sum_j \frac{P(G_j \cap A)}{P(G_j)} \cdot \mathbf{1}_{G_j}(\omega) \\ &= \sum_j P(A \mid G_j) \cdot \mathbf{1}_{G_j}(\omega) = \sum_j E(\mathbf{1}_A \mid G_j) \cdot \mathbf{1}_{G_j}(\omega) = \sum_j E(X \mid G_j) \cdot \mathbf{1}_{G_j}(\omega)\,. \end{aligned}$$

This relationship,

$$X_{\mathfrak{G}}(\omega) = \sum_j E(X \mid G_j) \cdot \mathbf{1}_{G_j}(\omega)\,, \tag{5.6}$$

between $X_{\mathfrak{G}}$ and conditional expectations of the form $E[X \mid G_j]$ can be extended by use of the ILMD Mehod to arbitrary nonnegative or integrable random variables X.

> Note that the right hand side of (5.6) is constant in ω on each partitioning event G_j of $\mathfrak{G}$:
>
> $$X_{\mathfrak{G}}(\omega) = E(X \mid G_j) \ \text{ for each } \omega \in G_j\,. \tag{5.7}$$

This formula gives us the justification to call $X_{\mathfrak{G}}$ (a random variable!) the conditional expectation of X with respect to $\mathfrak{G}$, the σ–algebra which is generated by those events G_j.

Proposition 5.2 which follows this introduction, will show that the integral equation

$$\int_G X_{\mathfrak{G}}\,dP = \int_G X\,dP \tag{5.8}$$

holds for all events $G \in \mathfrak{G}$, and that this property, together with its $\mathfrak{G}$–measurability, characterizes the random variable $X_{\mathfrak{G}}$. It will be the key to generalizing the definition of $X_{\mathfrak{G}}$ from σ–algebras which are generated by a finite or countable partition, $\Omega = G_1, \biguplus G_2 \biguplus \cdots$ of $\mathfrak{F}$–measurable sets G_j to arbitrary sub–σ–algebras of $\mathfrak{F}$.

We will find for any σ–algebra $\mathfrak{G} \subseteq \mathfrak{F}$ and nonnegative or integrable X a $\mathfrak{G}$–measurable $X_{\mathfrak{G}}$ which satisfies formula (5.8). Since this formula yields matching "averages"

$$\frac{1}{P(G)} \int_G X_{\mathfrak{G}}\, dP \;=\; \frac{1}{P(G)} \int_G X\, dP \tag{5.9}$$

for all events $G \in \mathfrak{G}$ which have positive probability, there is hope that this random variable $X_{\mathfrak{G}}$ is the answer to question **(A)** that was raised above. In fact, Theorem 5.6 on p.119 will show that $X_{\mathfrak{G}}$ is the best least–squares estimate of X among all $\mathfrak{G}$–measurable functions. □

Proposition 5.2. *We work under the assumptions of the introduction.*

(1) *Given are a probability space $(\Omega, \mathfrak{F}, P)$ and a finite or infinite sequence $G_1, G_2, \dots$ of elements of $\mathfrak{F}$ which constitute a partition of Ω. We write J for the finite or infinite index set for this sequence and J_1 for the set of those indices j such that $P(G_j) > 0$.*

(2) *Let $\mathfrak{G} := \sigma\{G_j : j \in J\}$. For an integrable or nonnegative random variable X on $(\Omega, \mathfrak{F}, P)$, we define again the $\mathfrak{G}$–measurable random variable $X_{\mathfrak{G}}$ via (5.5):*

$$X_{\mathfrak{G}}(\omega) \;:=\; \sum_{j \in J_1} \frac{1}{P(G_j)}\, E\left[X \mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j}(\omega).$$

Then formula (5.8) of the introduction, $\displaystyle\int_G X_{\mathfrak{G}} dP \;=\; \int_G X\, dP,$ *holds true for all $G \in \mathfrak{G}$.*

PROOF: ⋆ We employ the ILMD Mehod.

Step 1. If $X = \mathbf{1}_A$ for some $A \in \mathfrak{F}$, then for each $k \in J$,

$$\begin{aligned}
\int_{G_k} X_{\mathfrak{G}}\, dP \;&=\; \sum_{j \in J_1} \frac{1}{P(G_j)} \int_{G_k} E\left[\mathbf{1}_A \mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j}\, dP \\
&=\; \sum_{j \in J_1} \frac{1}{P(G_j)} \int_{G_k} P(A \cap G_j) \cdot \mathbf{1}_{G_j}\, dP \\
&=\; \sum_{j \in J_1} \frac{1}{P(G_j)} P(A \cap G_j) \cdot P(G_k \cap G_j)\, dP.
\end{aligned}$$

But the G_j are disjoint, thus $P(G_k \cap G_j) = 0$ for $k \neq j$, and $P(G_k \cap G_j) = P(G_k)$ for $k = j$. Therefore, all terms in the sum except the one for $j = k$ vanish, and we are left with

$$\begin{aligned}
\int_{G_k} X_{\mathfrak{G}}\, dP \;&=\; \frac{1}{P(G_k)} P(A \cap G_k) \cdot P(G_k)\, dP \;=\; P(A \cap G_k) \\
&=\; \int_{G_k} \mathbf{1}_A\, dP \;=\; \int_{G_k} X\, dP.
\end{aligned}$$

Since all elements of $\mathfrak{G}$ are a finite or infinite union $G_{j_1} \uplus G_{j_2} \uplus \cdots$ of the sets G_j, this last result extends for arbitrary events $G \in \mathfrak{G}$ to

$$\int_G X_{\mathfrak{G}} \, dP = \int_G X \, dP.$$

Step 2. If $X = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i}$ for some $m \in \mathbb{N}$, $A_1, \dots, A_m \in \mathfrak{F}$, and nonnegative $\alpha_1, \dots \alpha_m$, we obtain by first using the definition of $X_{\mathfrak{G}}$, then linearity of expectations, then using the result obtained in **Step 1** for each random variable $\mathbf{1}_{A_i}$, then linearity of the integral,

$$\begin{aligned}
\int_G X_{\mathfrak{G}} \, dP &\overset{(5.5)}{=} \int_G \sum_{j \in J_1} \frac{1}{P(G_j)} E\left[X \mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j} \, dP = \int_G \sum_{j \in J_1} \frac{1}{P(G_j)} E\left[\sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i} \mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j} \, dP \\
&= \sum_{i=1}^{m} \alpha_i \int_G \left(\sum_{j \in J_1} \frac{1}{P(G_j)} E\left[\mathbf{1}_{A_i} \mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j} \right) dP \overset{(5.5)}{=} \sum_{i=1}^{m} \alpha_i \int_G \left(\mathbf{1}_{A_i}\right)_{\mathfrak{G}} dP \\
&\overset{\text{Step 1}}{=} \sum_{i=1}^{m} \alpha_i \int_G \mathbf{1}_{A_i} \, dP = \int_G \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i} \, dP = \int_G X \, dP.
\end{aligned}$$

This proves the proposition for all simple functions.

Step 3: Monotone convergence allows us to extend the result from simple functions to any nonnegative random variable.

Step 4: If X is integrable then we apply the result obtain step 3 to X^+ and X^-. It follows from $\int_G X_{\mathfrak{G}}^+ dP = \int_G X^+ dP$, and $\int_G X_{\mathfrak{G}}^- dP = \int_G X^- dP$, that

$$\begin{aligned}
\int_G X_{\mathfrak{G}} dP &= \int_G \left(X^+ - X^-\right)_{\mathfrak{G}} dP = \int_G X_{\mathfrak{G}}^+ dP - \int_G X_{\mathfrak{G}}^- dP \\
&= \int_G X^+ dP - \int_G X^- dP = \int_G (X^+ - X^-) dP = \int_G X dP
\end{aligned}$$

holds true for all $G \in \mathfrak{G}$. ■

5.3 Conditional Expectations in the General Setting

What we have seen in the previous section was just of a motivational nature. We are ready now to attack the general case of an arbitrary sub–σ–algebra $\mathfrak{G}$ of $\mathfrak{F}$.

Theorem 5.3 (Existence Theorem for Conditional Expectations)**.**

Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and let $\mathfrak{G}$ be a sub–σ–algebra of $\mathfrak{F}$.
(I) *Let X be a nonnegative random variable on $(\Omega, \mathfrak{F}, P)$, let ν be the measure $A \mapsto \int_A X dP$ on $\mathfrak{F}$. Let $P_{\mathfrak{G}} := P\big|_{\mathfrak{G}}$ be the restriction of P to $\mathfrak{G}$, and let $\nu_{\mathfrak{G}} := \nu\big|_{\mathfrak{G}}$ be the restriction of ν to $\mathfrak{G}$. (See Definition 2.21 on p.24.) In other words, $P_{\mathfrak{G}}$ and $\nu_{\mathfrak{G}}$ are the set functions defined as*

$$P_{\mathfrak{G}}(G) = P(G), \qquad \nu_{\mathfrak{G}}(G) = \nu(G), \qquad (G \in \mathfrak{G}).$$

Then $P_{\mathfrak{G}}$ is a probability measure and $\nu_{\mathfrak{G}}$ is a measure on the measurable space $(\Omega, \mathfrak{G})$ such that $\nu_{\mathfrak{G}} \ll P_{\mathfrak{G}}$. The Radon–Nikodým derivative

$$E\big[X \mid \mathfrak{G}\big] \;:=\; \frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}$$

plays the role of $X_{\mathfrak{G}}$ in formula (5.8) on p.111 in the following sense:
$E\big[X \mid \mathfrak{G}\big]$ is $\mathfrak{G}$–measurable and satisfies the partial averaging property:

(5.10) $$\int_G E\big[X \mid \mathfrak{G}\big]\,dP \;=\; \int_G X\,dP \quad \text{for all } G \in \mathfrak{G}.$$

(II) *Let X be an integrable random variable on $(\Omega, \mathfrak{F}, P)$. The random variables $E\big[X^+ \mid \mathfrak{G}\big]$ and $E\big[X^- \mid \mathfrak{G}\big]$ exist according to (**I**). Define*

$$E\big[X \mid \mathfrak{G}\big] \;:=\; E\big[X^+ \mid \mathfrak{G}\big] \;-\; E\big[X^- \mid \mathfrak{G}\big].$$

Then, $E\big[X \mid \mathfrak{G}\big]$ satisfies formula (5.10).

PROOF: ⋆

PROOF of **I**: It is trivial that $\nu_{\mathfrak{G}}$ and $P_{\mathfrak{G}}$ are measures on the shrunken domain $\mathfrak{G}$, since they assign the same function values $\nu(G)$ and $P(G)$ to their arguments G as ν and P.

We now show that $\nu_{\mathfrak{G}} \ll P_{\mathfrak{G}}$, i.e., if $G \in \mathfrak{G}$ such that $P_{\mathfrak{G}}(G) = 0$, then $\nu_{\mathfrak{G}}(G) = 0$. We obtain this from $\nu \ll P$ (see prop.4.20 on p.99) as follows.

$$P_{\mathfrak{G}}(G) \,=\, 0 \;\Rightarrow\; P(G) \,=\, P_{\mathfrak{G}}(G) \,=\, 0 \;\Rightarrow\; \nu(G) \,=\, 0 \;\Rightarrow\; \nu_{\mathfrak{G}}(G) \,=\, \nu(G) \,=\, 0.$$

The Radon–Nikodým theorem then guarantees the existence of the Radon–Nikodým derivative $\frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}$, determined uniquely P–a.s. [27] We decide to name it $E\big[X \mid \mathfrak{G}\big]$ rather than $\frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}$.
The next point is subtle and very important. Since the measures $\nu_{\mathfrak{G}}$ and $P_{\mathfrak{G}}$ live on the measurable space $(\Omega, \mathfrak{G})$ the Radon–Nikodým theorem applies to this space, thus $E\big[X \mid \mathfrak{G}\big]$ is $\mathfrak{G}$–measurable and not just $\mathfrak{F}$–measurable!

Next, we prove formula (5.10). Let $G \in \mathfrak{G}$. Since the function $\omega \mapsto E\big[X \mid \mathfrak{G}\big](\omega)\mathbf{1}_G(\omega)$ is $\mathfrak{G}$–measurable, it follows from $P_{\mathfrak{G}} = P\big|_{\mathfrak{G}}$ that

(5.11) $$\int_G E\big[X \mid \mathfrak{G}\big]\,dP \;=\; \int E\big[X \mid \mathfrak{G}\big]\mathbf{1}_G\,dP \;=\; \int E\big[X \mid \mathfrak{G}\big]\mathbf{1}_G\,dP_{\mathfrak{G}} \;=\; \int_G E\big[X \mid \mathfrak{G}\big]\,dP_{\mathfrak{G}}.$$

(See Exercise 4.6 on p.105 for the second equation.) Further,

(5.12) $$E\big[X \mid \mathfrak{G}\big] \;=\; \frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}, \quad \text{i.e.,} \;\; E\big[X \mid \mathfrak{G}\big]\,dP_{\mathfrak{G}} \;=\; \frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}\,dP_{\mathfrak{G}} \;=\; d\nu_{\mathfrak{G}}.$$

We obtain from equations (5.11) and (5.12) that

$$\int_G E\big[X \mid \mathfrak{G}\big]\,dP \;=\; \int_G d\nu_{\mathfrak{G}} \;=\; \nu_{\mathfrak{G}}(G) \;=\; \nu(G) \;=\; \int_G X\,dP$$

[27] For the a.s. uniqueness of the Radon–Nikodým derivative see Remark 4.27 on p.97.

The equation next to the last holds since the set functions $\nu_{\mathfrak{G}} = \nu\big|_{\mathfrak{G}}$ and ν are identical for arguments $G \in \mathfrak{G}$

PROOF of **II** (Outline): Formula (5.10) holds for X^+ and X^-. It is a straightforward exercise to show the validity of (5.10) from the linearity of the integral. ■

Remark 5.3. We state once more that the partial averaging property (5.10) determines the $\mathfrak{G}$–measurable random variable $E\big[X \mid \mathfrak{G}\big]$ P–a.e. in the following sense. If X^* is another $\mathfrak{G}$–measurable random variable such that

$$\int_G X dP \;=\; \int_G X^* dP \quad \text{for all } G \in \mathfrak{G}\,,$$

then $P\{X^* \neq E\big[X \mid \mathfrak{G}\big]\} \;=\; 0.$ □

This last remark allows us to make the following definition (see SCF2 Definition 2.3.1).

Definition 5.1 (Conditional Expectation w.r.t a sub–σ–algebra)**.**

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X a nonnegative or integrable random variable.

For a sub–σ–algebra $\mathfrak{G}$ of $\mathfrak{F}$ we call <u>any**(!)**</u> random variable X^* that satisfies

(a) **(Measurability):** X^* is $\mathfrak{G}$–measurable,

(b) **$\mathfrak{G}$–Partial averaging** or **Partial averaging:**

$$\int_G X^*\, dP \;=\; \int_G X\, dP \;\; \text{for all } G \in \mathfrak{G}, \tag{5.13}$$

a **conditional expectation of X with respect to $\mathfrak{G}$**.

In most cases it does not matter which version X^* that satisfies **(a)** and **(b)** is chosen. It is customary to let the symbol $E\big[X \mid \mathfrak{G}\big]$ denote any such X^* and refer to it as <u>**the**</u> conditional expectation of X with respect to $\mathfrak{G}$.

If Z is another random variable on $(\Omega, \mathfrak{F}, P)$ then $\sigma(Z) \subseteq \mathfrak{F}$, thus $E\big[X \mid \sigma(Z)\big]$ is defined. In this case we will generally use the notation

$$E\big[X \mid Z\big] \;:=\; E\big[X \mid \sigma(Z)\big].$$

We call $E\big[X \mid Z\big]$ the **conditional expectation of X with respect to Z.** □

Remark 5.4. We can think of $E[X \mid \mathfrak{G}]$ as an estimate of X based on only the information that is available in $\mathfrak{G}$. The collecton of averages

$$\frac{1}{P(G)} \int_G X\, dP, \quad \text{where } \; G \in \mathfrak{G} \text{ and } P(G) > 0,$$

is sufficient to represent all stochastically relevant information for the $\mathfrak{G}$–measurable $E\big[X \mid \mathfrak{G}\big]$. The word "partial" in "partial averaging" indicates that those averages only are a part of

$$\frac{1}{P(A)} \int_A X\, dP, \quad \text{where } \; A \in \mathfrak{F} \text{ and } P(A) > 0\,.$$

This larger collection constitutes the stochastically relevant information for X itself.

Partial averaging makes it plausible that $E[X|\mathfrak{G}]$ is a well chosen estimate of X since all its averages over sets in $\mathfrak{G}$ match those of X. The larger $\mathfrak{G}$, the better an estimate for X we obtain.

Consider in particular the case of the introduction 5.1 to this chapter on p.109 where $\mathfrak{G}$ was generated by a partitioning sequence $\Omega = G_1 \uplus G_2 \uplus \cdots$. In that case,

$$E[X \mid \mathfrak{G}](\omega) \;=\; \sum_{j \in J_1} \frac{1}{P(G_j)}\, E\left[X \mathbf{1}_{G_j}\right] \cdot \mathbf{1}_{G_j}(\omega), \tag{5.14}$$

where J_1 is the set of indices for which $P(G_j > 0)$. See formula (5.5) on p.111. So the estimate $E[X \mid \mathfrak{G}]$ of X is constant on each atom G_j of $\mathfrak{G}$. Moving to a partition with more sets with smaller probabilities will improve this estimate. □

Remark 5.5 (Composition of conditional expectations). ⋆ According to Proposition 5.1 (Doob Composition Lemma) on p.106 the $\sigma(Z)$–$\mathfrak{B}_1$ measurable function on Ω,

$$E\big[X \mid Z\big] : \Omega \to \mathbb{R}, \qquad \omega \mapsto E\big[X \mid Z\big](\omega),$$

can be written as a composite function

$$E\big[X \mid Z\big] \;=\; g \circ Z, \tag{5.15}$$

where $Z : z \mapsto g(z)$ is $\mathfrak{B}^1$–$\mathfrak{B}^1$ measurable. Very confusingly it is common to write

$$E[X \;\mid\; Z = \cdot\,] \;:\; z \;\mapsto\; E[X \;\mid\; Z = z] \tag{5.16}$$

for this function $g(z)$. With this notation the functional relationship $E\big[X \mid Z\big](\omega) \;=\; g\big(Z(\omega)\big)$ which is obtained by replacing the dummy variable z with the function value $Z(\omega)$, reads

$$E[X \mid Z](\omega) \;= E[X \mid Z = \cdot\,]\big(Z(\omega)\big) \;=\; E[X \mid Z = Z(\omega)]\,. \tag{5.17}$$ □

The following concrete example shows how to compute a conditional expectation for a σ–algebra which is generated by a finite partition of Ω.

Example 5.1. Let $\Omega :=]0,6], \mathfrak{F} := \mathfrak{B}(]0,6]) :=$ all Borel sets of $]0,6]$, $P :=$ uniform probability on $]0,6]$, i.e.,

$$P(]a,b]) \;:=\; \frac{b-a}{6} \;\text{ for all } 0 \;<\; a \;\leq\; b \;\leq\; 6.$$

Let $\mathfrak{G} := \sigma(]0,2],]2,6])$, and let X be the random variable defined by $X(\omega) := 5\omega$.

We compute the conditional expectation $\omega \mapsto E[X \mid \mathfrak{G}](\omega)$ as follows. According to Proposition 5.2 on p.112,

$$E[X \mid \mathfrak{G}](\omega) \;=\; \sum_{j=1,2} \frac{1}{P(G_j)}\, E[X \mathbf{1}_{G_j}](\omega) \;=\; \begin{cases} \frac{1}{P(G_1)}\, E\left[X \mathbf{1}_{G_1}\right] & \text{if } \omega \in G_1, \\ \frac{1}{P(G_2)}\, E\left[X \mathbf{1}_{G_2}\right] & \text{if } \omega \in G_2. \end{cases} \tag{5.18}$$

We have $P(G_1) = \dfrac{2}{6}$, $P(G_2) = \dfrac{4}{6}$, $\displaystyle\int_{]a,b]} X\,dP = \frac{5}{6}\int_a^b x dx = \frac{5}{6}\left(\frac{b^2}{2} - \frac{a^2}{2}\right)$ for all $0 < a \le b \le 6$.
Thus the solution is

$$\begin{aligned} 0 < \omega \le 2 \ &\Rightarrow\ E[X \mid \mathfrak{G}](\omega) = \frac{6}{2}\cdot\frac{5}{6}\left(\frac{2^2}{2} - \frac{0^2}{2}\right) = \frac{5}{2}(2 - 0) = 5,\\ 2 < \omega \le 6 \ &\Rightarrow\ E[X \mid \mathfrak{G}](\omega) = \frac{6}{4}\cdot\frac{5}{6}\left(\frac{6^2}{2} - \frac{2^2}{2}\right) = \frac{5}{4}(18 - 2) = 20, \ \text{i.e.,}\\ E[X \mid \mathfrak{G}] &= 5\cdot\mathbf{1}_{]0,2]} + 20\cdot\mathbf{1}_{]2,6]}\,. \end{aligned}$$

We are done, but here is a sanity check. It should be true that $E\big[E[X \mid \mathfrak{G}]\big] = E[X]$. We have

$$\begin{aligned} E\big[E[X \mid \mathfrak{G}]\big] &= 5\cdot P(]0,2]) + 20\cdot P(]2,6]) = \frac{2\cdot 5}{6} + \frac{4\cdot 20}{6} = \frac{90}{6} = 15,\\ E[X] &= \int_\Omega X\,dP = 5\int_0^6 x\,\frac{dx}{6} = \frac{5}{6}\left(\frac{6^2}{2} - \frac{0^2}{2}\right) = \frac{5}{6}\cdot 18 = 15. \ \square \end{aligned}$$

Theorem 5.4 (Monotony of Conditional Expectations)**.** *Let X and Y be two random variables on a probability space $(\Omega, \mathfrak{F}, P)$ which both are integrable or nonnegative. and let $\mathfrak{G}$ be a sub–σ–algebra of $\mathfrak{F}$.*

(5.19) *If* $X \le Y$ *a.s.* *then* $E[X \mid \mathfrak{G}] \le E[Y \mid \mathfrak{G}]$ *a.s.*

PROOF: ⋆ The proof is a repetition of that of Theorem 4.8 on p.85.

$$\text{Let } A := \{E[X \mid \mathfrak{G}] > E[Y \mid \mathfrak{G}]\} \quad\text{and}\quad A_n := \left\{E[X \mid \mathfrak{G}] > E[Y \mid \mathfrak{G}] + \frac{1}{n}\right\};\ \ (n \in \mathbb{N})\,.$$

We will prove (5.19) by showing that the assumption $P(A) > 0$ implies $\int_{A_n} X\,dP > \int_{A_n} Y\,dP$ for large n. This contradicts $X \le Y$ a.s., since that assumption implies $\int_B X dP \le \int_B Y dP$ for all $B \in \mathfrak{F}$.
The sets A_n are $\mathfrak{G}$–measurable, thus partial averaging implies that

$$(5.20)\qquad \int_{A_n} X\,dP = \int_{A_n} E[X \mid \mathfrak{G}]\,dP \quad\text{and}\quad \int_{A_n} Y\,dP = \int_{A_n} E[Y \mid \mathfrak{G}]\,dP\,.$$

Assume to the contrary that $P(A) > 0$. Since $A_n \uparrow A$, $P(A_n) \uparrow P(A)$. See Proposition 4.7 (Continuity properties of measures) on p.58. Thus there exists $\gamma > 0$ such that $P(A) = 2\gamma$ and hence some $n \in \mathbb{N}$ such that $P(A_n) \ge \gamma$. Since $E[X \mid \mathfrak{G}] > E[Y \mid \mathfrak{G}] + \frac{1}{n}$ on all of A_n,

$$\begin{aligned} \int_{A_n} X\,dP &\overset{(5.20)}{=} \int_{A_n} E[X \mid \mathfrak{G}]\,dP \ge \int_{A_n}\left(E[Y \mid \mathfrak{G}] + \frac{1}{n}\right)dP\\ &= \int_{A_n} E[Y \mid \mathfrak{G}]\,dP + \frac{1}{n}P(A_n) \ge \int_{A_n} E[Y \mid \mathfrak{G}]\,d\mu + \frac{\gamma}{n}\\ &> \int_{A_n} E[Y \mid \mathfrak{G}]\,dP = \int_{A_n} Y\,dP\,. \end{aligned}$$

As mentioned earlier this contradicts $X \leq Y$ a.s., and we conclude that $P(A) = 0$. Thus

$$E[X \mid \mathfrak{G}] \;\leq\; E[Y \mid \mathfrak{G}] \;\text{ a.s.}. \quad \blacksquare$$

The following is SCF2 Theorem 2.3.2 which I reproduce here essentially unaltered. In particular I use his phrase "Taking out what is known". It sounds awkward to me, but I would not know a better formulation: It expresses the fact that a $\mathfrak{G}$–measurable random variable (i.e., one for which $\mathfrak{G}$ contains all its stochastically relevant information,) can be pulled out of a conditional expectation $E[\cdots \mid \mathfrak{G}]$ the same way a constant number can be pulled out of an ordinary expectation $E[\dots]$.

Note that all equations and inequalities are uderstood to only hold P–a.s., since conditional expectations are defined only P–a.s.!

Theorem 5.5. *Let $(\Omega, \mathfrak{F}, P)$ be a probability space. let $\mathfrak{G}$ be a sub–σ–algebra of $\mathfrak{F}$.*

(a) ***(Linearity of conditional expectations)*** *If X and Y are integrable random variables and c_1 and c_2 are constants, then*

(5.21) $$E[c_1 X + c_2 Y | \mathfrak{G}] \;=\; c_1 E[X|\mathfrak{G}] + c_2 E[Y|\mathfrak{G}].$$

This equation also holds if we assume that X and Y are nonnegative (rather than integrable) and c_1 and c_2 are positive, although both sides may be $+\infty$.

(b) ***(Taking out what is known)*** *If X and Y are integrable random variables, if XY is integrable, and if X is $\mathfrak{G}$–measurable, then*

(5.22) $$E[X \cdot Y | \mathfrak{G}] \;=\; X \cdot E[Y|\mathfrak{G}].$$

This equation also holds if we assume that X is positive and Y is nonnegative (rather than integrable), although both sides may be $+\infty$.

(c) ***(Iterated conditioning)*** *If $\mathfrak{H}$ is a sub–σ–algebra of $\mathfrak{G}$ ($\mathfrak{H}$ contains less information than $\mathfrak{G}$), and if X is an integrable random variable, then*

(5.23) $$E\big[E[X|\mathfrak{G}] \,\big|\, \mathfrak{H}\big] \;=\; E[X|\mathfrak{H}].$$

This equation also holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

(d) ***(Independence)*** *If X is integrable and independent of $\mathfrak{G}$, then*

$$E[X|\mathfrak{G}] \;=\; E[X]. \tag{5.24}$$

This equation also holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

(e) ***(Conditional Jensen's inequality)*** *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, (see Definition 2.25 (Concave-up and convex functions) on p.31) and that X is integrable. Then*

$$\varphi\big(E[X \mid \mathfrak{G}]\big) \;\leq\; E[\varphi \circ (X) \mid \mathfrak{G}]. \tag{5.25}$$

PROOF: See the SCF2 text. ■

Proposition 5.3.

Let $(\Omega, \mathfrak{F}, P)$ be a probability space, $\mathfrak{G}$ a sub–σ–algebra of $\mathfrak{F}$, and X a nonnegative or integrable random variable. Then

$$E\big[E[X|\mathfrak{G}]\big] \;=\; E[X]. \tag{5.26}$$

PROOF: The proof is left as exercise 5.1. See p.121. ■

Note the significance of formula (5.26). It states that $E[X|\mathfrak{G}]$ is an **unbiased estimator** of X.

Example 5.2. Here is an example for the Jensen inequality for conditional expectations. Let W_t be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$, let $g(x) := 2x^6 - 8$, and let $Y_t := g(W_t)$. Then g is convex (concave–up). Thus, for any $t, h \geq 0$,

$$Y_t \;= g(W_t) \overset{\textbf{(a)}}{=} g\big(E[W_{t+h} \mid \mathfrak{F}_t]\big) \overset{\textbf{(b)}}{\leq} E\big[g\big(W_{t+h}\big) \mid \mathfrak{F}_t\big] \;=\; E\big[Y_{t+h} \mid \mathfrak{F}_t\big]\,.$$

In the above, **(a)** holds because W_t is a martingale, and **(b)** follows from the conditional form of Jensen's inequality. □

Theorem 5.6. *Let X be a square–integrable random variable on a probability space $(\Omega, \mathfrak{F}, P)$, i.e.,*

$$E[X^2] \;<\; \infty\,.$$

Let $\mathfrak{G}$ be a sub–σ–algebra of $\mathfrak{F}$. Then

$E[\,X \mid \mathfrak{G}\,]$ is the best possible estimate of X, since it minimizes the distance to X in the following sense. If $\mathscr{A} \;=\; \{\hat{X} : \hat{X}$is $\mathfrak{G}$–measurable and $E[\hat{X}^2 < \infty]\}$ then

$$E\left[\big(X \;-\; E[X \mid \mathfrak{G}]\big)^2\right] \;=\; \min\Big(E[(X - \hat{X})^2] \;:\; \hat{X} \in \mathscr{A}\Big)\,. \tag{5.27}$$

*In other words, $E[X \mid \mathfrak{G}]$ is the optimal **least squares estimate** of X among all $\mathfrak{G}$–measurable and square–integrable random variables.*

PROOF: ⋆ We first prove that

$$E\left[\left(X - E[X \mid \mathfrak{G}]\right)^2 \mid \mathfrak{G}\right] \leq E[(X - Z)^2 \mid \mathfrak{G}], \quad \text{for all } Z \in \mathscr{A}. \tag{5.28}$$

Let

$$\begin{aligned} X_1 &:= E\left[\left(X - E[X \mid \mathfrak{G}]\right)^2 \mid \mathfrak{G}\right]. \\ X_2 &:= E\left[\left(E[X \mid \mathfrak{G}] - Z\right)^2 \mid \mathfrak{G}\right]. \\ X_3 &:= E\left[\left(X - E[X \mid \mathfrak{G}]\right)\left(E[X \mid \mathfrak{G}] - Z\right) \mid \mathfrak{G}\right]. \end{aligned}$$

Then

$$\begin{aligned} E\left[(X - Z)^2 \mid \mathfrak{G}\right] &= E\left[\left((X - E[X \mid \mathfrak{G}]) + (E[X \mid \mathfrak{G}] - Z)\right)^2 \mid \mathfrak{G}\right] \\ &= X_1 + X_2 + 2X_3 \geq X_1 + 2X_3, \end{aligned} \tag{5.29}$$

(The inequality results from $X_2 \geq 0$ and the monotony of conditional expectations.)
We will show that $X_3 = 0$.

$$\begin{aligned} X_3 = \; & E[(X \cdot E[X \mid \mathfrak{G}]) \mid \mathfrak{G}] - E[\left(E[X \mid \mathfrak{G}] \cdot E[X \mid \mathfrak{G}]\right) \mid \mathfrak{G}] \\ & - E[X \cdot Z \mid \mathfrak{G}] + E\left[\left(E[X \mid \mathfrak{G}] \cdot Z\right) \mid \mathfrak{G}\right]. \end{aligned} \tag{5.30}$$

We apply the "pull out what is known" rule to terms #1 and #3 of (5.30) and obtain

$$\begin{aligned} E[(X \cdot E[X \mid \mathfrak{G}]) \mid \mathfrak{G}] &= E[X \mid \mathfrak{G}] \cdot E[X \mid \mathfrak{G}], \\ E[X \cdot Z \mid \mathfrak{G}] &= Z \cdot E[X \mid \mathfrak{G}]. \end{aligned}$$

For terms #2 and #4 of (5.30) we observe that $E[X \mid \mathfrak{G}] \cdot E[X \mid \mathfrak{G}]$ and $E[X \mid \mathfrak{G}] \cdot Z$ are $\mathfrak{G}$–measurable random variables, thus $E[\ldots \mid \mathfrak{G}]$ has no effect, thus

$$\begin{aligned} E[\left(E[X \mid \mathfrak{G}] \cdot E[X \mid \mathfrak{G}]\right) \mid \mathfrak{G}] &= E[X \mid \mathfrak{G}] \cdot E[X \mid \mathfrak{G}], \\ E\left[\left(E[X \mid \mathfrak{G}] \cdot Z\right) \mid \mathfrak{G}\right] &= E[X \mid \mathfrak{G}] \cdot Z. \end{aligned}$$

We substitute those four indentities into formula (5.30) and obtain

$$X_3 = E[X \mid \mathfrak{G}] \cdot E[X \mid \mathfrak{G}] - E[X \mid \mathfrak{G}] \cdot E[X \mid \mathfrak{G}] - Z \cdot E[X \mid \mathfrak{G}] + E[X \mid \mathfrak{G}] \cdot Z.$$

This proves $X_3 = 0$. It follows from (5.29) that $E\left[(X - Z)^2 \mid \mathfrak{G}\right] \geq X_1$, i.e.,

$$E\left[(X - Z)^2 \mid \mathfrak{G}\right] \geq E\left[\left(X - E[X \mid \mathfrak{G}]\right)^2 \mid \mathfrak{G}\right].$$

We have shown that (5.28) is true.
Formula (5.27) now is obtained easily. Let $Z \in \mathscr{A}$. Since

$$E\big[E[Y \mid \mathfrak{G}]\big] = E[Y] \quad \text{and} \quad Y_1 \leq Y_2 \text{ a.s.} \Rightarrow E[Y_1] \leq E[Y_2]$$

for any integrable or non–negative random variables Y, Y_1, Y_2, it follows from (5.28) that

$$E\left[(X - Z)^2\right] = E\left[E[(X - Z)^2 \mid \mathfrak{G}]\right] \geq E\left[E[(X - E[X \mid \mathfrak{G}])^2 \mid \mathfrak{G}]\right].$$

But this is the assertion of formula (5.27). ■

The next theorem, which Shreve calls the **Independence Lemma**, can be very useful to actually compute conditional expectations. This is SCF2 Lemma 2.3.4.

Theorem 5.7 (Independence Lemma)**.**

Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and let $\mathfrak{G}$ be a sub–σ–algebra of $\mathfrak{F}$. Assume that
- *the random variables $X_1, \dots, X_K$ are $\mathfrak{G}$–measurable,*
- *the random variables $Y_1, \dots, Y_L$ are independent of $\mathfrak{G}$.*

Let $f(x_1, \dots, x_K, y_1, \dots, y_L)$ be a function of the dummy variables $x_1, \dots, x_K$ and $y_1, \dots, y_L$. Let $g(x_1, \dots, x_K)$ be the function

$$g(x_1, \dots, x_K) \;=\; Ef(x_1, \dots, x_K, Y_1, \dots, Y_L).$$

Then $$E[f(X_1, \dots, X_K, Y_1, \dots, Y_L) | \mathfrak{G}] \;=\; g(X_1, \dots, X_K).$$

PROOF: See the outline given in the text. ■

5.4 Exercises for Ch.5

Exercise 5.1. Prove prop.5.3 on p.119 of this document: Let $(\Omega, \mathfrak{F}, P)$ be a probability space, $\mathfrak{G}$ a sub–σ–algebra of $\mathfrak{F}$, and X a nonnegative or integrable random variable. Then

$$E[\,E[X \mid \mathfrak{G}]\,] \;=\; E[X]\,. \;\square$$

6 Brownian Motion

Key properties of Brownian Motion will be that this process is both a martingale and a Markov process. We start out this chapter by discussing those two concepts. We follow closely the SCF2 text.

6.1 Martingales and Markov Processes

Introduction 6.1. We will see that the pricing of stock options and other financial derivatives with the help of tools from stochastic calculus fundamentally depends on the following.

(1) Consider the filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$, in which the filtration element $\mathfrak{F}_t$ represents the financial market information that accrued until the time t. Then the "real world" probability P can be replaced by a "risk–neutral" probability $\widetilde{P}$ which is characterized as follows: Let S_t be the price of a stock at time t. How much would we be willing to pay at $t = 0$ for the asset if the bank pays interest at a rate $R(s)$ at time s? Certainly not the full amount S_t, since, if we invest S_t dollars in the bank instead, then compound interest would grow that money to $e^{\int_0^t R(s)ds} S_t$. Rather, the fair price of the stock at $t = 0$ would be the discounted stock price, $M_t := e^{-\int_0^t R(s)ds} S_t$.
The risk–neutral world, the one governed by $\widetilde{P}$, is characterized as follows: the future development of the discounted price process $M_t(\omega)$ of the stock shows no trend that can be inferred from the information $\mathfrak{F}_t$ that is available at time t.

In other words, the best possible prediction of this process at a future time $t + h$ in risk–neutral terms is its present state, M_t:

$$\text{Best estimate of } M_{t+h} \text{ given } \mathfrak{F}_t \;=\; M_t \qquad (h > 0)\,. \tag{6.1}$$

We have seen in Theorem 5.6 on p.119 that the best estimate based on the information contained in $\mathfrak{F}_t$ is the conditional expectation w.r.t. $\mathfrak{F}_t$. Thus (6.1) is made mathematically precise by the formula

$$\widetilde{E}[\, M_{t+h} \mid \mathfrak{F}_t] \;=\; M_t, \qquad (h > 0). \tag{6.2}$$

Here $\widetilde{E}[\dots]$ is the expectation $\int \dots d\widetilde{P}$ with respect to risk–neutral probability $\widetilde{P}$.
Stochastic processes M_t that satisfy (6.2) are called martingales. We will discuss some of their properties.

(2) The future development of any function $\varphi(M_t)$ of that discounted price process M_t, but also of other processes such as stock price S_t itself, does not depend on the entire past information, i.e., not on all of $\mathfrak{F}_t$. Rather the knowledge of the present information concerning those processes will be sufficient. The formal mathematical definition is that of a Markov process, a process X_t which satisfies

$$E[\varphi\big(X_{t+h}\big)|\mathfrak{F}_t] \;=\; E[\varphi\big(X_{t+h}\big)|X_t]\,, \qquad (h > 0) \tag{6.3}$$

for all reasonable, i.e., nonnegative and measurable, functions $\varphi(x)$ □

We now give the formal definition of a martingale.

Definition 6.1 (Martingale)**.** Let $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\in I}\,, P)$ be a filtered probability space.

We assume that I is the index set of an extended real–valued, adapted, continuous time or discrete time process X that satisfies $E[|X_t|] < \infty$ for all t. We call X

(a) a **martingale** if $E[X_t \mid \mathfrak{F}_s] = X_s$ a.s., for all $s \leq t$ such that $s, t, \in I$,

(b) a **submartingale** if $E[X_t \mid \mathfrak{F}_s] \geq X_s$ a.s., for all $s \leq t$ such that $s, t, \in I$,

(c) a **supermartingale** if $E[X_t \mid \mathfrak{F}_s] \leq X_s$ a.s., for all $s \leq t$ s.t. $s, t, \in I$. □

Remark 6.1. A simple proof by induction shows that, if $I = \mathbb{N}$ then

(a) X is a martingale $\Leftrightarrow$ $E[X_{n+1} \mid \mathfrak{F}_n] = X_n$ a.s., for all $n \in \mathbb{N}$,

(b) X is a submartingale $\Leftrightarrow$ $E[X_{n+1} \mid \mathfrak{F}_n] \geq X_n$ a.s., for all $n \in \mathbb{N}$,

(c) X is a supermartingale $\Leftrightarrow$ $E[X_{n+1} \mid \mathfrak{F}_n] \leq X_n$ a.s., for all $n \in \mathbb{N}$. □

Remark 6.2.

Comparisons on an ω–by–ω basis involving conditional expectations can generally only be asserted to hold almost surely since such conditional expectations only are determined up to a set of probability zero. We will follow the example of Shreve and often drop the "a.e." in such statements. □

Proposition 6.1.

A martingale X satisfies $E[X_s] = E[X_t]$ for any $s, t \in I$.

PROOF:

Let $s < t$. We apply the partial averaging property for conditional expectations. Integration over the set Ω (which belongs to $\mathfrak{F}_s$) results in

$$E[X_t] = \int_\Omega X_t \, dP = \int_\Omega E[X_t \mid \mathfrak{F}_s] \, dP = \int_\Omega X_s \, dP = E[X_s] . \blacksquare$$

The following connection between sums of independent variables and submartingales is worthwhile remembering.

Lemma 6.1. *If X_n are $\mathfrak{F}_n$ adapted and independent, and if $S_n = \sum_{j=1}^{n} X_j$, then*

$$\begin{aligned} E[S_n] \text{ increasing} &\Rightarrow S_n \text{ is a submartingale} , \\ E[S_n] = \text{const} &\Rightarrow S_n \text{ is a martingale} . \end{aligned}$$

PROOF:

$$E[S_{n+k} \mid \mathfrak{F}_n] = S_n + E\left[\sum_{j=n+1}^{n+k} X_j \mid \mathfrak{F}_n\right]$$

(independence of X_j and $\mathfrak{F}_n$)

$$= S_n + E\left[\sum_{j=n+1}^{n+k} X_j\right]$$

$$= S_n + \left(E[S_{n+k}] - E[S_n]\right).$$

Since $E[S_{n+k}] - E[S_n] \geq 0$ for submartingales and $E[S_{n+k}] - E[S_n] = 0$ for martingales, the assertion follows. ■

Definition 6.2 (SCF2 Definition 2.3.6 - Markov Process)**.** Let $(\Omega, \mathfrak{F}, P)$ be a probability space, let T be a fixed positive number, let $(\mathfrak{F}_t)_{t\in[0,T]}$, be a filtration of sub-a-algebras of $\mathfrak{F}$.

Let $X = (X_t)_{t\in[0,T]}$, be an adapted stochastic process for which the codomain Ω' of the random variables $\omega \mapsto X_t(\omega)$ is the real numbers or $\mathbb{R}^n$. (It is thus more appropriate to write $x = X_t(\omega)$ instead of $\omega' = X_t(\omega)$.)

Assume that for all $0 \leq s \leq t \leq T$ and for every nonnegative, Borel–measurable function $f_t : x \mapsto f_t(x)$, one can find another Borel–measurable function $f_s : x \mapsto f_s(x)$ such that

(6.4) $$E[f_t(X_t) \mid \mathfrak{F}_s] = f_s(X_s).$$

Then we call X a **Markov process** (with respect to the filtration $(\mathfrak{F}_t)_{t\in[0,T]}$. □

There is yet another alternate definition of the Markov property which has the advantage of being very well suited to determine in practical terms whether a process is Markov:

Proposition 6.2.

A process X is a Markov process if and only if the following is satisfied.
Let $0 \leq s \leq t \leq T$, and let φ be an arbitrary, nonnegative, Borel–measurable function $x \mapsto \varphi(x)$. Then,

(6.5) $$E[\varphi(X_t)|\mathfrak{F}_s] = E[\varphi(X_t)|X_s].$$

The interpretation is as follows: [28]

The future development of a Markov process does not depend on the past, only on the present.

PROOF: The equivalence of (6.4) and (6.5) is not hard to see.

First, assume that (6.4) holds true. Let φ be nonnegative and Borel–measurable. Setting $f_t(x) := \varphi(x)$ in (6.4), we see that there is a Borel measurable function $x \mapsto f_s(x)$ that satisfies

$$E[\varphi(X_t) \mid \mathfrak{F}_s] = f_s(X_s).$$

[28]https://en.wikipedia.org/wiki/Markov_property

Since the right–hand side is a function of X_s, the same must be true for the left–hand side, i.e., $E[\varphi(X_t) \mid \mathfrak{F}_s]$ is $\sigma(X_s)$–measurable. This yields the first equation in

$$E[\varphi(X_t) \mid \mathfrak{F}_s] = E\big[E[\varphi(X_t) \mid \mathfrak{F}_s] \mid X_s\big] = E[\varphi(X_t) \mid X_s].$$

The second equation follows from the Iterated Conditioning property. See Theorem 5.5 on p.118.

Now assume that (6.5) is satisfied. Let f_t be nonnegative and Borel–measurable and $s \leq t$. Then

$$E[f_t(X_t)|\mathfrak{F}_s] = E[f_t(X_t)|X_s].$$

We argue as before and see that $E[f_t(X_t)|X_s]$ is $\sigma(X_s)$–measurable, since it equals, by definition, $E[f_t(X_t)|\sigma(X_s)]$. We use Doob composition and conclude that we can write this as a function $f_s \circ X_s$ for a suitable Borel measurable function f_s. In other words,

$$E[f_t(X_t)|\mathfrak{F}_s] = f_s \circ X_s.$$

This is formula (6.4). ■

Remark 6.3. If X_t is a real–valued or n–dimensional Markov process, then we apply the previous proposition to the function $\varphi(x) = x$ in the onedimensional case, or the coordinate functions $\varphi(x^{(1)}, \dots, x^{(n)}) = x^{(j)}$. We obtain

$$\begin{aligned} &E[X_{t+h} \mid \mathfrak{F}_t] = E[X_{t+h} \mid X_t]\,; \quad (t, h \geq 0) \qquad \text{one dimensional case}\,,\\ &E[X^{(j)}_{t+h} \mid \mathfrak{F}_t] = E[X^{(j)}_{t+h} \mid X_t]\,; \quad (t, h \geq 0) \qquad n\text{–dimensional case}\ : \end{aligned} \tag{6.6}$$

Conditioning of the position at a future time $t + h$ with respect to the position at time t is equivalent to conditioning with respect to the entire past $\mathfrak{F}_t$ up to time t. □

Proposition 6.3 (Processes with independent increments are Markov). [29]

An $\mathfrak{F}_t$–adapted exended real–valued process with independent increments is Markov.

PROOF: ⋆ The proof can be found in many graduate level books on probability theory, e.g., [3] Bauer, Heinz: Probability Theory. ■

Remark 6.4. The concept of a Markov process also exists for discrete time stochastic processes. Just replace the index set $[0, T]$ with the set I of the countable set of times and adjust the conditions for such indices.

For example, the condition "for all $0 \leq s \leq t$" becomes "for all $s, t \in I$ such that $s \leq t$".

The above applies in particular to random sequences $X_1, X_2, X_3, \dots$. If such a random sequence satisfies one of the equivalent conditions (6.4) or (6.5), then it is customary to speak of a **Markov chain** rather than a time discrete Markov process. □

[29] Adapted from [8] Calin, O., An Introduction to Stochastic Calculus with Applications to Finance

Example 6.1. Here are two informal examples of Markov chains.

(1) The random sequence $X = X_n, n = 0, 1, 2, 3, \dots$, is defined as follows. We assume that $X_0(\omega) = n_0$ for some fixed $n_0 \in \mathbb{Z}$ and all ω, and

$$X_n(\omega) \;=\; \begin{cases} X_{n-1}(\omega) \,+\, 1 & \text{with probability } 0 < p < 1, \\ X_{n-1}(\omega) \,-\, 1 & \text{with probability } 1-p. \end{cases}$$

Clearly, this sequence satisfies (6.5), since the value of $X_n(\omega)$ does not depend on any $X_j(\omega)$ for $j < n-1$. This Markov chain is called a **random walk** on the integers. In the special case $p = q = \frac{1}{2}$ we speak of a **symmetric random walk**. The beginning sections of SCF2 Chapter 3 are about the symmetric random walk.

(2) The price $S = S_n$ of a stock at times $n = 0, 1, 2, 3, \dots$ develops according to the following rules: $S_0(\omega) = s_0$ for some fixed real number s_0 and all (ω), and

$$S_n(\omega) \;=\; \begin{cases} u \cdot S_{n-1}(\omega) & \text{with probability } 0 < p < 1, \\ d \cdot S_{n-1}(\omega) & \text{with probability } 1-p, \end{cases}$$

for two fixed real numbers $0 < d < u$. Typically we will have $d < 1 < u$ so that u signifies an upward movement in stock price and d signifies a downward movement. This sequence also satisfies (6.5), since the value of $S_n(\omega)$ does not depend on the stock price at times prior to $n-1$.
We will examine this process as part of the binomial asset model in Chapter 8 (The Binomial Asset Model). $\square$

6.2 Basic Properties of Brownian Motion

Definition 6.3 (Brownian motion)**.** Given are the index set $I := [0, \infty[$, a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ with $t \in I$, and a stochastic process $W = (W_t)_{t \in I}$.

We call W a **Brownian motion** with respect to the filtration $\mathfrak{F}_t$, if it satisfies the following.

(1) W is adapted to $\mathfrak{F}_t$.
(2) $P\{W_0 = 0\} \;=\; 1$.
(3) $P\{t \mapsto W_t \text{ is continuous for ALL } t\} \;=\; 1$.
(4) Let $0 \le s < t < \infty$. Then the increment $W_t - W_s$ is independent of the σ–algebra $\mathfrak{F}_s$.
(5) Let $0 \le s < t < \infty$. Then $W_t - W_s \sim \mathcal{N}(0, t-s)$, i.e., $W_t - W_s$ is normal with

$$E[W_t \,-\, W_s] \;= 0, \qquad \text{Var}\,[W_t \,-\, W_s] \;= t \,-\, s. \quad \square \tag{6.7}$$

Remark 6.5. If W_t is a Brownian motion with respect to a filtration $\mathfrak{F}_t$ then it also is one with respect to its own filtration $\mathfrak{F}^W = \left(\mathfrak{F}_t^W\right)_{t \in I}$, defined as

$$\mathfrak{F}_t^W \;=\; \sigma(W_s : 0 \le s \le t).$$

In this case we simply speak of Brownian motion without mentioning the filtration $\mathfrak{F}_t^W$.
One can prove that the increments are independent w.r.t.$\mathfrak{F}_t^W$, if

(4') For any finite selection of times $0 \leq t_0 < t_1 < \cdots < t_m < \infty$ the increments $W_{t_1} - W_{t_0},\ W_{t_2} - W_{t_1},\ \dots,\ W_{t_m} - W_{t_{m-1}}$ are independent. □

A proof acceptable to mathematicians that definition 6.3 is free of contradictions and Brownian motion actually exists (the tough part is proven continuity at all times t for the trajectories $t \mapsto W_t(\omega)$ belonging to a set of probability 1) was first given by Norbert Wiener. For this reason you will find books which refer to Brownian motion as **Wiener process**.

The consequences of the next theorem, which we include without proof, are profound. We cannot define integrals

$$\int_{t_0}^{t_1} Z_t(\omega) W_t'(\omega)\, dt\,,$$

since there is no derivative $W_t'(\omega)$.

Theorem 6.1.

For a Brownian motion $(W_t)_{t\geq 0}$ on a filtered probability space $(\Omega, \mathfrak{F}), \mathfrak{F}_t, P)$,

the paths $t \mapsto W_t(\omega)$ are nowhere differentiable with probability 1.

In other words,

$$P\left\{\omega\ :\ \frac{dW_t(\omega)}{dt} \text{ exists for at least one } t \geq 0\right\}\ =\ 0\,.$$

PROOF: Out of scope. A proof can be found, e.g., in [3] Bauer, Heinz: Probability Theory. ■

For the next definition, note the following.

(a) If X be a random variable and $u \in \mathbb{R}$, then the random variable $\omega \mapsto e^{uX(\omega)}$ is nonnegative as an exponential. Thus, its expected value $E\left[e^{uX}\right]$ is always defined (though it may be infinite).

Here is the multidimensional analogue.

(b) If $\vec{X} = (X_1, \dots, X_n)$ is a random vector and $\vec{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$, then the expected value of the random variable

$$\omega\ \mapsto\ e^{\vec{u}\bullet\vec{X}(\omega)}\ =\ \exp\left[\sum_{j=1}^{n} u_j\, X_j(\omega)\right]$$

is always defined (though it may be infinite). In the above, as usual,

$$\text{if}\ \ \vec{a} = (a_1, \dots, a_n) \in \mathbb{R}^n,\ \vec{b} = (b_1, \dots, b_n) \in \mathbb{R}^n, \quad \text{then}\ \ \vec{a} \bullet \vec{b}\ =\ \sum_{j=1}^{n} a_j b_j$$

denotes the standard inner product of $\mathbb{R}^n$

These obervations allow us to make the next definition for any random variable X and any random vector $\vec{X}$.

Definition 6.4 (Moment–generating function)**.** Let X be a random variable and let $\vec{X} = (X_1, \dots, X_n)$ be a random vector on a probability space $(\Omega, \mathfrak{F}, P)$. We define

(6.8) $$\Phi_X : \mathbb{R} \longrightarrow [0, \infty]; \qquad u \mapsto \Phi_X(u) := E\left[e^{uX}\right],$$

(6.9) $$\Phi_{\vec{X}} : \mathbb{R}^n \longrightarrow [0, \infty], \qquad \vec{u} \mapsto \Phi_{\vec{X}}(\vec{u}) := E\left[e^{\vec{u} \bullet \vec{X}}\right].$$

We call Φ_X (resp., $\Phi_{\vec{X}}$), the **moment–generating function** aka **MGF** of X (resp., of $\vec{X}$). In the multidimensional case we also call $\Phi_{\vec{X}}$ the **joint moment–generating function** aka **joint MGF** of $\vec{X}$. □

Proposition 6.4.

Let Z be a normal random variable with mean α and variance σ^2 on a probability space $(\Omega, \mathfrak{F}, P)$. Then its moment–generating function is

(6.10) $$\Phi_Z(u) = e^{\alpha u + \frac{1}{2}\sigma^2 u^2}.$$

PROOF: I was not able to locate the proof in [16] Wackerly, Mendenhall and Scheaffer: Mathematical Statistics with Applications). but it can be found in most text books on probability theory You can find it for the case $\mu = 0$ in the proof of SCF2, Theorem 3.2.1. ■

Proposition 6.5.

Let $W_t, 0 \le t < \infty$ be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. If $s, t \in [0, \infty[$, then

(6.11) $$E[W_t] = 0,$$

(6.12) $$Cov[W_s, W_t] = E[W_s W_t] = \min(s, t).$$

PROOF: See SCF2, ch.3.3.2 ■

Proposition 6.6. ⋆

Let $W_t, 0 \le t < \infty$ be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Let

$$0 \le t_0 < t_1 < \cdots < t_m .$$

Then the covariance matrix for the m–dimensional random vector $\left(W_{t_1}, W_{t_2} \dots, W_{t_m}\right)$ is

(6.13) $$\begin{bmatrix} E[W_{t_1}W_{t_1}] & E[W_{t_1}W_{t_2}] & \dots & E[W_{t_1}W_{t_m}] \\ E[W_{t_2}W_{t_1}] & E[W_{t_2}W_{t_2}] & \dots & E[W_{t_2}W_{t_m}] \\ \vdots & \vdots & \ddots & \vdots \\ E[W_{t_m}W_{t_1}] & E[W_{t_m}W_{t_2}]\,; & \dots & E[W_{t_m}W_{t_m}] \end{bmatrix} = \begin{bmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \dots & t_m \end{bmatrix}$$

Moreover the moment–generating function for $(W_{t_1}, W_{t_2} \dots, W_{t_m})$ is

$$\begin{aligned} \varphi(u_1, \dots, u_m) &= E\left[\exp\left\{u_m W_{t_m} + u_{m-1} W_{t_{m-1}} + \dots + u_1 W_{t_1}\right\}\right] \\ &= \exp\left\{\frac{1}{2}(u_1 + u_2 + u_m)^2 t_1 + \frac{1}{2}(u_2 + u_3 + u_m)^2 (t_2 - t_1) + \cdots \right. \\ &\qquad \left. \cdots + \frac{1}{2}(u_{m-1} + u_m)^2 (t_{m-1} - t_{m-2}) + \frac{1}{2} u_m^2 (t_m - t_{m-1})\right\}. \end{aligned} \tag{6.14}$$

PROOF: See SCF2, ch.3.3.2 ■

It is well known that moment–generating functions uniquely determine the distribution of random variables and random vectors. Thus we have the following.

Theorem 6.2 (SCF2 Theorem 3.3.2 – Characterizations of Brownian motion). ⋆

Let $(\Omega, \mathfrak{F}, P)$ be a probability space with a process $W_t, 0 \le t < \infty$ satisfying
- $W_0(\omega) = 0$,
- *the trajectories $t \mapsto W_t(\omega)$ are continuous functions of t P–a.s.*

Then we have equivalence ***(1)*** $\Leftrightarrow$ ***(2)*** $\Leftrightarrow$ ***(3)*** *of the following three statements:*

(1) *For all $0 = t_0 < t_1 < \dots < t_m$, the increments*

$$W_{t_1} - W_{t_0},\ W_{t_2} - W_{t_1},\ \dots,\ W_{t_m} - W_{t_{m-1}},$$

are independent, and each of these increments is normally distributed with mean zero and variance $Var[W_{t_m} - W_{t_{m-1}}] = t_m - t_{m-1}$.

(2) *For all $0 = t_0 < t_1 < \dots < t_m$, the random variables $W_{t_1}, W_{t_2} \dots, W_{t_m}$ are jointly normal with means $E[W_{t_j}] = 0$ and covariance matrix (6.13).*

(3) *For all $0 = t_0 < t_1 < \dots < t_m$, the random variables $W_{t_1}, W_{t_2} \dots, W_{t_m}$ have the joint moment–generating function (6.14).*

Further, if one of ***(1)***, ***(2)***, ***(3)*** *is satisfied, then $(W_t)_{t \ge 0}$ is a Brownian motion with respect to $\mathfrak{F}_t^W$.*

PROOF: ■

The following is SCF2 Theorem 3.3.4.

Theorem 6.3 (Brownian motion is a martingale).

Let $W = (W_t)_{t \ge 0}$, be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Then W is an $\mathfrak{F}_t$–martingale.

PROOF: For $0 \le s \le t$, we have

$$\begin{aligned} E[W_t \,|\, \mathfrak{F}_s] &= E[(W_t - W_s) + W_s \,|\, \mathfrak{F}_s] = E[(W_t - W_s) \,|\, \mathfrak{F}_s] + E[W_s \,|\, \mathfrak{F}_s] \\ &= E[W_t - W_s] + W_s = W_s. \end{aligned}$$

The third equation results **a)** from the independence of $W_t - W_s$ and $\mathfrak{F}_s$, and **b)** from the $\mathfrak{F}_s$–measurability of W_s. ■

6.3 Digression: L^1 and L^2 Convergence ⋆

In this section we use the same symbol $\|\cdot\|$ for very different ways to define the size of an item, and the same symbol $d(\cdot,\cdot)$ for very different ways to define the distance of two items.

Example 6.2. Here we give six examples of measuring sizes and distances. The first is well known from linear algebra.

(a) For vectors $\vec{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\vec{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$, we easily accept that

$$\|\vec{x}\|_2 := \sqrt{\sum_{j=1}^{n} {x_j}^2} \qquad \text{and} \qquad d_2(\vec{x}, \vec{y}) := \|\vec{x} - \vec{y}\|_2 \tag{6.15}$$

are a good way to measure the size of $\vec{x}$ and the distance between $\vec{x}$ and $\vec{y}$. If $n = 2$ then $\vec{x}$ and $\vec{y}$ are ε–close, i.e., have distance less than ε, $\Leftrightarrow$ $\vec{y}$ lies within a circle of radius ε around $\vec{x}$.

(b) The following is not quite as plausible, but we might also be willing to accept

$$\|\vec{x}\|_1 := \sum_{j=1}^{n} |x_j| \qquad \text{and} \qquad d_1(\vec{x}, \vec{y}) := \|\vec{x} - \vec{y}\|_1 \tag{6.16}$$

as a way to measure the size of $\vec{x}$ and the distance between $\vec{x}$ and $\vec{y}$. Now, if $n = 2$, the vectors $\vec{x}$ and $\vec{y}$ are ε–close $\Leftrightarrow$ $\vec{y}$ lies within the tilted rectangle with edges $(x_1 \pm \varepsilon, y_2)$ and $(x_1, y_2 \pm \varepsilon)$.

(c) For real–valued functions $f, g : [a,b] \to \mathbb{R}$, defined on an interval $[a,b] \subseteq \mathbb{R}$, we could measure the size $\|f\|_{L^1}$ of f by the area enclosed by the graph of f, the x–axis, and the vertical lines, $y = a$ and $y = b$, and we could measure the distance $d(f,g)$ between f and g by the area which is enclosed by the graphs of f and g, and the vertical lines, $y = a$ and $y = b$. In other words,

$$\|f\|_{L^1} := \int_a^b |f(t)|\,dt \qquad \text{and} \qquad d_{L^1}(f,g) := \|f - g\|_{L^1}. \tag{6.17}$$

(d) This time working with squares is not quite as plausible as what we did in **(c)**, but we might also be willing to accept for $f, g : [a,b] \to \mathbb{R}$ to measure the size $\|f\|$ of f and the distance $d(f,g)$ between f and g as follows.

$$\|f\|_{L^2} := \sqrt{\int_a^b f(t)^2\,dt} \qquad \text{and} \qquad d_{L^2}(f,g) := \|f - g\|_{L^2}. \tag{6.18}$$

In the remaining examples we extend **(d)** to integrals of a more general type. The reader can easily do the corresponding generalizations of **(c)**.

(e) We replace $\int \dots dt$ with $\int \dots \varphi(t)dt$ for some fixed, measurable, nonnegative, $\varphi : \mathbb{R} \to \mathbb{R}$. This includes the case of an interval $-\infty < a < b < \infty$, since we can chose the "density" φ to be zero outside $[a, b]$. We now define for $f, g : \mathbb{R} \to \mathbb{R}$, size and difference as follows.

$$\|f\|_{L^2} := \sqrt{\int_{-\infty}^{\infty} f(t)^2 \varphi(t)\, dt} \qquad \text{and} \qquad d_{L^2}(f, g) := \|f - g\|_{L^2}. \tag{6.19}$$

This last example shows how to make the transition from functions defined for real arguments to functions defined on an abstract domain Ω by replacing $\int\limits_{-\infty}^{\infty} \dots \varphi(t)\, dt$ with the abstract integral $\int_\Omega \dots d\mu(\omega)$.

(f) Let $(\Omega, \mathfrak{F}, \mu)$ be a measurable space with a σ–finite measure μ, and assume that f and g are real–valued and Borel measurable functions on Ω. We define size and difference as follows.

$$\|f\|_{L^2} := \sqrt{\int_\Omega f(\omega)^2\, d\mu(\omega)} \qquad \text{and} \qquad d_{L^2}(f, g) := \|f - g\|_{L^2}. \ \square \tag{6.20}$$

It can be shown that the functions $\|\cdot\|$ which occur in all the examples above satisfy the properties of the following definition if we exclude elements x for which $\|x\| = \infty$.

Definition 6.5 (Seminorm)**.**

Let V be a vector space (in the abstract sense). A function

$$\|\cdot\| : V \longrightarrow \mathbb{R}, \qquad x \mapsto \|x\|$$

is called a **seminorm** on V if it satisfies the following.

(6.21a)	$\|x\| \geq 0$ for all $x \in V$ and $\|0\| = 0$	**positive semidefiniteness**
(6.21b)	$\|\alpha x\| = \|\alpha\| \cdot \|x\|$ for all $x \in V$, $\alpha \in \mathbb{R}$	**absolute homogeneity**
(6.21c)	$\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$	**triangle inequality** $\square$

It can also be shown that the functions $d(\cdot, \cdot)$ in all examples satisfy the properties of the following definition if we exclude elements x, y for which $d(x, y) = \infty$. Matter of fact, they are satisfied whenever we set

$$d(x, y) := \|y - x\|$$

for a seminorm $\|\cdot\|$ as defined above.

Definition 6.6 (Pseudometric spaces)**.** Let X be an arbitrary, nonempty set.

A **pseudometric** on X is a real–valued function of two arguments

$$d(\cdot,\cdot): X \times X \to \mathbb{R}, \qquad (x,y) \mapsto d(x,y)$$

satisfying the following three properties:

(6.22a)	$d(x,y) \geq 0$ and $d(x,x) = 0$ for all $x,y \in X$	**positive semidefiniteness**
(6.22b)	$d(x,y) = d(y,x)$ for all $x,y \in X$	**symmetry**
(6.22c)	$d(x,z) \leq d(x,y) + d(y,z)$ for all $x,y,z \in X$	**triangle inequality**

Let $x,y \in X$ and $\varepsilon > 0$. We say that x and y are ε**–close**, if $d(x,y) < \varepsilon$. ☐

There is a fundamental difference between the cases **(a)**, **(b)** and the cases **(c)–(f)**. In the first two cases it is easy to see that positive semidefiniteness can be strengthened to "positive definiteness"

(6.23) $$\|\vec{x}\| = 0 \;\Leftrightarrow\; \vec{x} = 0 \qquad \text{and} \qquad d(\vec{x},\vec{y}) = 0 \;\Leftrightarrow\; \vec{x} = \vec{y}.$$

On the other hand, regardless whether we interpret $\int \dots dt$ as Riemann integral or Lebesgue integral, if $f(t) = 1$ for $t = \frac{a+b}{2}$ and zero else, and if $g(t) = 0$ for all $t \in [a,b]$, then

$$\|f\| = 0 \qquad \text{and} \qquad d(f,g) = 0\,,$$

even though $f \neq 0$ and $f \neq g$.
One can actually show the following for σ–finite measures μ.

(6.24) $$\int |f|\,d\mu = 0 \;\Leftrightarrow\; \int f^2\,d\mu = 0 \;\Leftrightarrow\; f = 0 \;\; \mu\text{–a.e.},$$

and thus

(6.25) $$\int |f-g|\,d\mu = 0 \;\Leftrightarrow\; \int (f-g)^2\,d\mu = 0 \;\Leftrightarrow\; f = g \;\; \mu\text{–a.e.}$$

There is another difference but it is of more of a technical nature. It will never happen in exampless **(a)**, **(b)** that $\|\vec{x}\| = \infty$ or $d(\vec{x},\vec{y}) = \infty$. In contrast to this note that, for example, $\int_0^1 \ln(x)dx = \infty$ and $\int_0^1 \big(\ln(x)\big)^2 dx = \infty$.

Before we continue, note that there is no substantial difference between examples **c** and **d**. Moreover **d** and **e** are specific cases of example **f**. We thus focus our attention on **a**, **b**, **f**.

The "positive definiteness" property of formula 6.23 is so important that it leads to the following definitions which are a lot more important than those of seminorms and pseudometrics.

Definition 6.7 (Norm)**.**

Let V be a vector space (in the abstract sense). A function

$$\|\cdot\| : V \longrightarrow \mathbb{R}, \qquad x \mapsto \|x\|$$

is called a **norm** on V if it satisfies the following.

(6.26a) $\|x\| \geq 0$ for all $x \in V$ and $\|x\| = 0 \Leftrightarrow x = 0$ — **positive definiteness**

(6.26b) $\|\alpha x\| = |\alpha| \cdot \|x\|$ for all $x \in V$, $\alpha \in \mathbb{R}$ — **absolute homogeneity**

(6.26c) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$ — **triangle inequality**

The pair $(V, \|\cdot\|)$ is called a **normed vector space** □

Definition 6.8 (Metric spaces)**.**

Let X be an arbitrary, nonempty set. A **metric** on X is a real–valued function of two arguments

$$d(\cdot,\cdot) : X \times X \to \mathbb{R}, \qquad (x,y) \mapsto d(x,y)$$

with the following three properties:

(6.27a) $d(x,y) \geq 0$ for all $x, y \in X$ and $d(x,y) = 0 \Leftrightarrow x = y$ — **positive definite**

(6.27b) $d(x,y) = d(y,x)$ for all $x, y \in X$ — **symmetry**

(6.27c) $d(x,z) \leq d(x,y) + d(y,z)$ for all $x, y, z \in X$ — **triangle inequality**

The pair $(X, d(\cdot,\cdot))$, usually just written as (X, d), is called a **metric space**. We write X for short, if it is clear which metric is referred to. □

Remark 6.6. ⋆

From the perspective of advanced mathematics there are tremendous advantages to having norms and metrics rather than seminorms and semimetrics. The mechanism to enforce positive definiteness is to call two measurable functions f and g equivalent if $f = g\,\mu$–a.e. and work with those equivalence classes $[f]$ rather than with the original functions f. We do not worry about such sophistication. We usually write f for those equivalence classes $[f]$. □

6.4 Quadratic Variation of Brownian Motion

Notation 6.1. In the following the letter Π will not denote the pricing function of a contingent claim as will be the case when we discuss financial markets, e.g., in Chapter 8 (The Binomial Asset Model). Rather, it will denote a **partition**

$$\Pi := \Pi_t := \{t_0, t_1, \dots, t_n\}, \quad \text{where } 0 = t_0, < t_1, < \cdots < t_n = t; \ (0 \leq t \leq T).$$

Such a partition is interpreted as a set of times for a stochastic process with index set $I = [0, T]$ for some fixed $T > 0$ and $0 \le t \le T$. We will often write Π for Π_t if this does not lead to confusion.

The step sizes $t_j - t_{j-1}$ are not assumed to be of equal size. We denote by

$$\|\Pi_t\| \;:=\; \max\{t_{j+1} - t_j \;:\; j = 0, \dots, n-1\}\,.$$

the maximum step size (difference of neighboring times) of the partition. We will refer to $\|\Pi_t\|$ as the **mesh** of Π_t. □

SCF2 defines the first–order variation of a function $[0, T] \to \mathbb{R}$, but we have no use for it Instead we directly introduce the quadratic variation of such functions. The following is SCF2 Definition 3.4.1

Definition 6.9 (Quadratic Variation)**.**

Let $f : [0, T] \to \mathbb{R}$ be a (Borel measurable) function of time t, and let $0 \le t \le T$. We call

(6.28) $$[f, f](t) \;:=\; \lim_{\|\Pi_t\| \to 0} \sum_{j=0}^{n-1} [\,f(t_{j+1}) \;-\; f(t_j)\,]^2,$$

the **quadratic variation of f up to time** t.

Here the limit $\lim\limits_{\|\Pi\| \to 0}$ is to be understood in the same way as

$$\int_a^b f(u)\,du \;=\; \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} f(t_j^*)(t_j \;-\; t_{j-1})\,, \quad t_{j-1} \le t_j^* \le t_j,$$

in the definiton of the Riemann integral. In other words, the limit is taken along partitions $\Pi_t = \{0 = t_0 < t_1 < \dots < t_n = t\}$ in such a way that the mesh becomes smaller and smaller. □

Remark 6.7 (Notation for quadratic variation of stochastic processes)**.** Quadratic variation makes sense for any function that depends on "time" t, including the paths $t \mapsto X_t(\omega)$ of a stochastic process $X_t, 0 \le t \le T$.

We will often write $[X, X]_t$ and $[X, X]_t(\omega)$ rather than $[X, X](t)$ and $[X, X](t, \omega)$. □

Remark 6.8. Let $f : [0, T] \to \mathbb{R}$ be a (Borel measurable) function with a continuous derivative. Let $0 \le t \le T$. Then $[f, f](t) = 0$.

You will find a proof of this in SCF2 Remark 3.4.2. □

SCF2 Theorem 3.4.3 states the following. Let W be a Brownian motion. Then, for almost surely all $\omega \in \Omega$,

$$[W, W]_t(\omega) \;=\; t \quad \text{for all } 0 \le t \le T\,.$$

He actually proves a lot less:

Theorem 6.4.

Let W be a Brownian motion. For $0 \leq t \leq T$ and a partition $\Pi = \{t_0, t_1, \dots, t_n\}$ of $[0, t]$, let

$$Q_\Pi(t) := \sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j})^2.$$

Then,

$$\lim_{\|\Pi\| \to 0} E\left[(Q_\Pi(t) - t)^2\right] = 0.$$

PROOF: See the proof of SCF2 Theorem 3.4.3. ■

Remark 6.9. SCF2 Remark 3.4.4 and 3.4.5 are to a large degree about making plausible the extremely important relations

- $dt\,dt = 0$,
- $dt\,dW_t = dW_t\,dt = 0$,
- $dW_t\,dW_t = dt$.

Even though I can follow those remarks line by line I fail to see understand how they make it easier to understand this so called **multiplication table for Brownian motion differentials**. I will explain them differently later in the course.

Here is one thing he says that should be clear to all.

Brownian motion accumulates quadratic variation at rate one per unit time. □

6.5 Brownian Motion as a Markov Process

Theorem 6.5 (SCF2 Thm.3.5.1)**.**

Let W be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Then W is a Markov process.

PROOF (outline): Let $0 \leq s \leq t \leq T$ and $f_t : \mathbb{R} \to [0, \infty,\ x \mapsto f_t(x)$ Borel–measurable. According to Definition 6.2 which corresponds to SCF2 Definition 2.3.6 of a Markov process one must find another Borel–measurable function $f_s : x \mapsto f_s(x)$ such that

(6.29) $$E[f_t(W_t) \mid \mathfrak{F}_s] = f_s(W_s).$$

It can be shown that

(6.30) $$f_s : \mathbb{R} \longrightarrow \mathbb{R}, \qquad x \mapsto E\left[f_t(x + W_t - W_s)\right]$$

is the sought after function. For the proof see SCF2 ch.3.5. Note that that proof does not require the normality of W_t. It entirely relies on the fact that the increments $W_{t+h} - W_t$ are independent of $\mathfrak{F}_t$. ■

We will show that Brownian motion has a transition density according to the next definition.

Definition 6.10. ⋆

Let $X = X_t$ be a real–valued and adapted Markov process on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Assume there exists a Borel measurable function

$$p :]0, \infty[\times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}\,; \quad (\tau, x, y) \mapsto p(\tau, x, y) \tag{6.31}$$

such that $x \mapsto p(\tau, x, y)$ is Borel measurable for each fixed τ and y, and which satisfies, for every nonnegative Borel measurable function $f : \mathbb{R} \to \mathbb{R}$ and $s \geq 0$ and $\tau > 0$ the relation,

$$E[f(X_{s+\tau}) \mid \mathfrak{F}_s] \;=\; \int_{-\infty}^{\infty} f(y)\, p(\tau, X_s, y)\, dy\,. \tag{6.32}$$

We call $p(\tau, x, y)$ the **transition density** for X. □

Remark 6.10. ⋆ Formula (6.32) is an equation of two random variables which holds true almost surely. We supply the argument ω to emphasize this aspect and obtain for $s \geq 0$ and $\tau > 0$.

$$E[f(X_{s+\tau}) \mid \mathfrak{F}_s](\omega) \;=\; \int_{-\infty}^{\infty} f(y)\, p(\tau, X_s(\omega), y)\, dy\,, \quad \text{a.s.} \tag{6.33}$$

In particular, let $B \subseteq \mathbb{R}$ be a Borel subset and $f(x) := \mathbf{1}_B(x)$. Then (6.33) becomes

$$P\{X_{s+\tau} \in B \mid \mathfrak{F}_s\}(\omega) \;= E[\mathbf{1}_B(X_{s+\tau}) \mid \mathfrak{F}_s](\omega) \;=\; \int_B p(\tau, X_s(\omega), y)\, dy\,, \quad \text{a.s.} \tag{6.34}$$

We recall from Proposition 6.2 on p.124 that the expressions above are $\sigma(X_s)$–measurable. This can also be seen directly since the random variable

$$\omega \;\mapsto\; \int_{\mathbb{R}} f(y)\, p(\tau, X_s(\omega), y)\, dy$$

is, for frozen τ, a function of $X_s(\omega)$ only and hence $\sigma(X_s)$ measurable. Thus conditioning with respect to $\mathfrak{F}_s$ is the same as conditioning with respect to X_s. Thus, from (6.34),

$$P\{X_{s+\tau} \in B \mid X_s\}(\omega) \;=\; \int_B p(\tau, X_s(\omega), y)\, dy\,, \quad \text{a.s.}\,. \tag{6.35}$$

As in Remark 5.5 on p.116, Doob composition applied to $P\{\cdots \mid X_s\}$ yields a Borel measurable function $x \mapsto g(x)$ such that $P\{X_{s+\tau} \in B \mid X_s\} = g \circ X_s$. Again, it is customary to write

$$P\{X_{s+\tau} \in B \mid X_s = x\}$$

instead of $g(x)$ for this function, and this turns out to be the ordinary conditional probability when discrete random variables or random variables with joint density functions are involved. Under this convention we obtain the following for fixed x: If $X_s(\omega) = x$, then (6.34) and (6.35) yield

$$P\{X_{s+\tau} \in B \mid X_s = x\} \;=\; \int_B p(\tau, x, y)\, dy. \tag{6.36}$$

Thus $y \mapsto p(\tau, x, y)$ is exactly that "ordinary" conditional density for the probability of X ending up at time $s+\tau$ in a set B, under the condition that its trajectory was at time s in x.

The time s of conditioning does not appear in the expression on the right hand. Thus this conditional probability is equal to that of starting at time zero in x and ending up at time τ in B. This is informally stated as follows. If I know the postion of X at time s then I can consider s as my new start time. The trajectories $\tau \mapsto X_{s+\tau}$ will behave in terms of all probabilistic aspects just the same as the trajectories X_τ that had originally started at time zero in x. □

Proposition 6.7.

The transition density for a Brownian motion is

$$p(\tau, x, y) \;=\; \frac{1}{\sqrt{2\pi\tau}}\, e^{-\frac{(y-x)^2}{2\tau}}\,.$$

PROOF: The proof is given as part of SCF2 Theorem 3.5.1. ■

6.6 Additional Properties of Brownian Motion

We are skipping all of SCF2 Chapter 3.4.3 (Volatility of Geometric Brownian Motion) except for the following definition.

Definition 6.11 (Geometric Brownian Motion)**.**

Let W be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Let S_0, α, σ be real numbers such that $S_0, \sigma > 0$ We call the process

$$S_t \;:= S_0\, \exp\left[\sigma W_t \,+\, \left(\alpha - \frac{1}{2}\sigma^2\right) t\right]. \tag{6.37}$$

geometric Brownian motion or also **GBM**. We will see in Example 9.1 on p.194 how GBM is obtained as the solution of a SDE (stochastic differential equation) which models the price of the risky asset (stock) in the Black–Scholes option pricing framework. □

Definition 6.12 (Exponential martingale)**.**

Let $W = W_t, t \geq 0$, be a Brownian motion on a filtered probability space $\Omega, \mathfrak{F}, \mathfrak{F}_t, P$, and $\sigma \in \mathbb{R}$. We call the process $Z = Z_t, t \geq 0$, defined as

$$Z_t \;:= \exp\left[\sigma W_t \,-\, \frac{1}{2}\sigma^2 t\right], \tag{6.38}$$

the level σ **exponential martingale** of W. □

Z_t derives its name from the following theorem (SCF2 Theorem 3.6.1).

Theorem 6.6.

Let $W = W_t, t \geq 0$, be a Brownian motion on a filtered probability space $\Omega, \mathfrak{F}, \mathfrak{F}_t, P$ and $\sigma \in \mathbb{R}$. Then the level σ exponential martingale of W is an $\mathfrak{F}_t$–martingale.

PROOF: See SCF2 Theorem 3.6.1 for the proof. ■

The SCF2 text contains an entire chapter 3.2 on discrete time versions $X_t^{(n)}$, defined only for times $t_j = 2^{-n}j$ and called **symmetric random walks**. In a sense, one can represent (continuous time) Brownian motion as a limit of properly scaled and linearly interpolated symmetric random walks. We now briefly discuss a small part of this material.

Definition 6.13 (Scaled symmetric random walk)**.**

Let B_j be an iid sequence of random variables with two possible outcomes, 1 and -1. Assume that

$$p := P\{B_j = 1\} = \frac{1}{2}; \qquad q := 1 - p = \frac{1}{2} = P\{B_j = -1\}.$$

Let

$$X_0 := 0, \qquad X_k := \sum_{j=1}^{k} X_j,\ k = 1, 2, \dots \tag{6.39}$$

Then the process M_k lives on the grid of the integers, and, at each time k, it is equally likely that the process moves one unit to the left or to the right. For this reason we call this process a **symmetric random walk**. □

To approximate a Brownian motion, we speed up time and scale down the step size of a symmetric random walk. More precisely, we proceed as follows.

Definition 6.14 (Scaled symmetric random walk)**.**

Let $n \in \mathbb{N}$. For $t \geq 0$ let the integer k be determined by $k \leq nt \leq k+1$. Let

$$W_t^{(n)} := \begin{cases} \frac{1}{\sqrt{n}} X_{nt} \text{ if } nt \text{ is an integer,} \\ \text{the linear interpolation of } \frac{1}{\sqrt{n}} X_k \text{ and } \frac{1}{\sqrt{n}} X_{k+1} \text{ otherwise.} \end{cases} \tag{6.40}$$

We call the continuous time process $W_t^{(n)}$ the n**–th scaled symmetric random walk**. □

Theorem 6.7 (SCF2 Theorem 3.2.1 - Central Limit Theorem for scaled random walk)**.** ⋆

Let $t > 0$. As $n \to \infty$, the distribution of the scaled random walk $W^{(n)}(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t.

PROOF: See SCF2. ■

6.7 Exercises for Ch.6

Exercise 6.1. Prove the assertions of Remark 6.1 on p.123 of this document.

Hint: Use induction to prove the remark for a submartingale X_n. Apply this result to $-Y_n$ to obtain a proof for the case of a supermartingale Y_n. The result for a martingale is then immediate. □

Exercise 6.2. Prove prop.6.1 on p.123 of this document:

$$\text{A martingale } X \text{ satisfies} \quad E[X_s] \;=\; E[X_t] \quad \text{for any } s,t \in I\,. \;\square$$

7 Financial Models - The Basics

7.1 Interest Bearing Financial Assets

Before we discuss financial markets, we list some facts about interest payments. We assume that the student has some basic knowledge about interest paid on interest bearing financial instruments such as a bank account or a bond and how to discount such an instrument, if it pays a certain amount at a future date, to its present value.

This material that is taught, e.g., in Math 346 (Introduction to Financial Mathematics), a course which is officially listed as a prerequisite for Math 454. Students who did not attend Math 346 are expected to study this material on their own in a text like, e.g., [1] Anthony, Martin and Biggs, Norman: Mathematics for Economics and Finance - Methods and Modelling.

Unless something else is stated explicitly, we assume the following.

- The unit of time is one year, so $t = 3.5$ means 3 years and 6 months after $t = 0$, which denotes the initial point in time.
- All interest rates are annual interest rates, so $r = 0.048$ means an annual interest rate of 4.8%.
- The unit of currency is one dollar, as opposed to one dime or 15 yen or one renminbi or ...

7.1.1 Interest Compounded at Discrete Points in Time

Remark 7.1. We start with the following observation. Let $s < t$. Consider a bank account that contains x_s dollars at time s, an interest rate that is constant = r between s and t, and assume that interest is not compounded continously, but computed at t based on x_s, the account balance at s.

(a) What is x_t, the balance at time t, if no changes have been made to the original investment?

The answer to this question is as follows. Each dollar earns interest in height of $r(t-s)$ during $t-s$ years. Since the interest paid on x_s dollars is $x_s \cdot r(t-s)$. Accordingly, x_s has grown to

$$x_t \;=\; x_s + x_s \cdot r(t-s) \;=\; x_s\big(1 + r(t-s)\big)\,. \tag{7.1}$$

(b) In reverse, if the account balance at time t is x_t dollars, how much money had to be invested at time s? We obtain the answer from **(a)** by solving (7.1) for x_s:

$$x_s \;=\; \frac{x_t}{1 + r(t-s)}\,. \tag{7.2}$$

(b′) Note that (7.2) also answers the following question, since it is a reformulation of the one asked in **(b)**: If an investment in an interest bearing asset pays the amount x_t at the future time t, how much money is it worth today, at time $s < t$? The answer is given by (7.2),

$$x_s \;=\; \frac{x_t}{1 + r(t-s)}\,, \tag{7.3}$$

since only x_s dollars need to be invested today to achieve a payout of x_t dollars at time t.

The process of removing from x_t the interest to be earned during the interval $]s,t]$ is referred to as **discounting** x_t to its **present value**, x_s (at time s). □

Example 7.1. Assume that the annual interest rate is 4%, interest is compounded semiannually, and an amount of $x_{t_0} = 600$ dollars has been invested today, at time t_0. Interest will be paid as follows.

- The first interest payment happens after half a year, at $t_1 := t_0 + \frac{1}{2}$. The investment has grown to
$$x_{t_1} \;=\; 600 \,+\, 600 \cdot \frac{1}{2} \cdot 0.04 \;=\; 600\big(1 + 0.02\big) \;=\; 612 \text{ dollars}\,.$$
- Interest is paid again at $t_2 := t_1 + \frac{1}{2}$. Since it is based on $x_{t_1} = \$612$, the original investment has grown to
$$x_{t_2} \;=\; 612 \,+\, 612 \cdot \frac{1}{2} \cdot 0.04 \;=\; 600\big(1 + 0.02\big)\big(1 + 0.02\big) \;=\; 624.24 \text{ dollars}\,.$$
- At $t_3 := t_2 + \frac{1}{2}$, the time of the third interest payment, the account balance is
$$x_{t_3} \;=\; x_{t_2} \,+\, x_{t_2} \cdot 0.02 \;=\; 600\big(1 + 0.02\big)^3 \;=\; 636.7248 \text{ dollars}\,,$$

and so on. Since interest is only added at $t_1, t_2, \dots$, we see that
$$x_t \;=\; x_{t_0} \text{ for } t_0 \le t < t_1, \qquad x_t \;=\; x_{t_1} \text{ for } t_1 \le t < t_2, \qquad x_t \;=\; x_{t_2} \text{ for } t_2 \le t < t_3, \qquad \dots \quad \square$$

Example 7.1 easily generalizes to the case where interest can vary but remains constant between consecutive interest payments. We do this next.

Example 7.2. Assume that an interest bearing account is opened at time t_0 and that interest is compounded only at the specific times
$$t_1 \;<\; t_2 \;<\; t_3 \;<\; \cdots$$
Further, assume that the interest rate for the period $]t_{j-1}, t_j]$ is constant and equals r_j. The calculations that were done in Example 7.1 show the following: Due to interest added, each dollar that was invested at time t_0 will have grown at time t to the following amount b_t:
$$\begin{aligned}
t_0 \le t < t_1: &\quad b_t \;=\; 1 \quad \text{(no interest has been added yet)},\\
t_1 \le t < t_2: &\quad b_t \;=\; b_0\big(1 + r_1(t_1 - t_0)\big) \;=\; 1 + r_1(t_1 - t_0),\\
t_2 \le t < t_3: &\quad b_t \;=\; b_1\big(1 + r_2(t_2 - t_1)\big) \;=\; \big(1 + r_1(t_1 - t_0)\big)\big(1 + r_2(t_2 - t_1)\big),\\
&\quad \cdots\cdots\cdots\cdots\cdots\\
t_j \le t < t_{j+1}: &\quad b_t \;=\; b_{t_{j-1}}\big(1 + r_{t_j}(t_{t_j} - t_{t_{j-1}})\big).\\
&\quad\;\;\;\; =\; \big(1 + r_1(t_1 - t_0)\big)\big(1 + r_2(t_2 - t_1)\big)\cdots\big(1 + r_{j-1}(t_{j-1} - t_{j-2})\big)\big(1 + r_j(t_j - t_{j-1})\big)
\end{aligned}$$
As in Example 7.1, x_t remains constant during the intervals $]t_{j-1}, t_j[$, since interest is not added other than at $t_1, t_2, \dots$. $\square$

This course focuses on financial models in which interest is compounded continuously. We discuss discrete time interest payments only in connection with the binomial asset model. [30] There we assume that the initial investment happens at $t_0 = 0$ and that interest is compounded at times $1, 2, 3, \dots$ at one and the same rate r. The formulas of Example 7.2 become much simpler under those conditions:

Example 7.3. Assume that an interest bearing account is opened at time $t_0 = 0$ and that interest is compounded only at the specific times
$$t_1 = 1,\; t_2 = 2,\; t_3 = 3,\; \dots$$

[30] See Chapter 8 (The Binomial Asset Model).

Further, assume that the interest rate is constant and equals r at all times t. Then the results of Example 7.2 read as follows:

(a) x_0 dollars invested at $t = 0$ grow to

$$\begin{array}{rlll} 0 \le t < 1: & x_t &=& x_0 \quad \text{(no interest has been added yet)}, \\ 1 \le t < 2: & x_t &=& 1 + r, \\ 2 \le t < 3: & x_t &=& x_1(1+r) \;=\; x_0(1+r)^2, \\ & & \ldots & \\ j \le t < j+1: & x_t &=& x_{j-1}(1+r) \;=\; x_0(1+r)^j. \end{array}$$

In particular, for a time which is an integer n,

$$x_n \;=\; x_0(1+r)^n\,. \tag{7.4}$$

(b) Let $m, n \in [0, \infty[_{\mathbb{Z}}$ and $m < n$. An investment of x_n at time n is discounted to its present value at m as follows:

$$x_m \;=\; \frac{x_n}{(1+r)^{n-m}}\,.$$

In particular, if $m = 0$,

$$x_0 \;=\; \frac{x_n}{(1+r)^n}\,. \quad \square \tag{7.5}$$

- There is a special name and symbol for the interest accrual of (7.4), if the initial investment is one monetary unit, i.e., $x_0 = 1$ dollar. [31]
- There also is a special name and symbol for the discounting of (7.5), if the account balance at the future time n is one monetary unit, i.e., $x_n = 1$ dollar.

Definition 7.1. We assume the settings of Example 7.3 above.

- We write B_t for the amount of money to which an investment of one dollar at time $t_0 = 0$ in an interest bearing account has increased at time t. We call $t \mapsto B_t$ its **money market account price** aka **money market account price process**.
- We write D_t for the amount of money that needs to be invested at time $t_0 = 0$ in an interest bearing account, so that it will grow by means of interest accrual to one dollar at time t. We call $t \mapsto D_t$ the **discount process** of the account.

Clearly, if an investment of $x_0 = 1$ at $t_0 = 0$ grows to B_t at time t and D_t invested at $t_0 = 0$ grows to 1 at time t, then then an investment of x_0 at $t_0 = 0$ grows to $x_0 B_t$ at time t and $x_t D_t$ invested at $t_0 = 0$ grows to x_t at time t.

- We call the dollar amount $x_t D_t$ the **present value** of x_t (with respect to the time t), and we say that x_t has been **discounted** to x_s, the **present value** of the investment (at time s). $\square$

[31] We express monetary amounts in dollars, but we could also have chosen another currency instead, e.g., euros, renminbi, yen, rubels, ...

Proposition 7.1.

We assume that an interest bearing account compounds interest at the times $1, 2, \dots$ (years) at a constant, annual interest rate, given by $r \geq 0$. Then, money market account price B_n and discounted value D_n are

$$B_n = (1+r)^n \,, \tag{7.6}$$

$$D_n = \frac{1}{B_n} = (1+r)^{-n} \,. \tag{7.7}$$

PROOF: This is a trivial consequence of (7.4) and (7.5). □

7.1.2 Continuously Compounded Interest

Remark 7.2. As in Remark 7.1 on p.140, we consider a bank account that contains x_s dollars at time s, and an interest rate that is constant = r between s and t ($s < t$).

(A) We assume that the interval from s to t is partitioned into k equally sized subintervals of size $(t-s)/k$,

$$x_s, \;\; x_s + \frac{t-s}{k}, \;\; x_s + \frac{2(t-s)}{k}, \;\dots,\; x_t = x_s + \frac{k(t-s)}{k}, \tag{7.8}$$

and that interest is compounded at each time $x_s + \frac{j(t-s)}{k}$, where $j = 1, \dots, k$.
The calculations in Example 7.3 on p.141 show the following, if one considers that r, the interest rate, is based on an interval of length 1 (year), and one must adjust it to $r(t-s)/k$, for each partition interval of length $(t-s)/k$.

- **(1)** Total interest accrued between s and t is $x_s \left(1 + \frac{r(t-s)}{k}\right)^k$.
- **(2)** Thus, continuously compounded interest (the limit as $k \to \infty$) is $x_s \, e^{r(t-s)}$.

(B) Next, we assume that interest is compounded continuously and the interest rate varies, but it is constant and equals r_j, on each subinterval

$$\left[x_s + \frac{(j-1)(t-s)}{k} \,,\, x_s + \frac{j(t-s)}{k} \right[\,.$$

We apply **(2)** to each one of those and see that the total interest accrued between s and t is

$$x_s \cdot e^{r_1(t-s)} \cdot e^{r_2(t-s)} \cdots e^{r_k(t-s)} = x_s \cdot \exp\left\{ \sum_{j=1}^{k} r_k(t-s) \right\}. \tag{7.9}$$

(C) Finally, we assume that interest is compounded continuously and the interest rate is not piecewise constant on $[s,t]$, but modeled by a Riemann integrable function $u \mapsto r(u)$. For $k \in \mathbb{N}$, we subdivide $[s,t]$ as we did in **(B)** and we set

$$r_j := r(u_j) \,, \quad \text{for some } x_s + \frac{(j-1)(t-s)}{k} \leq u_j \leq x_s + \frac{j(t-s)}{k} \,.$$

Now, the right–hand side exponent of (7.9) denotes a Riemann sum. We obtain that the interest that accrues between s and t is

(7.10) $$x_s \cdot e^{\int_s^t r(u)du}\,. \quad \square$$

We recall from Definition 7.1 on p.142 the following.

- B_t, the money market account price of an interest bearing account, is is the worth of an investment of one dollar at time 0
- D_t, the value of the discount process at time t, is the amount of money one must deposit at time 0 so that interest earned increases it to one dollar at time t.

Proposition 7.2.

We assume that an interest bearing account compounds interest continuously, at a varying interest rate which is given by $t \mapsto r(t)$. Then, money market account price B_t and discounted value D_t are

(7.11) $$B_t = e^{\int_0^t r(u)du}\,,$$

(7.12) $$D_t = \frac{1}{B_t} = e^{-\int_0^t r(u)du}\,.$$

PROOF: This is a trivial consequence of (7.10) and the fact that

$$x_t = x_s\, e^{\int_0^t r(u)du} \quad \Leftrightarrow \quad x_s = \frac{x_t}{e^{\int_0^t r(u)du}}\,. \quad \blacksquare$$

7.2 Assets and Contingent Claims, Trades, Portfolios and Arbitrage

The remainderof this entire chapter 7 (Financial Models - The Basics) closely follows the book [7] Björk, Thomas: Arbitrage Theory in Continuous Time. We use to a large degree the notation found there.

Everything happens in the context of a once and for all given probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. We interpret the filtration $(\mathfrak{F}_t)_t$ as the information available up to time t for a given financial market. We call this filtration the **information filtration** or also simply the **filtration** of the financial market.

Before you continue with this chapter, we suggest that you review chapters 4.4 (Stochastic Processes and Filtrations) and 6.1 (Martingales and Markov Processes) about the following:

- For the exact definition of a stochastic Process see Definition 4.16 on p.73.
- For the exact definition of a filtration see Definition 4.20 on p.77.
- For the exact definition of an adapted Process see Definition 4.21 on p.77.
- The definition of a Markov process is precise. See Proposition 6.2 on p.124. $\square$

Introduction 7.1. The finance part of this course is about pricing **financial derivatives** which are financial instruments defined in terms of (derived from) one or more underlying assets like stocks and bonds. Such financial derivatives are also called **contingent claims**. A prime example is a **European call** option for which the underlying asset is a stock. This option is a contract written at some time t_0. It specifies that, at the time of expiration $T > t_0$, the holder of this option has the right, but not the obligation, to buy a share of this stock for the price of K (dollars), the so called strike price, regardless of the market price S_T of that stock at time T.

We see several features in this example.

- The stock price S is a stochastic process $S_t(\omega)$ since it depends on time t and is non–deterministic, i.e., it depends on randomness ω.
- The value of this contract at time of expiration is a function of the stock price $S_T(\omega)$ at that time: The contract allows us to make a profit $X_T(\omega) - K$ if the price of the stock at time T exceeds the strike price, and it is worthless (but does not lead to a loss) otherwise.
- We call this contract value at time T the contract function $\mathcal{X}(\omega)$ of this option. For a European call, it is

$$\mathcal{X}(\omega) \,=\, \Phi\big(S_T(\omega)\big), \quad \text{where } \Phi(x) \,=\, (x-K)^+ \,=\, \max(x-K,0).$$

  We write $\Pi_t(\mathcal{X})$ for the price process of a contingent claim $\mathcal{X}$. In other words, $\Pi_t(\mathcal{X})(\omega)$ is the price of the financial derivative at time t. It is obvious that

$$\Pi_T(\mathcal{X}) \,=\, \mathcal{X}\,,$$

  since paying more for the claim at expiration time would be an unwise decision by the buyer, whereas offering the option for less would lead to a loss by the seller.
- Not so obvious: What is the appropriate price $\Pi_t(\mathcal{X})$ at a time t prior to T? In particular, what should be the price of this contract at the time t_0, when it is written? □

Definition 7.2 (Financial Market)**.** A **financial market model** aka **financial market** consists of the following.

(1) A collection of financial assets $\vec{\mathscr{A}} \,=\, \big(\mathscr{A}^{(0)}, \mathscr{A}^{(1)}, \dots, \mathscr{A}^{(n)}\big)$, e.g., stocks, bonds, options written on stocks, ... We distinguish between **riskless assets** such as bank accounts or zero coupon bonds where the money will grow according to an underlying interest rate and **risky assets** such as stocks which will fluctuate in value for a variety of reasons. Of course the real world is more complex and this distinction has been made for conceptual simplicity.

(2) Unit prices $\vec{S}_t(\omega) \,=\, \big(S_t^{(0)}(\omega), S_t^{(1)}(\omega), \dots, S_t^{(n)}(\omega)\big)$ of the assets $\vec{\mathscr{A}}$.

(3) Trading times $t \geq 0$ at which the assets $\mathscr{A}^{(j)}$ may be bought or sold. We speak of a **continuous time financial market**, if those trading times form an interval $[t_0, T]$ or $[t_0, \infty[$. We speak of a **discrete time financial market**, if those trading times form a finite or infinite sequence $t_0 < t_1 < t_2 < \dots$ In either case, usually $t_0 = 0$.

- We consider the trading times t_j of a discrete time market as special times, i.e., as real numbers. We follow this convention even if the trading times happen to be integers $n_0, n_0+1, n_0+2, \dots$.
- Thus, $[t_j, t_n[= \{t \in \mathbb{R} : t_j \le t < t_n\}$, **NOT** $[t_j, t_n[= \{t_j, t_{j+1}, \dots, t_{n-1}\}$.
- In particular, $[t_{j-1}, t_j[$ denotes the times from the time of trade t_{j-1} until "just before" the time t_j of the next trade. This is not the empty set!

(4) **Interest** is earned by holdings in a riskless asset and increases their value as time progresses. Recall that accrual of interest defines two processes B_t and D_t as follows.
• x dollars invested at time zero will have increased at time t to $B_t \cdot x$ dollars. In other words, B_t is the money market account price process of the riskless asset.
• To have y dollars in the account at time t, only $D_t \cdot y = (1/B_t) \cdot y$ dollars need to be invested at time zero. So, D_t is the discount process of the riskless asset.

- In a discrete time financial market, we assume that interest is added to the money in a bank account only at the trading times t_j, and that the value of the holdings in that account remains constant during the interval $[\,t_{j-1}, t_j[$.
- In a continuous time financial market, we assume that interest is compounded continuously. ☐

Remark 7.3. The focus of these lecture notes is on continuous time financial markets. Only Chapter 8 (The Binomial Asset Model) is about discrete time financial markets. There, we assume that the trading times are the non–negative integers $0, 1, 2, \dots$ ☐

Notation 7.1.

- We use the term "stock" as a synonym for "risky asset".
- We use the terms "bond", "bank account", "money market account" as synonyms for "riskless asset". We do this even though there are differences. For example, bonds have risks if one intends to sell them before maturity, since their price will fall if interest rates rise.
- There usually will be a single riskless asset. We reserve slot zero for that asset and often write B_t rather than $S_t^{(0)}$ for the price of this asset to improve readability. ☐

Remark 7.4. The reader may have noted that Definition 7.1 on .p142 employs the symbol B_t for the money market account price at time t. Does it coincide with $S_t^{(0)}$? The answer is affirmative. This topic will be discussed in Section 7.3 (The Holdings Process of a Riskless Asset). ☐

We list here a few more financial derivatives in addition to the European call.

Definition 7.3.

- A **European put** option is a contract written at some time t_0. It specifies that, at the time of expiration $T > t_0$, the holder of this option has the right, but not the obligation, to sell a share of an underlying security for the price of K (strike price). Note that the contract function, which specifies the value of this derivative at time T to the contract holder, is
$$\Phi\big(S_T(\omega)\big), \quad \text{where } \Phi(x) = (K - x)^+ = \max(K - x, 0).$$

- An **American call** option is a contract written at some time t_0. It specifies that, at any time up to the time of expiration $T > t_0$, the holder of this option has the right, but not the obligation, to buy a share of an underlying security stock for the price of K (strike price).

- An **American put** option is a contract written at some time t_0. It specifies that, at any time up to the time of expiration $T > t_0$, the holder of this option has the right, but not the obligation, to sell a share of an underlying security for the price of K (strike price).

- A **forward contract** is a contract between two parties **A** (the seller of the contract) and **B** (the buyer), written at some time t_0. It specifies that, at the time of expiration $T > t_0$, **A** has the obligation to sell a share of an underlying security for the price of K (strike price), and **B** has the obligation to buy at this price. Clearly the value of the option to the buyer at time T is
$$\Phi\big(S_T(\omega)\big), \quad \text{where } \Phi(x) = x - K. \ \square$$

Trade happens in this market, so people will have portfolios which list for each asset how many units are being held. We have access to the market information $\mathfrak{F}_t^{\vec{S}}$ up to the time t of the trade, i.e., we can base our trades on the development of the asset prices up to that time, but we cannot peek into the future.

Definition 7.4 (Portfolio strategy)**.**

A **portfolio** or **portfolio strategy** is a stochastic process

$$(7.13) \qquad \vec{H} = \vec{H}_t(\omega) = \big(H_t^{(0)}(\omega), H_t^{(1)}(\omega), \dots, H_t^{(n)}(\omega)\big)$$

which denotes the holdings (quantity) $H_t^{(j)}$ someone has in asset $\mathscr{A}^{(j)}$ at time t. Negative values indicate that this quantity is not owned but owed. We speak of a **Markovian portfolio** if $\vec{H}$ is a Markov process. In other words, a Markovian portfolio depends on current stock price $\vec{S}_t$ only and not on $\mathfrak{F}_t^{\vec{S}}$, the stock price of the past.

We say that $\vec{H}$ denotes a **long position** in the asset $\mathscr{A}^{(j)}$ at time t if $H_t^{(j)} > 0$. We say that $\vec{H}$ denotes a **short position** in this asset if $H_t^{(j)} < 0$.

We have to make some distinctions between continuous time and discrete time models:

Continuous case:
- We assume that $\vec{H}_t$ is $\mathfrak{F}_t^{\vec{S}}$–adapted.

Discrete case, with trading times $t_0 < t_1 < t_2 < \dots$:
- **(1)** we assume that $\vec{H}_t(\omega)$ is constant on each interval $[t_{k-1}, t_k[$,
- **(2)** we assume that $\vec{H}_{t_k}$ is $\mathfrak{F}_{t_{k-1}}^{\vec{S}}$–adapted ($k > 0$),
- **(3)** We define $\vec{H}_{t_0} := \vec{H}_{t_1}$. □

Notation 7.2 (Alternate notation for assets and portfolios)**.** We almost exclusively deal with only one riskless asset. Also, most of the financial models we study only have one risky asset. The major exception is Chapter 13 (Black–Scholes Model Part II: Risk–neutral Valuation). There, the case of multiple risky assets is examined. These considerations justify to introduce the following notation.

Given is a financial market with assets vector $\vec{\mathscr{A}} = \big(\mathscr{A}^{(0)}, \mathscr{A}^{(1)}, \dots, \mathscr{A}^{(n)}\big)$ and associated portfolio process $\vec{H}_t = \big(H_t^{(0)}, H_t^{(1)}, \dots, H_t^{(n)}\big)$.
- We assume that there is only one riskless asset and that $\mathscr{A}^{(0)}$ denotes it. We write

$$\mathscr{A}^B := \mathscr{A}^{(0)}, \qquad \text{and} \qquad H_t^B := H_t^{(0)}$$

  for this riskless asset and its holdings in the associated portfolio.

- If there is only one risky asset, we write

$$\mathscr{A}^S := \mathscr{A}^{(1)} \qquad \text{and} \qquad H_t^S := H_t^{(1)},$$

  for this risky asset and its holdings in the associated portfolio.
- Otherwise, there will be n risky assets $\mathscr{A}^{(1)}, \mathscr{A}^{(2)}, \dots \mathscr{A}^{(n)}$, with stock prices $\vec{S}_t = \big(S_t^{(1)}, \dots, S_t^{(n)}\big)$. We use vector notation for the risky assets and write

$$\mathscr{A}^{\vec{S}} := \big(\mathscr{A}^{(1)}, \dots \mathscr{A}^{(n)}\big) \qquad \text{and} \qquad H_t^{\vec{S}} := \big(H_t^{(1)}, \dots H_t^{(n)}\big)\,.$$

□

Remark 7.5. Everybody understands the meaning of shares and price in the context of a stocks. For example owning 800 shares of a stock that currently has a price of $\$25.00$ per share means that the value of the holdings in that asset is $800 \cdot 25.00 = 20,000.00$ dollars.

But what about the riskless asset? What could be the meaning of someone owning a bank account with 20,000 "bank shares" which are valued at $\$0.83$ each? An answer to that question will be given in Section 7.3 (The Holdings Process of a Riskless Asset). □

Definition 7.5 (Self–financing portfolio)**.**

A portfolio is a **self–financing portfolio strategy** (simply, **self–financing portfolio**), if money can be shifted around at times of trade by selling some assets and reinvesting the proceeds into other assets, subject to the following:

- It is not allowed to move any proceeds out of the portfolio to finance, e.g., the purchase of consumer goods or the next vacation.
- There is no infusion of external money to purchase additional shares.

In other words, the acquisition of additional shares in such portfolios must be financed through the sale of shares in some other asset or assets. □

Remark 7.6. The above definition of a self–financing portfolio is not very mathematical. We make it precise by formulating what is called a **Budget equation**. We will see later that discrete time trading models such as the multiperiod binomial asset model (Chapter 8.2) and continuous time trading models such as the Black–Scholes market (Chapter 10) have completely different budget equations. □

Definition 7.6 (Portfolio value)**.**

Given a portfolio $\vec{H} = \left(H_t^{(0)}, \dots, H_t^{(n)}\right)$, its **portfolio value** aka **portfolio value process** is

$$
\begin{aligned}
V^{\vec{H}} \; := \; V_t^{\vec{H}}(\omega) \; := \vec{H}_t(\omega) \bullet \vec{S}_t(\omega) \; &= \; \sum_{j=0}^{n} H_t^{(j)}(\omega) \cdot S_t^{(j)}(\omega) \\
&= H_t^{(0)} S_t^{(0)} \; + \; H_t^S S_t^{(1)} \; + \; \cdots \; + \; H_t^{(n)} S_t^{(n)}.
\end{aligned}
\tag{7.14}
$$

Since $H_t^{(j)} \cdot S_t^{(j)}$ equals

$$\text{number of units owned of asset } \mathscr{A}^{(j)} \quad \times \quad \text{unit price of asset } \mathscr{A}^{(j)},$$

$V_t^{\vec{H}}$ equals the total worth at time t of all holdings in that portfolio.

To simplify the notation, we often write V_t^H for $V_t^{\vec{H}}$. If it is clear from the context which portfolio is referenced, we omit the superscript and write V_t for $V_t^{\vec{H}}$. □

Remark 7.7. Note that determining the portfolio value in a tiscrete time market presents a conceptual difficulty. At each time of trade $t_k \neq t_0$, two portfolios exist, since the trade can be thought of as the sale of the entire old portfolio (H_t^B, H_t^S) which was purchased at time t_{k-1}, followed by the purchase of the new portfolio (H_{t+1}^B, H_{t+1}^S).

- If the investor decides to liquidate some of the investments to finance, e.g, a cruise aroound the world, the new portfolio will be worth less than the old one.
- Conversely if new money is used to purchaseadditional holdings, the new portfolio will be worth more than the old one.
- So, what should V_t, the portfolio value at such a time of trade be?

Fortunately, we will deal almost exclusively with self–financing portfolios. There is no ambiguity concerning the portfolio value of such portfolios, since both the sales value of the old portfolio and the purchase price of the new portfolio, must coincide for a self–financing portfolio. □

Definition 7.7 (Arbitrage Portfolio)**.**

A portfolio $\vec{H}_t$ is an **arbitrage portfolio**, if it allows with zero probability of risk to create money out of nothing with positive probability and does so without the infusion or withdrawal of money at any trading time $t > 0$.
In other words, $\vec{H}_t$ must be self–financing, and its value process $V_t^{\vec{H}}$ must satisfy

(7.15) $$V_0^{\vec{H}} = 0,$$

(7.16) $$P\{V_T^{\vec{H}} \geq 0\} = 1,$$

(7.17) $$P\{V_T^{\vec{H}} > 0\} > 0. \ \square$$

Note that the above is equivalent to replacing T with some $0 < t \leq T$, since we can invest the positive amount $V_t^{\vec{H}}$ entirely into the bond and have at least that much profit at time T.

Remember that we are designing a model and it is natural to make some simplifying assumptions even though they may be unrealistic in the real world.

Assumption 7.1. Unless stated differently, the market adheres to the following:

- Shares $H_t^{(j)}$ can equal any real number, and asset price per share $S_t^{(j)}$ can equal any strictly positive number. In particular we allow fractions of shares and asset prices.
- There is no **bid–ask spread**: The trading institution will not charge you more when it sells you an asset than the price at which it would buy it from you.
- There are no costs for executing a trade.
- The market is completely liquid: one can buy and/or sell unlimited quantities of any asset. In particular one can borrow unlimited amounts from the bank (by acquiring a short position in the bond).

The last condition is so central to the market model that we list it separately for emphasis.

- The market is efficient and thus **free of arbitrage**, i.e., it does not allow the existence of arbitrage portfolios. □

Definition 7.8 (Contingent Claim)**.**

A **contingent claim**, also called a **financial derivative**, is a $\mathfrak{F}_T^S$–measurable random variable $\mathcal{X}(\omega)$. We call $\mathcal{X}$ a **simple claim** if there is a function $s \mapsto \Phi(s)$ of asset price s or a function $\vec{s} \mapsto \Phi(\vec{s})$ of an asset price vector $\vec{s}$ such that

$$\mathcal{X} = \Phi \circ S_T.$$

We occasionally refer to Φ as the **contract function** of that claim. □

Definition 7.9 (Hedging/Replicating Portfolio)**.** Given are a contingent claim $\mathcal{X}$ and a portfolio $\vec{H}$.

(a) We say that $\vec{H}$ is a **hedging portfolio** or a **hedge** or a **replicating portfolio** for $\mathcal{X}$, and we say that $\mathcal{X}$ is **reachable** by $\vec{H}$, if $\vec{H}$ is self–financing and

$$V_T^H \;=\; \mathcal{X} \text{ almost surely.}$$

(b) If all claims can be replicated then we say that the market is **complete**. □

Remark 7.8. We stress that part of the definition of a replicating portfolio is the condition that it be self–financing. □

Part of Assumption 7.1 about a market is that there be no arbitrage. The next theorem states that in such a market all hedgeable contingent claims can be priced correctly (without admitting arbitrage) by means of their replicating portfolios. Björk refers to the next theorem as a **pricing principle**.

Theorem 7.1 (Pricing principle)**.**

Given is a contingent claim $\mathcal{X}$ with a replicating portfolio strategy $\vec{H}$.
For $\vec{H}$ to be free of arbitrage it it necessary that the option price process $\Pi(\mathcal{X})$ for that claim satisfies

$$\Pi(\mathcal{X}) \;=\; V^H, \quad \textit{i.e.,} \quad \Pi_t(\mathcal{X}) \;=\; V_t^H, \textit{ for all trading times } t.$$

PROOF:

The case $t = T$ is immediate: We mentioned already in the introduction 7.1 to Chapter 7.2 on Assets and Contingent Claims, Trades, Portfolios and Arbitrage (see p.144) that we must have $\Pi_T(\mathcal{X}) = \mathcal{X}$ since otherwise we could borrow money to purchase the lesser valued item and immediately sell it at the higher price.

It follows from the definition of a replicating portfolio that $\mathcal{X} = V_T^H$. This proves in conjunction with $\Pi_T(\mathcal{X}) = \mathcal{X}$ that $\Pi_T(\mathcal{X}) = V_T^H$.

Let us now assume that there is some $0 \le t_0 < T$ such that $\Pi_{t_0}(\mathcal{X}) \neq V_{t_0}^H$. We examine separately the cases $\Pi_{t_0}(\mathcal{X}) < V_{t_0}^H$ and $\Pi_{t_0}(\mathcal{X}) > V_{t_0}^H$ and show that each one allows for arbitrage opportunities.

Case I: $\Pi_{t_0}(\mathcal{X}) > V_{t_0}^H$

1. $t = t_0$: We sell short a claim $\mathcal{X}$ at a price of $\Pi_{t_0}(\mathcal{X})$.
2. $t = t_0$: We use the proceeds to purchase a replicating portfolio $\vec{H}_{t_0}$ at its value, $V_{t_0}^H$.
3. We create a separate portfolio by investing the difference $\Delta := \Pi_{t_0}(\mathcal{X}) - V_{t_0}^h$ in the riskless asset.
4. Compounded interest will make that investment grow to $\Delta' \geq \Delta$ at time $t = T$. The specific value of Δ' will depend on the interest rate process.
5. The original portfolio will grow in value from $V_{t_0}^H$ at time $t = t_0$ to V_T^H at time $t = T$. We then sell the portfolio and use that money to buy one unit of the claim. We use that security to cover the short sale that happened at $t = t_0$.
6. We have made a profit of Δ' without investing any of our own money.

Case II: $\Pi_{t_0}(\mathcal{X}) < V_{t_0}^H$

1. $t = t_0$: We sell short a hedge $\vec{H}_{t_0}$ for $\mathcal{X}$ at a price of $V_{t_0}^H$.
2. $t = t_0$: We use the proceeds to purchase a claim $\mathcal{X}$ at a price of $\Pi_{t_0}(\mathcal{X})$.
3. We create a separate portfolio by investing the difference $\Delta := V_{t_0}^h - \Pi_{t_0}(\mathcal{X})$ in the riskless asset.
4. That investment will grow to Δ' at time $t = T$.
5. $\mathcal{X}$ will be worth V_T^H at time $t = T$ since $\vec{H}$ replicates this claim. We then sell the claim, buy $\vec{H}$ from the proceeds, and use $\vec{H}$ to cover the short sale that happened at time $t = t_0$.
6. We have made a profit of Δ' without investing any of our own money. ■

7.3 The Holdings Process of a Riskless Asset

In this subchapter we assume, for simplicity, that the financial market consists of

- a single riskless asset (e.g., bank account) $\mathscr{A}^B$,
- a aingle risky asset (e.g., stock) $\mathscr{A}^S$.

We write B_t for the money market account price and D_t for the discount process of $\mathscr{A}^B$.

For example, if trading happens at continuous time and $\mathscr{A}^B$ is governed by the interest rate process R_t, we have

For example, in a continuous time market with interest rate process R_t

- $B_t = \exp\left[\int_0^t R_s\,ds\right]$, • $D_t = 1/B_t = \exp\left[-\int_0^t R_s\,ds\right]$.

On the other hand, if trades occur at the years $0, 1, 2, \dots$ with a fixed, annual interest rate r,

- $B_t = (1+r)^n$, • $D_t = 1/B_t = (1+r)^{-n}$.

Introduction 7.2. Associated with the assets vector $\vec{\mathscr{A}} = \left(\mathscr{A}^B, \mathscr{A}^S\right)$ is the portfolio $\vec{H}_t = \left(H_t^B, H_t^S\right)$. The value of this portfolio is

$$V_t = H_t^B\, S_t^{(0)} + H_t^S\, S_t. \tag{7.18}$$

Here, $S_t^{(0)}$ is the price per share of the riskless asset and $S_t = S_t^{(1)}$ is the price of the risky asset.

We announced in Remark 7.4 on p.146 that $S_t^{(0)}$ equals the money market account price and that this justifies to write $B_t = S_t^{(0)}$.[32]

For convenience, we define the symbols

$$V_t^B := H_t^B\, S_t^{(0)}, \qquad V_t^S := H_t^S\, S_t\,. \tag{7.19}$$

[32] See Notation 7.1 on p.146.

Thus V_t^B is the money value of the bank account holdings, and V_t^S is the money value of the stock holdings of the portfolio $\vec{H}_t$. Note that $V_t = V_t^B + V_t^S$.

It is clear how to interpret the equation $V_t^S = H_t^S S_t$. If today's stock price is S_t dollars and I hold H_t^S shares of that stock, then those shares contribute V_t^S to my overall portfolio value V_t. For example, if I hold 20 shares of stock and each share's current value is 30, then my holdings in that stock are worth 600.

Bank account holdings are approached completely differently. Consider a balance of $1,000$ dollars in that account. Then V_t^B should certainly be \$1,000. But what about H_t^B and $S_t^{(0)}$?

The obvious approach is to say that a dollar is a dollar, so one dollar should be one share in the bank account, and H_t^B, the number of shares, should be $1,000$. Let us assume the account was established about a year ago with a balance of \$980, no money was deposited or withdrawn ever since, and the \$20 increase is due to interest earned on the deposit. Then our approach would imply that the unit value per share remained the same (one dollar), and the holdings increased from 980 shares to $1,000$ shares. This is completely different from investments in a stock, where the number of shares remains unchanged if no trades are executed.

It turns out that it is more advantageous for finance modeling to take another, much less obvious approach and to define bank shares and bank share prices in such a way that they behave like stock shares and stock price.

- If I own 100 shares of the stock at $t_0 = 0$, i.e., $H_0^S = 100$, and if I do not make any trades in that stock until time t_1, then my holding in that asset remain constant: $H_t^S = H_0^S = 100$, for $0 \le t \le t_1$.
- However, the (money) value V_t^S of those holdings will change, since $V_t^S = H_t^S \cdot S_t$, and the stock price does not remain constant.

The analogous situation for investing in the riskless asset would be as follows.

- I consider one currency unit (one dollar), [33] invested AT TIME $t = 0$, to be one **bank share**, i.e., one share of that asset.
- If I own 100 bank shares at $t_0 = 0$, i.e., $H_0^B = 100$, and if I do not make any trades in the riskless asset until time t_1, then my bank account holding (the number of bank shares) should remain constant: $H_t^B = H_0^B = 100$, for $0 \le t \le t_1$.

However, the (money) value V_t^B of those 100 dollars that were invested at time 0 will change, since interest is added, at certain times or continually, to that investment.

- Due to interest, today's value of one unit is B_t, today's money market account price. Accordingly, today's balance in that account is $100B_t = H_0^B\, B_t = H_0^B\, B_t$.
- Interest payments have increased, today's value of each bank share to B_t, today's money market account price. Accordingly, today's balance in the account is
$$V_t^B = 100B_t = H_0^B\, B_t = H_0^B\, B_t.$$
- In reverse, if I invest at time t 100 dollars into my bank account, this will not buy me 100 bank shares. (We just saw that $V_t^B = B_t \cdot 100$ dollars are needed for that.)
- Rather, the 100 dollars must be discounted to time 0 by multiplying this amount with $D_t = 1/B_t$, since today's 100 dollars only buy $H_t^B = D_t \cdot 100$ bank shares.

[33] if you prefer, Euro or Chinese Yuan or Rubel or ...

- Of course, the value of those H_t^B bank shares at time t is, due to accrued interest,
$$H_t^B \cdot B_t \;=\; (D_t \cdot 100) \cdot B_t \;=\; (B_t^{-1} \cdot 100) \cdot B_t \;=\; 100 \text{ dollars},$$
just as it should be.

In summary, good choices for shares in and price per share of a riskless asset would be
- $H_t^B = D_t V_t^B$, where V_t^B denotes today's worth of the investment in that asset,
- $S_t^{(0)} = B_t =$ today's money market account price. ☐

All that remains is to formalize the definition of bank shares and the finding that it results in the price per such a share being equal to the money market account price.

Definition 7.10 (Bank shares).

Assume that a riskless asset has a money market acccount price process $B_t(\omega)$ and discount process $D_t(\omega) = 1/B_t(\omega)$.
If the monetary value of that investment at time t is V_t^B dollars, then we call

(7.20) $$H_t^B(\omega) \;:=\; D_t(\omega) \cdot V_t^B(\omega) \;=\; \frac{V_t^B(\omega)}{B_t(\omega)}$$

the number of **bank shares** which the investor holds in that asset. ☐

Proposition 7.3. *Given the assumptions of this section, it follows that*

the price of a bank share at time t equals the money market account price of the asset, i.e.,

(7.21) $$S_t^{(0)} \;=\; B_t\,, \qquad \textit{for } t \;\geq\; 0.$$

PROOF: Follows from the material presented in the introduction to this section ■

Remark 7.9. Stock price S_t and stock holdings H_t^S have the following analogies for riskless assets:

(1) asset price per unit at a given time $t \;=\; B_t \;=\;$ money market account price at t,
(2) Holdings $H_t^B \;=\; D_t V_t^B \;=\;$ today's value discounted to time zero.

☐

7.4 Discrete Time Financial Markets

The following examples are about interest in discrete markets.

Example 7.4.

- In a discrete time financial market, interest is added only at the trading times t_j. See Definition 7.2**(4)** (Financial Market) on p.145.
- The most complicated case occurs when the trading times are not equally spaced, and different interest rates r_j per unit of time may occur during different intervals $[t_{j-1}, t_j[$. Then the account value jumps at t_j by a factor $1+(t_j-t_{j-1})r$, from $H^B_{t_j} B_{t_{j-1}}$ to $H^B_{t_j} B_{t_j} = H^B_{t_j} B_{t_{j-1}}\big(1+(t_j-t_{j-1})r\big)$.
- The most complicated case occurs when the trading times are not equally spaced, and different interest rates r_j per unit of time may occur during different intervals $[t_{j-1}, t_j[$. Then the money market account price jumps at t_j by a factor $1+(t_j-t_{j-1})r$, from $B_{t_{j-1}}$ to $B_{t_j} = B_{t_{j-1}}\big(1+(t_j-t_{j-1})r\big)$.
- For example, if the annual interest rate is 4% and trades occur one per day and t_j denotes day j, then the daily interest rate is $r = \frac{4}{365}\%$, and interest earned increases the money in the account at each day j by a factor of $1 + \frac{4}{36{,}500}$. $\square$

Here are some remarks concerning the portfolio value process $V^{\vec{H}}$ [34] of a discrete market.

Remark 7.10. Recall that $\vec{H}_{t_0} = \vec{H}_{t_1}$ by the definition of a discrete market portfolio. Thus

$$V^{\vec{H}}_{t_0} = \vec{H}_{t_1} \bullet \vec{S}_{t_0} = \sum_{j=0}^{n} H^{(j)}_{t_1} S^{(j)}_{t_0}, \quad \square \tag{7.22}$$

Remark 7.11. Portfolio value is interpreted differently in discrete and continuous trading models. In discrete time markets there are two cases to consider.

A. The case $t_k > t_0$, i.e., $k > 0$.

We interpret, for each trading time $t_k > t_0$, $\vec{H}_{t_k}$ as the holdings in asset $\mathscr{A}^{(j)}$ during the interval $[t_{k-1}, t_k[$. In other words, the quantities $\vec{H}_{t_k}$ are bought and sold at time t_{k-1} and held constant until the next time of trade t_k. The times $t_1, t_2, \dots$ are genuine times of trade.

The following happens at $t = t_k$:

(a) The entire "old" portfolio $\vec{H}_{t_k}$, which was purchased at time t_{k-1} at prices $\vec{S}_{t_{k-1}}$, is sold at current prices $\vec{S}_{t_k}$.
The money received from that sale is $V_{t_k} = \vec{H}_{t_k} \bullet \vec{S}_{t_k}$.

(b) This amount V_{t_k} now is used to purchase the new portfolio $\vec{H}_{t_{k+1}}$. This purchase also happens at current prices $\vec{S}_{t_k}$.
Since money spent = money received = V_{t_k}, we have $V_{t_k} = \vec{H}_{t_{k+1}} \bullet \vec{S}_{t_k}$.

> **Important**: The "obvious" portfolio value equation
>
> $$V_{t_k} = \vec{H}_{t_k} \bullet \vec{S}_{t_k}$$
>
> applies to the sale of the old portfolio $\vec{H}_{t_k}$, but NOT to the purchase of the new portfolio $\vec{H}_{t_{k+1}}$! $\square$

The equation

$$V_{t_k} = \vec{H}_{t_k} \bullet \vec{S}_{t_k} = \vec{H}_{t_{k+1}} \bullet \vec{S}_{t_k} \tag{7.23}$$

[34] See Definition 7.6 (Portfolio value) on p.149

expresses that no money is added or removed when the old portfolio $\vec{H}_{t_k}$ is traded for the new portfolio $\vec{H}_{t_{k+1}}$ Thus

money spent = money received.

B. The case $k = 0$.

The time t_0 is the setup time for the initial portfolio $\vec{H}_{t_0}$. There is no old portfolio which can be traded for this initial portfolio. Rather, the first time of trade is t_1.

Recall that $\vec{H}_{t_0} = \vec{H}_{t_1}$ by definition. The following happens at t_0:

- The amount V_{t_0} is available to setup (buy) the initial portfolio $\vec{H}_{t_0}$. This purchase takes place at current prices $\vec{S}_{t_0}$. Since the money spent at setup is V_{t_0}, this is the value of the portfolio $\vec{H}_{t_0}$. In other words,

(7.24) $$V_{t_0} \;=\; \text{Portfolio setup value} \;=\; \vec{H}_{t_0} \bullet \vec{S}_{t_0} \;=\; \vec{H}_{t_1} \bullet \vec{S}_{t_0}\,. \;\square$$

We refer to (7.23), the equation which expresses the "money spent = money received" balance when a portfolio is traded for a new one, as the budget equation for the portfolio. This equation also allows us to amend Definition 7.5 (Self–financing portfolio) on p.149 to one that is mathematically more precise.

Definition 7.11 (Discrete time budget equation and self–financing portfolios).

(A.) The **budget equation** for a portfolio $\vec{H}_t$ in a discrete time financial market is

(7.25) $$\sum_{j=0}^{n} H^{(j)}_{t_{k+1}} S^{(j)}_{t_k} \;=\; V^{\vec{H}}_{t_k} \;=\; \sum_{j=0}^{n} H^{(j)}_{t_k} S^{(j)}_{t_k} \quad \text{for } t_k > t_0.$$

(B.) We call $\vec{H}_t$ a **self–financing portfolio strategy** aka **self–financing portfolio**, if it satisfies this budget equation. $\square$

7.5 Continuous Time Financial Markets

We recall from part **(4)** of Definition 7.2 (Financial Market) on p.145, that interest is compounded continuously in a continuous time market. Clearly, this implies the following.

Proposition 7.4.

*Assume that R_t is an **interest rate process**, for the riskless asset $\mathscr{A}^{(0)}$, i.e., $R_t(\omega)$ is the interest rate given at time t. Then each dollar invested into the asset at time zero will increase to*

(7.26) $$B_t \;:=\; \exp\left[\int_0^t R_s\,ds\right]$$

at time t. Thus, B_t is the money market account price, and

(7.27) $$D_t \;:=\; \frac{1}{B_t} \;=\; \exp\left[-\int_0^t R_s\,ds\right]$$

is the discount process of $\mathscr{A}^{(0)}$.

Remark 7.12.

- More generally, between times $t_1 < t_2$, the holdings in a money market account increase during the interval $[t_1, t_2[$ by the factor

$$\frac{B_{t_2}}{B_{t_1}} = \exp\left[-\int_{t_1}^{t_2} R_s\,ds\right].$$

- In the special case of a constant interest rate r on the interval $[0,T]$, those holdings increase during the interval $[t_1, t_2[$ by the factor $e^{(t_2-t_1)r}$. $\square$

Here are some remarks concerning the portfolio value process $V^{\vec{H}}$ [35] of a continuous time market.

Example 7.5. If $\mathscr{A}^{(3)}$ denotes IBM stock which is traded at time t at a price of $S_t^{(3)} = \$120.15$ per share and $H_t^{(3)} = -27.78$ shares, (a short position!) then IBM stock contributes -3337.767 dollars to the value $V_t^{\vec{H}}$ of that portfolio. $\square$

Remark 7.13. Portfolio value is interpreted differently in discrete and continuous trading models. In continuous time markets, each time t is a trading time. We interpret $\vec{H}_t$ as the holdings (number of shares) in asset $\mathscr{A}^{(j)}$ at that time t. The value of those $\mathscr{A}^{(j)}$–holdings is

$$\text{quantity} \times \text{price} = H_t^{(j)} \cdot S_t^{(j)}.$$

Thus the sum of those holdings, $\sum\limits_{j=0}^{n} H_t^{(j)} S_t^{(j)}$, is the value of the entire portfolio at time t. $\square$

Remark 7.14. Definition 7.11 (Discrete time budget equation and self–financing portfolios) on p.156 gave the budget equation for a discrete trading times financial market. Such a budget equation can also be formulated for continuous trading times, It turns out to be

$$dV_t^H = \vec{H}_t \bullet d\vec{S}_t = \sum_{i=1}^{N} H_t^{(i)}\,dS_t^{(i)}, \tag{7.28}$$

where $dS_t^{(i)} = dS_t^{(i)}(\omega)$ is a "stochastic differential". We need knowledge of stochastic calculus to understand the meaning of (7.28), so we will defer dealing with continuous time budget equations until Chapter 10.1 (Prologue: The Budget Equation in Continuous Time Markets). [36] $\square$

7.6 Exercises for Ch.7

[35] See Definition 7.6 (Portfolio value) on p.149

[36] See Definition 10.1 on p.204.

8 The Binomial Asset Model

A very simple financial market model is the binomial model. It is characterized as follows.

Assumption 8.1 (Binomial Asset Model)**.**

Trading only happens at times $t = 0, 1, 2, \dots$ (we have a discrete time financial market in the sense of Definition 7.2 (Financial Market) on p.145), and there are only two assets:

(1) $\mathscr{A}^B$ is a bond/bank account. We denote its money market account price at time t by B_t. Interest is compounded only at the trading times $t = 1, 1, 2, \dots$ (no interest is due yet at start time zero), and the interest rate r is fixed and deterministic. Thus

(8.1) $$B_1 = (1+r)B_0, \quad \dots, \quad B_n = (1+r)B_{n-1} = (1+r)^n B_0.$$

(2) $\mathscr{A}^S$ is a stock. We denote its price process by S_t.

(3) S_t remains unchanged between trading times. At the next such time it will either go up by a factor u with a probability p_u, or it will do down by a factor d with a probability p_d. Thus the dynamics for S_t are

(8.2) $$S_n = S_{n-1} \cdot Z_n = \begin{cases} u \cdot S_{n-1}, & \text{with probability } p_u > 0, \\ d \cdot S_{n-1}, & \text{with probability } p_d > 0\,, \end{cases}$$

(8.3) Here $$Z_n := \begin{cases} u, & \text{with probability } p_u > 0, \\ d, & \text{with probability } p_d > 0\,. \end{cases}$$

is an iid sequence of binomial random variables with success probability p_u.

(4) We assume that $B_0 = 1$ and S_0 has the deterministic value $S_0 = s$.

(5) We assume that trading ends at time T (an integer). The meaning of T will often be the time of expiry of a contingent claim. □

Remark 8.1 (Portfolio Strategy for the binomial model)**.**

According to Definition 7.4 (Portfolio Strategy) on p.147

a portfolio strategy for the binomial asset model is a process

(8.4) $$\vec{H}_t(\omega) = \left(H_t^B(\omega), H_t^S(\omega)\right), \quad t = 1, 2, \dots, T$$

which denotes the holdings H_t^B in $\mathscr{A}^B$ and H_t^S in $\mathscr{A}^S$ of an investor during the interval $[t-1, t]$. Negative values indicate that this quantity is not owned but owed.

Its portfolio value is

(8.5) $$\begin{aligned} V_0^{\vec{H}} &= H_1^B B_0 + H_1^S S_0, \\ V_t^{\vec{H}} &= H_t^B B_t + H_t^S S_t \text{ if } t > 0, \text{ at time of sale.} \ \square \end{aligned}$$

Note that, according to Definition 7.4**(3)**, $\vec{H}_0$ is defined by $\vec{H}_0 = \vec{H}_1$.

We next specify the budget equation that must be satisfied by a self–financing portfolio. See Definition 7.11 (Budget Equation) on p.156.

Proposition 8.1 (Budget equation in the binomial asset model)**.** *A portfolio strategy*

$$\vec{H}_t(\omega) \ = \ \big(H_t^B(\omega), H_t^S(\omega)\big), \quad t = 1, 2, \dots, T$$

for the binomial asset model is self–financing if and only if the following condition holds.

Budget equation:

$$H_t^B(1+r)^t \ + \ H_t^S S_t \ = H_{t+1}^B(1+r)^t \ + \ H_{t+1}^S S_t \quad (t = 1, \dots, T-1). \tag{8.6}$$

PROOF: H_t^B "bank shares" is the amount of money one would have had to deposit at time 0 to obtain, due to compound interest, the bank account balance $H_t^B B_{t-1} = H_t^B(1+r)^{t-1}$ that belongs to the new portfolio $\vec{H}_t$ purchased at time $t-1$.
This money in the bank increases during the interval $[t-1, t]$ by a factor $1+r$ to $H_t^B(1+r)^t$. In other words, the bank account portion of $\vec{H}_t$ has become $H_t^B(1+r)^t$ at time t.
Clearly, the value of the stock shares was $H_t^S S_{t-1}$ at time $t-1$ and has changed to $H_t^S S_t$ at time t.
Thus the sales value of $\vec{H}_t$ is $V_t^{\vec{H}} \ = \ H_t^B(1+r)^t \ + \ H_t^S S_t$.
We use that money to purchase (still at time t) the new portfolio $\vec{H}_{t+1}$. Its bank account portion is worth $H_{t+1}^B B_t = H_{t+1}^B(1+r)^t$, the H_{t+1}^S shares of stock are worth $H_{t+1}^S S_t$, thus the value of the new portfolio is $H_{t+1}^B(1+r)^t \ + \ H_{t+1}^S S_t$.
The budget equation states that this amount must equal the sales value of the old portfolio. Hence,

$$V_t^{\vec{H}} \ = \ H_t^B(1+r)^t \ + \ H_t^S S_t \ = \ H_{t+1}^B(1+r)^t \ + \ H_{t+1}^S S_t \ \blacksquare$$

One of the key properties of the binomial asset model will be that, if it does not admit arbitrage, one can replace the probabilities p_u and p_d which were introduced in Assumption 8.1**(3)** made about the binomial asset model (p.158), with different probabilities $\tilde{p}_u$ and $\tilde{p}_d$. Those two numbers then define a probability $\widetilde{P}$ on $\mathfrak{F}^S = \sigma\{S_0, S_1, \dots\}$ which is equivalent to P and makes discounted stock price $(1+r)^{-n} S_n$ a $\widetilde{P}$–martingale. We collect here some material which will help establish that fact.

Proposition 8.2.

If $(\Omega, \mathfrak{F}, P)$ *is a probability space,* $n \in \mathbb{N}$ *and* $A_1, \dots, A_n \in \mathfrak{F}$*, then*

$$\begin{aligned} P(A_n \cap A_{n-1} \cap \dots \cap A_1) \ &= P(A_n \mid A_{n-1} \cap \dots \cap A_1)\, P(A_{n-1} \mid A_{n-2} \dots \cap A_1) \cdots \\ &\cdots P(A_3 \mid A_2 \cap A_1)\, P(A_2 \mid A_1)\, P(A_1). \end{aligned} \tag{8.7}$$

PROOF:

Repeated use of $P(U \cap V) = P(U \mid V)P(V)$ with $U = A_j$ and $V = A_{j-1} \cap \cdots \cap A_1$ yields

$$\begin{aligned}
&P(A_n \cap A_{n-1} \cap \cdots \cap A_1)\\
&= P(A_n \mid A_{n-1} \cap \cdots \cap A_1)\, P(A_{n-1} \cap \cdots \cap A_1)\\
&= P(A_n \mid A_{n-1} \cap \cdots \cap A_1)\, P(A_{n-1} \mid A_{n-2} \cdots \cap A_1)\, P(A_{n-2} \cdots \cap A_1)\\
&= \ldots\ldots\ldots\ldots\ldots\ldots\\
&= P(A_n \mid A_{n-1} \cap \cdots \cap A_1)\, P(A_{n-1} \mid A_{n-2} \cdots \cap A_1) \cdots P(A_3 \mid A_2 \cap A_1)\, P(A_2 \mid A_1)\, P(A_1). \blacksquare
\end{aligned}$$

Proposition 8.3.

*Let the process $X = \left(X_j\right)_{j=0,1,\ldots}$ follow a **binomial tree model**, i.e., there exist $x_0, u, d, \pi_u, \pi_d \in \mathbb{R}$ such that $u < d$ and*

(8.8) $$X_0 = x_0 = \text{const},$$

(8.9) $$\pi_u > 0, \quad \pi_d > 0, \quad \pi_u + \pi_d = 1,$$

(8.10) $$\begin{aligned} \text{either } X_{n+1} &= X_n u \ \text{ with probability } \pi_u \quad (\text{“upward move”}),\\ \text{or } X_{n+1} &= X_n d \ \text{ with probability } \pi_d \quad (\text{“downward move”}). \end{aligned}$$

Then π_u and π_d determine a probability P on the measurable space $(\Omega, \sigma\{X_0, X_1, \ldots\})$.

This probability is characterized as follows.

Assume that the path $x_1 = X_1(\omega), x_2 = X_2(\omega), \ldots, x_n = X_n(\omega)$ *consists of k upward moves $x_{j+1} = x_j u$ and of $n-k$ upward moves $x_{j+1} = x_j d$. Then*

(8.11) $$P\{X_0 = a_0, X_1 = x_1, \ldots, X_n = x_n\} = \pi_u^k\, \pi_d^{n-k},$$

(8.12) $$P\{X_n = x_0\, u^k\, d^{n-k}\} = \binom{n}{k} \pi_u^k \pi_d^{n-k}.$$

In particular, the number of upward moves of X_n has a binom$(n; \pi_u)$ distribution.

PROOF: ⋆

The process X has been constructed in such a fashion that X_n will be one of $x_0\, u^j\, d^{n-j}$ where $j = 0, 1, \ldots, n$. X_{n+1} only depends on X_n and not on the prior values $X_0, \ldots, X_{n-1}$, thus

(8.13) $$P(X_{n+1} = a \mid \sigma(X_1, \ldots, X_n)) = P(X_{n+1} = a \mid X_1, \ldots, X_n) = P(X_{n+1} = a \mid X_n)$$

for any number a. It follows from (8.10) that

(8.14) $$P\{X_n = x_n \mid X_{n-1} = x_{n-1}\} = \begin{cases} \pi_u & \text{if } x_n = x_{n-1}u,\\ \pi_d & \text{if } x_n = x_{n-1}d,\\ 0 & \text{else}. \end{cases}$$

Let $x_0, \ldots, x_n$ such that

(8.15) $$x_j = u\, x_{j-1} \quad \text{or} \quad x_j = d\, x_{j-1} \qquad (j = 1, 2, \ldots, n).$$

Then (8.13) and (8.14) yield

$$\begin{aligned} P\{X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} &= P\{X_n = x_n \mid X_{n-1} = x_{n-1}\} \\ &= \begin{cases} \pi_u & \text{if } x_n = x_{n-1}u, \\ \pi_d & \text{if } x_n = x_{n-1}d. \end{cases} \end{aligned} \tag{8.16}$$

The condition (8.15) is necessary for the following reason: If it is not satisfied then $P\{X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} = 0$, and the leftmost conditional probability is not defined.

Case 1: Assume that the numbers $x_0, \dots, x_n$ satisfy the condition (8.15). We apply (8.16) to formula (8.7) of Proposition 8.2 on p.159 with $A_j = \{X_j = x_j\}$ $(j = 0, 1, \dots, n)$. We obtain

$$\begin{aligned} &P\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} \\ &= P\{X_0 = x_0\}\, P\{X_1 = x_1 \mid X_0 = x_0\} \cdots\; P\{X_n = x_n \mid X_{n-1} = x_{n-1}\}\,. \end{aligned} \tag{8.17}$$

If the event A describes k upward moves and thus $n-k$ downward moves of the process, i.e., there are k indices j such that $x_j = u\, x_{j-1}$ and $n-k$ indices j such that $x_j = u\, x_{j-1}$, then the above equals, since $P\{X_0 = x_0\} = 1$,

$$P\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} \;=\; \pi_u^k \pi_d^{n-k}.$$

We have derived formula (8.11) of this proposition.

Case 2: If $x_0, \dots, x_n$ do not satisfy (8.15) then $P\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\} = 0$. Let

$$B \;:=\; \big\{(x_0, \dots, x_n) : P\{X_0 = x_0, \dots, X_n = x_n\} = 0\big\}\,.$$

By construction, each X_k can only take one of the $k+1$ values $x_0 u^j d^{k-j}$ where $j = 0, 1, \dots, k$. Thus the size of B is finite, thus

$$P\{(X_0, \dots, X_n) \in B\} \;=\; \sum \big[\, P\{X_0 = x_0, \dots, X_n = x_n\} \,:\, (x_0, \dots, x_n) \in B \,\big] \;=\; \sum 0 \;=\; 0$$

Both cases together show that the finite distributions of the process X are determined by formula (8.11) and thus by π_u and π_d.

The proof of (8.12) is obtained as follows. Observe that, for any $0 \le j \le n$,

$$X_n(\omega) \;=\; x_0\, u^j\, d^{n-j} \quad \Leftrightarrow \quad \text{there were } j \text{ upward moves and } n-j \text{ downward moves}\,,$$

and that there as many combinations of k upward moves and $n-k$ downward moves as there are ways to select k items from n items. According to (8.11) each one of those combinations occurs with the same probability $\pi_u^k \pi_d^{n-k}$. It follows that

$$P\{X_n \;=\; x_0\, u^j\, d^{n-j}\} \;=\; \binom{n}{k} \pi_u^k\, \pi_d^{n-k}\,. \;\blacksquare$$

Corollary 8.1. *In the settings of Proposition 8.3 we assume that $\Phi(x)$ is a function which is defined for all values x which can be assumed by X_n, i.e., for $x \in \{x_0 u^k d^{n-k} : k = 0, 1, \dots, n\}$. Then*

$$E[\Phi(X_n)] \;=\; \sum_{k=0}^{n} \binom{n}{k} \pi_u^k \pi_d^{n-k}\, \Phi(x_0 u^k d^{n-k}). \tag{8.18}$$

PROOF:
We have seen in Proposition 8.3 that the number of upward moves of X_n follows a binom$(n;\pi_u)$ distribution, i.e.,

(8.19) $$P\{X_n = x_0\, u^k\, d^{n-k}\} = \binom{n}{k}\pi_u^k \pi_d^{n-k} \quad (k = 0, 1, \dots, n)\,.$$

For $X_n(\omega) = x_0 u^k d^{n-k}$ we obtain

$$\Phi\big(X_n(\omega)\big) = \Phi(x_0 u^k d^{n-k}) = \psi(k)\,.$$

It follows for the expected value of $\mathcal{X}$ that

$$\begin{aligned} E[\Phi(X_n)] &= \sum_x \Phi(x)\, P\{X_n = x\} \\ &= \sum_{k=0}^{n} \Phi(x_0 u^k d^{n-k})\, P\{X_n = x_0 u^k d^{n-k}\} \\ &= \sum_{k=0}^{n} \Phi(x_0 u^k d^{n-k}) \binom{n}{k}\pi_u^k \pi_d^{n-k}\,. \quad \blacksquare \end{aligned}$$

(see (8.19))

It will be almost immediate from the next proposition that if $M_t = D_t S_t$ (S_t = stock price) then M_t is a $\mathfrak{F}_t^S$–martingale under risk–neutral probability.

Proposition 8.4.

Let the process X and the probability measure P constructed on $(\Omega, \sigma\{X_0, X_1, \dots\})$ be as defined in Proposition 8.3 (see p.160). We write E_P for the expectation with respect to that probability and, as usual, $\mathfrak{F}_n^X = \sigma\{X_0, X_1, \dots, X_n\}$. Then

(8.20) $$E_P[X_{n+1} \mid \mathfrak{F}_n^X] = (u\,\pi_u + d\,\pi_d) X_n\,.$$

PROOF:
According to formula (5.6) on p.111:

$$X_{\mathfrak{G}}(\omega) = \sum_j E(X \mid G_j) \cdot \mathbf{1}_{G_j}(\omega)\,,$$

applied to $\mathfrak{G} := \sigma\{X_j : j \le n\}$. $\mathfrak{G}$ is generated by the sets $\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$. Such a set has probability zero unless $x_j = x_{j-1}u$ or $x_j = x_{j-1}d$ for each $j = 1, 2, \dots, n$.
Since conditional expectations are determined only up to a set of probability zero, (8.20) is valid if we can prove the following. Let

$$\begin{aligned} A &:= \{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \text{ such that } P(A) > 0, \text{ i.e.,} \\ x_j &= x_{j-1}u \text{ or } x_j = x_{j-1}d \quad \text{for all } j = 1, 2, \dots, n\,. \end{aligned}$$

Then

(**A**) $$E_P[X_{n+1} \mid A] = (u\,\pi_u + d\,\pi_d)\, X_n(\omega) \quad \text{for all } \omega \in A\,.$$

To prove **(A)** we observe that

$$\begin{aligned}
E_P[X_{n+1} \mid A] &= \sum_x x \cdot P\{X_{n+1} = x \mid A\} \\
&= (x_n u) P\{X_{n+1} = x_n u \mid A\} + (x_n d) P\{X_{n+1} = x_n d \mid A\} \\
&= (x_n u) P\{X_{n+1} = x_n u \mid X_n = x_n, \dots, X_1 = x_1\} \\
&\quad + (x_n d) P\{X_{n+1} = x_n d \mid X_n = x_n, \dots, X_1 = x_1\}
\end{aligned}$$

It follows from the definition of π_u and π_d that

$$\begin{aligned}
P\{X_{n+1} &= x_n u \mid X_n = x_n, \dots, X_1 = x_1\} = \pi_u \,, \\
P\{X_{n+1} &= x_n d \mid X_n = x_n, \dots, X_1 = x_1\} = \pi_d \,,
\end{aligned}$$

thus $E_P[X_{n+1} \mid A] = (x_n u)\pi_u + (x_n d)\pi_d = x_n(u\,\pi_u + d\,\pi_d)$.
Since $A \subseteq \{X_n = x_n\}$, we have $X_n(\omega) = x_n$ and thus $E_P[X_{n+1} \mid A] = (u\,\pi_u + d\,\pi_d) X_n(\omega)$ for all $\omega \in A$. This proves **(A)** and, hence, (8.20). ■

Remark 8.2. Except for item **(1)**, this remark will be about stock price S_t and discounted stock price $D_t S_t$ rather than about a general binomial tree X_t.

(1) If $x_0 > 0$ and $d > 0$ (hence, $u > 0$), then $X_n(\omega) > 0$ for all n and all ω.

(2) Stock price S_n follows a binomial tree model for which the above proposition applies if we restrict the events of Ω to $\mathfrak{F}^S = \sigma\{S_0, S_1, \dots\}$. This is true for the real world probabilities p_u, p_d of upward and downward moves which thus define a probability P on $(\Omega, \mathfrak{F}^S)$ via

$$P\{S_{n+1} = au \mid S_n = a\} := p_u \,, \qquad P\{S_{n+1} = ad \mid S_n = a\} := p_d \,.$$

Note though that (8.9) explicitly requires that both $p_u > 0$ and $p_d > 0$.

(3) Discounted stock price $M_n := D_n S_n$, where $D_n = (1+r)^{-n}$, also follows a binomial tree model under the real world probabilities p_u and p_d. To see this we write

$$u' := \frac{u}{1+r} \,, \quad d' := \frac{d}{1+r} \,, \quad Z_n(\omega) := \begin{cases} u & \text{if } S_{n+1}(\omega) = S_n(\omega) u \,, \\ d & \text{if } S_{n+1}(\omega) = S_n(\omega) d \,. \end{cases}$$

Then

$$M_{n+1} = D_{n+1} S_{n+1} = D_1 D_n S_n Z_n = M_n (D_1 Z_n) = \begin{cases} M_n u' & \text{with probability } p_u \,, \\ M_n d' & \text{with probability } p_d \,. \end{cases}$$

Thus we have the following: If we replace S_n with M_n, u with u', d with d' and keep B_n, p_u, p_d unchanged, then the new system satisfies formula (8.2) on p.158. Thus we have again a binomial asset model.

(4) We claim that $\mathfrak{F}^M = \mathfrak{F}^S$. The proof is as follows. $\mathfrak{F}^S_n$ is generated by the events

$$\{S_0 = s_0,\, S_1 = s_1,\, \dots, S_n = s_n\} \quad \text{where} \quad s_k = u\,s_{k-1} \text{ or } s_k = d\,s_{k-1} \ (k = 1, 2, \dots, n) \,.$$

See the proof of Proposition 8.3 above. Since there is some $0 \le j \le k$ such that $s_k = s_0 u^j d^{k-j}$ (the case where s_k represents j upward moves and $k - j$ downward moves) for each $k = 1, 2, \dots, n$, and that is the case if and only if

$$M_k(\omega) = s_0 (u')^j (d')^{k-j} = D_k s_0 u^j d^{k-j} = D_k s_k \,,$$

it follows that

$$\{S_0 = s_0,\, S_1 = s_1,\, \dots, S_n = s_n\} \ = \{M_0 = s_0,\, M_1 = D_1 s_1,\, \dots, M_n = D_n s_n\}\,,$$

and thus that $\mathfrak{F}^M = \mathfrak{F}^S$.

(5) All that was discussed in **(2)** and **(3)** remains in force if we replace the real world probabilities p_u and p_d with the risk neutral probabilities $\tilde{p}_u$ and $\tilde{p}_d$. as long as both $\tilde{p}_u > 0$ and $\tilde{p}_d > 0$, i.e.,

$$d \,<\, 1 + r \,<\, u\,.$$

Note that the resulting probability $\widetilde{P}$ on $(\Omega, \mathfrak{F}^{\mathfrak{S}})$ is equivalent to P since

$$\begin{aligned} & P\{S_0 = s_0,\, S_1 = s_1,\, \dots, S_n = s_n\} \ > \ 0 \\ \Leftrightarrow\ & P\{S_0 = s_0,\, S_1 = s_1,\, \dots, S_n = s_n\} \ = \ p_u^k p_d^{n-k} \ \text{ for some } \ k = 1, 2, \dots, n \\ \Leftrightarrow\ & \widetilde{P}\{S_0 = s_0,\, S_1 = s_1,\, \dots, S_n = s_n\} \ = \ \tilde{p}_u^k \tilde{p}_d^{n-k} \ \text{ for some } \ k = 1, 2, \dots, n\,. \ \square \end{aligned}$$

We now return to examining the properties of a general binomial tree.

Theorem 8.1. *With the same definitions as before we have the following.*

Let the process M be defined as

$$M_n \ := \ \frac{1}{(u\pi_u + d\pi_d)^n}\, X_n\,.$$

Then M is both an $\mathfrak{F}_t^X$–martingale and an $\mathfrak{F}_t^M$–martingale)

PROOF: Let $\alpha := u\pi_u + d\pi_d$. Then $M_n = \frac{1}{\alpha^n} X_n$ and $X_n = \alpha^n M_n$. Deterministic expressions can be moved in and out of conditional expectations. Further, according to Proposition 8.4 on p.162, $E_P[X_{n+1} \mid \mathfrak{F}_n^X] = \alpha X_n$. Thus

$$E_P[M_{n+1} \mid \mathfrak{F}_n^X] \ = \ \alpha^{-(n+1)} E_P[X_{n+1} \mid \mathfrak{F}_n^X] \ = \ \alpha^{-(n+1)}(\alpha X_n) \ = \ \alpha^{-n} X_n \ = \ M_n\,.$$

It follows that M is an $\mathfrak{F}_t^X$–martingale. We have seen in Remark 8.2 that $\mathfrak{F}_t^M = \mathfrak{F}_t^X$. Thus M also is an $\mathfrak{F}_t^M$–martingale.

■

Considering that the stock price joint probabilities are given by

$$\begin{aligned} P\{S_0 = a_0, S_1 = s_1, \dots, S_n = s_n\} \ &= \ p_u^k p_d^{n-k} \qquad \text{in the real world}, \\ \widetilde{P}\{S_0 = a_0, S_1 = s_1, \dots, S_n = s_n\} \ &= \ \tilde{p}_u^k \tilde{p}_d^{n-k} \qquad \text{in the risk–neutral world}, \end{aligned}$$

and the number of upward moves of stock price at time T follow a binomial distribution in both worlds (see (8.11) and (8.12) on p.160 and Remark 8.2**(2)** on p.163), it should not come as a surprise that the options price process $\Pi_T(\mathcal{X})$ for a simple claim $\mathcal{X}$, and thus also the identical portfolio value process V_t^H for a replicating portfolio $\vec{H}_t$, have a close connection with the binomial distribution.

Corollary 8.2 (Expectation of a simple claim in the binomial tree model)**.** *Let π_u and π_d be the risk–neutral probabilities for up and down moves of stock price. Then the expected value of a simple claim $\mathcal{X} = \Phi(S_T)$ is*

$$E[\mathcal{X}] = \sum_{k=0}^{T} \binom{T}{k} \pi_u^k \pi_d^{T-k} \Phi(s u^k d^{T-k}). \tag{8.21}$$

PROOF:

This follows from Corollary 8.1 on p.161. ■

8.1 The One Period Model

In the one period model there are only two times $t = 0$ and $t = 1$. A portfolio $\vec{H}_1 = (H_1^B, H_1^S)$ is purchased at $t = 0$. [37]

We follow the notation of [7] Björk, Thomas: Arbitrage Theory in Continuous Time and write

$$x := H_1^B, \qquad y := H_1^S.$$

According to assumption 8.1, parts **(4)** and **(3)**, the value process is

- $V_0 = x \cdot B_0 + y \cdot S_0 = x + y \cdot s,$
- $V_1 = x(R+1) + ysZ.$

Proposition 8.5.

The model above is free of arbitrage if and only if the following conditions hold:

$$d < (1+r) < u. \tag{8.22}$$

Informal PROOF that if (8.22) does not hold then there will be arbitrage portfolios:

First case – We assume $u > d \geq 1 + r$: We borrow money from the bank and invest it in the stock, with a return at least as high as the interest we must pay on our loan. There is positive probability p_u that $Z = u$, and in this case we will not just break even but make a profit.

Second case – We assume $d < u \leq 1 + r$: We sell short the stock and invest the proceeds in the bank with a return guaranteed to be high enough to buy that stock on the market and deliver it to the buyer. There is positive probability p_d that $Z = d$, and in this case we will not just break even but make a profit.

The proof of the reverse direction is left as exercise 8.1. See p.183. ■

We focus on the stock price process $S = (S_0, S_1)$ and the discounted stock price $D_1 S_1$. Since $S_0 = s$ = const, $\sigma(S_0) = \{\emptyset, \Omega\}$. Let $A := \{S_1 = su\}$. Since either $S_1 = su$ or $S_1 = sd$, we obtain

$$A^{\complement} = \{S_1 = sd\}, \quad \sigma(S_1) = \{\emptyset, \Omega, A, A^{\complement}\}, \quad \sigma(S_0, S_1) = \sigma(S_1) = \{\emptyset, \Omega, A, A^{\complement}\}.$$

We thus have completely determined the filtration $(\mathfrak{F}_t^S)_{t=0,1}$ generated by S as

$$\mathfrak{F}_0^S = \{\emptyset, \Omega\}, \qquad \mathfrak{F}_1^S = \{\emptyset, \Omega, A, A^{\complement}\}.$$

[37] Recall that $\vec{H}_0 = \vec{H}_1$ = portfolio holdings established at time $t = 0$!

Let $\mathfrak{F} := \sigma(S_0, S_1) = \mathfrak{F}_1^S$, i.e., we restrict the probability space $(\Omega, \mathfrak{F}, P)$ to the events known by S. Then P is completely specified by p_u as follows.

$$P(\emptyset) = 0, \;\; P(\Omega) = 1, \;\; P(A) = p_u, \;\; P(A^{\complement}) = p_d = 1 - p_u.$$

The relation $d < (1+r) < u$ yields a unique number $\tilde{p}_u$ such that $1+r$ is the convex combination

$$1 + r = (1 - \tilde{p}_u)d + \tilde{p}_u u = \tilde{p}_u u + \tilde{p}_d d \quad (\text{define } \tilde{p}_d := 1 - \tilde{p}_u). \tag{8.23}$$

This pair of numbers, $\tilde{p}_u$ and $\tilde{p}_d$, defines a probability measure $\widetilde{P}$ on $(\Omega, \mathfrak{F})$ via

$$\widetilde{P}(\emptyset) := 0, \;\; \widetilde{P}(\Omega) := 1, \;\; \widetilde{P}(A) := \tilde{p}_u, \;\; \widetilde{P}(A^{\complement}) := \tilde{p}_d = 1 - \tilde{p}_u. \tag{8.24}$$

To summarize, absence of arbitrage allows us to define a probability measure $\widetilde{P}$ on the information σ–algebra $\sigma(S_0, S_1) = \mathfrak{F}_1^S = \mathfrak{F}$ of stock price S such that

$$\tilde{p}_u u + \tilde{p}_d d = 1 + r\,.$$

It can easily be seen that $\widetilde{P}$ is equivalent to P. See Exercise 8.2 on p.183.

Now a reminder about the discount process. We have seen in formula (8.1) on p.158 that the interest factor by which a hank account holding increases between times zero and n is

$$B_n = (1+r)^n\,.$$

We can turn this around and see that an asset worth V_n at time n has to be discounted to $\frac{1}{(1+r)^n} V_n$ if one wants to determine how many units of the riskless asset $\mathscr{A}^B$ are needed at $t = 0$ to generate the amount V_n at time n. It follows that the discount process in the binomial model is given by

$$D_0 = 1, \;\; D_1 = \frac{1}{1+r}, \; ; \dots, \; D_n = \frac{1}{(1+r)^n}, \;\; \dots \tag{8.25}$$

This is, of course, just as it must be, since discount process D_t and price of money market account B_t are always reciprocal to each other.

Proposition 8.6. *The measure $\widetilde{P}$ defined by $\tilde{p}_u$ (and $\tilde{p}_d = 1 - \tilde{p}_u$) of formula (8.23) on $\mathfrak{F}_1^S$ satisfies*

(a) *The present stock price is obtained from its price in the future by discounting that one and taking its expectation with respect to the measure $\widetilde{P}$:*

$$S_0 = \frac{1}{1+r} \cdot \widetilde{E}[S_1], \tag{8.26}$$

(b) *The discounted stock price $M_n = D_n S_n, n = 0, 1$, is an $\mathfrak{F}_n^S$–martingale.*

PROOF: Since

$$D_n = \frac{1}{(1+r)^n} = \frac{1}{(u\tilde{p}_u + d\tilde{p}_d)^n}\,,$$

we obtain **(b)** from Theorem 8.1 on p.164 by setting

$$\pi_u := \tilde{p}_u\,, \;\; \pi_d := \tilde{p}_d\,, \;\; X_n := S_n \;\; (n = 0, 1)\,.$$

For the proof of (8.26) we proceed as follows. For $n = 0, 1$, let $M_n := S_n/(1+r)^n$.
Since $S_0 = s = \text{const}$, $\widetilde{E}[S_0] = S_0$. Since M_n is a $\widetilde{P}$–martingale, $\widetilde{E}[M_0] = \widetilde{E}[M_1]$. Thus

$$S_0 \;=\; \widetilde{E}[S_0] \;=\; \widetilde{E}[M_0] \;=\; \widetilde{E}[M_1] \;=\; \widetilde{E}\left[\frac{1}{1+r}S_1\right] \;=\; \frac{1}{1+r}\,\widetilde{E}[S_1]\,. \;\blacksquare$$

We give some definitions in the sequel which will be restated later in a more general context.

Definition 8.1 (Martingale Measure)**.**

We call a probability measure $\widetilde{P}$ for which discounted stock price D_tS_t is a martingale, a **martingale measure**. We also call $\widetilde{P}$ a **risk–neutral measure**, since, the equation

$$\widetilde{E}\left[D_{t+h}S_{t+h} \mid \mathfrak{F}_t\right] \;=\; D_tS_t \;\text{ for } h > 0\,,$$

has the following interpretation: On average, when we account for the riskless ("risk–neutral") growth by discounting S_{t+h} to $t = 0$, this discounted value must equal the (known) present value S_t of the asset if we also discount that one to $t = 0$. □

We now compute the probabilities $\tilde{p}_u$ and $\tilde{p}_d$ which determine the martingale measure $\widetilde{P}$.

Proposition 8.7.

The martingale probabilities $\tilde{p}_u$ and $\tilde{p}_d$ of formula (8.23) on p.166 can be explicitly computed as

(8.27) $$\tilde{p}_u \;=\; \frac{(1+r)-d}{u-d}\,, \qquad \tilde{p}_d \;=\; \frac{u-(1+r)}{u-d}\,.$$

PROOF: Trivial. ■

Remark 8.3 (Contingent Claim)**.** Since the expiration time is $T = 1$, a contingent claim (Definition 7.8 on p.151) in the one period model is a $\mathfrak{F}_1^S$–measurable random variable $\mathcal{X}(\omega)$. Note that

$$\mathfrak{F}_1^S \;=\; \sigma(S_0, S_1) \;=\; \sigma(S_1), \;\text{ since } S_0 = s = \text{ const.}$$

Thus, by Doob's composition lemma, there is a function $x \mapsto \Phi(x)$ of stock price x such that

$$\mathcal{X} = \Phi \circ S_1.$$

In other words, any contingent claim in the one period binomial model possesses a contract function Φ and thus is a simple claim. In a more general setting it will not always be true that all contingent claims are simple. □

To find an answer to the question how, in the one period model, a derivative $\mathcal{X}$ expiring at time $t = 1$ should be priced today, we work with replicating portfolios. In the general case a portfolio was the entire collection (process) $\vec{H} = \vec{H}_t$ since assets can be traded at any time t. In the discrete case $t = t_0 < t_1 < t_2 < \cdots < T$ trades only happen at times t_{j-1}, and those holdings

$$\vec{H}_{t_j} \;=\; \left(H^0_{t_j}, H^1_{t_j}, \dots, H^n_{t_j}\right)$$

remain constant until t_j. In the discrete case $t = t_0 < t_1 < t_2 < \cdots < t_m = T$, there is no more trade at expiration time $t_m = T$. Thus things are very simple in the one period model.

- Since $T = 1$, the only trade that influences V_T^H takes place at $t = 0$.
- There are only two assets, the bond (risk free asset) with prices $B_t = B_0, B_1$ (where $B_0 = 1$), and the stock (risky asset) $S_t = S_0, S_1$.

Our entire portfolio strategy can be described by two numbers $\vec{H}_0 = (x, y)$ which are deterministic since this portfolio is established at $t = 0$, and we know today what our holdings are today.

We recall our assumption that the market is efficient and that there is no arbitrage.

The next proposition shows us how to build a hedging portfolio for an arbitrary contract function.

Proposition 8.8.

Let the one period binomial model be free of arbitrage, i.e., $d < 1 + r < u$. Let $\mathcal{X}$ be an arbitrary claim with contract function Φ, i.e.,

$$\mathcal{X} = \Phi \circ S_1$$

Then this contract is hedged by the following portfolio $\vec{H}_1 = \big(H_1^B, H_1^S\big)$:

$$\begin{aligned} H_1^B &= \frac{1}{1+r} \cdot \frac{u\Phi(sd) - d\Phi(su)}{u-d}, \\ H_1^S &= \frac{1}{s} \cdot \frac{\Phi(su) - \Phi(sd)}{u-d}. \end{aligned} \tag{8.28}$$

Note for the above that $\Phi(x)$ is a function of stock price at $t = 1$, i.e., Φ is given by its two function values $\Phi(sd)$ and $\Phi(su)$.

PROOF: For convenience, let

$$x := H_1^B, \qquad y := H_1^S$$

be the portfolio which was established at $t = 0$. Thus we claim that that the portfolio $\vec{H}_1 = (x, y)$, given by

$$\begin{aligned} x &= \frac{1}{1+r} \cdot \frac{u\Phi(sd) - d\Phi(su)}{u-d}, \\ y &= \frac{1}{s} \cdot \frac{\Phi(su) - \Phi(sd)}{u-d}. \end{aligned} \tag{8.29}$$

is a hedge for $\mathcal{X}$. Rather than doing this the mathematically elegant way and showing that this choice of x and y will lead to the equation $V_1^H(\omega) = \mathcal{X}(\omega)$, we proceed the opposite way.

We recall from formulas (8.1) and (8.2) on p.158 that, since $S_0 = \text{const} = s$, and since money market investments will increase by a factor $1 = R$, the portfolio $\vec{H}_1 = (x, y)$ yields at time $t = 1$ a value

$$V_1^h = x(1+r) + y(sZ_1) = \begin{cases} x(1+r) + ysu, & \text{if } Z_1 = u, \\ x(1+r) + ysd, & \text{if } Z_1 = d. \end{cases}$$

On the other hand

$$V_1^h = \mathcal{X} = \Phi(S_1) = \Phi(sZ_1) = \begin{cases} \Phi(su), & \text{if } Z_1 = u, \\ \Phi(sd), & \text{if } Z_1 = d. \end{cases}$$

We equate the right–hand sides separately for $Z_1 = u$ and $Z_1 = d$ and obtain

$$(1+r)x + suy = \Phi(su),$$
$$(1+r)x + sdy = \Phi(sd).$$

This is a linear system of equations with determinant $(1+r)s(d-u)$ which is not zero since $d < u$ and $s > 0$. Thus there is a unique solution (x, y). It is easy to see that

(8.30)
$$x = \frac{1}{1+r} \cdot \frac{u\Phi(sd) - d\Phi(su)}{u-d}, \qquad y = \frac{1}{s} \cdot \frac{\Phi(su) - \Phi(sd)}{u-d}. \quad \blacksquare$$

We have computed a replicating portfolio for an arbitrary simple contract function in a one period binomial market which satisfies $d < 1+r < u$. In other words, such a financial market is complete. [38] Thus we have the following corollary.

Corollary 8.3.

If the one period binomial model is free of arbitrage then it is complete.

PROOF: Immediate from the preceding proposition. ■

Complete markets have the following benefit: We know how to correctly price an arbitrary claims at any point in time if we know how to construct a corresponding hedge, since this price equals the value of that hedge at the given time.

We have seen in Proposition 8.6 on p.166 that discounted stock price is a martingale with respect to risk–neutral measure $\widetilde{P}$. The next proposition states that the same is true for (arbitrage free) pricing of contingent claims.

Proposition 8.9.

In the one period binomial model, the discounted, arbitrage free, price process $D_t \cdot \Pi_t(\mathcal{X})$ of a contigent claim $\mathcal{X}$ is a $\widetilde{P}$–martingale. In particular, we have risk–neutral valuation

(8.31)
$$\Pi_0(\mathcal{X}) = \frac{1}{1+r} \cdot \widetilde{E}[\mathcal{X}].$$

PROOF: Let $\vec{H}$ be a hedging portfolio for $\mathcal{X}$. Since trading only takes place at $t = 0$, $\vec{H}$ is determined by $(x, y) := \vec{H}_1$, i.e., $x = H_1^B$ and $y = H_1^S$. Moreover,

$$\Pi_0(\mathcal{X}) = V_0^H = x \cdot 1 + y \cdot s$$

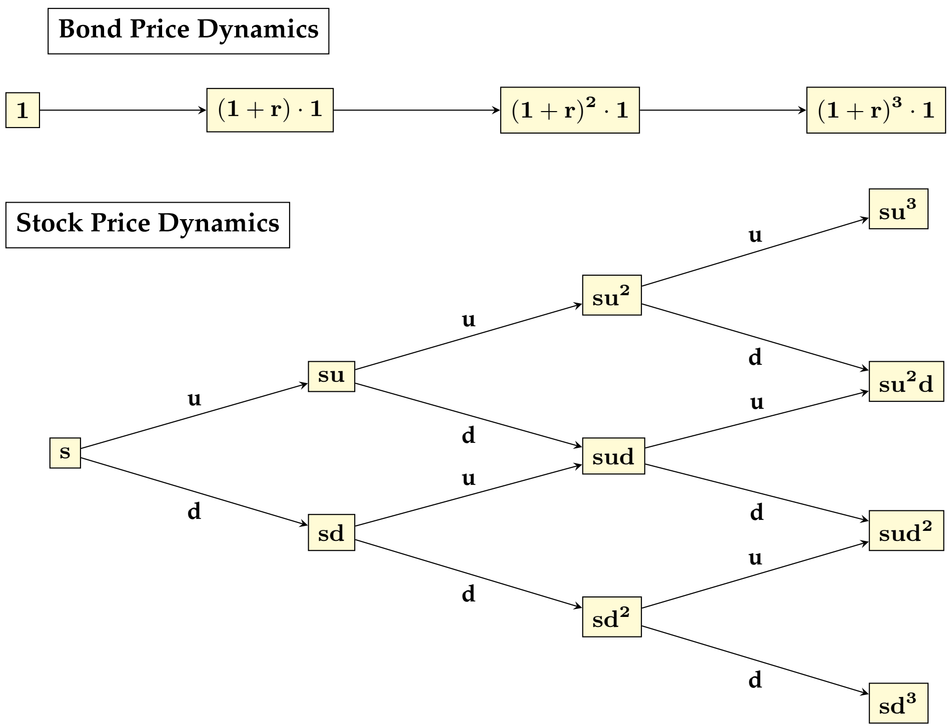
We use the expressions (8.30) for x and y and afterwards the expressions (8.27) for the martingale probabilities $\tilde{p}_u$ and $\tilde{p}_d$. We obtain

$$\Pi_0(\mathcal{X}) = \frac{1}{1+r} \cdot \Big[\frac{(1+r)-d}{u-d}\,\Phi(su) + \frac{u-(1+r)}{u-d}\,\Phi(sd)\Big]$$
$$= \frac{1}{1+r} \cdot \big(\Phi(su)\cdot\tilde{p}_u + \Phi(sd)\cdot\tilde{p}_d\big) = \frac{1}{1+r}\,\widetilde{E}\,[\Phi\circ S_1] = \frac{1}{1+r}\,\widetilde{E}\,[\mathcal{X}]. \quad \blacksquare$$

[38] See Definition 7.9 (Hedging/Replicating Portfolio) on p.151.

8.2 The Multiperiod Model

After having given special attention to the one period model, we now continue with the general binomial asset moded where expiration time T may be greater than one. We recall from Assumption 8.1 for the binomial model that the dynamics that govern the development of the price B_t of the riskless asset (the bond) and the price of the risky asset (the stock) S_t for $t = 0, 1, \dots, T$) are, for $T = 3$, described by the following diagrams.



8.1 (Figure). **Stock price dynamics**

Notation 8.1.

A. We look at a vertical slice of the diagram in Figure 8.1 by fixing a time t_0 and name its $t_0 + 1$ nodes, starting at the bottom, $\mathfrak{N}_{t_0,0}, \mathfrak{N}_{t_0,1}, \dots, \mathfrak{N}_{t_0,t_0}$. This way, the node $\mathfrak{N}_{t_0,k}$ is reached at $t = t_0$ $\Leftrightarrow$ exactly k of the t_0 stock price movements were upward and $t_0 - k$ of them were downward.

Thus $\mathfrak{N}_{t_0,k}$ is the node in the t_0–slice of the diagram with stock price $S_{t_0} = su^k d^{t_0 - k}$.

Clearly, stock price uniquely identifies the t_0–node since $d < u$.

Assuming that the arbitrage free prices for a given simple claim exist, we further write $\Pi\big(\mathfrak{N}_{t_0,k}\big)$ for this arbitrage free price belonging to that node, i.e., associated with $S_{t_0} = su^k d^{t_0 - k}$. We will see in Theorem 8.2 on p.175, that in an arbitrage free market every simple claim has such prices for every node in the tree.

B. Remember for the following that $\vec{H}_t = (H_t^B, H_t^S)$ is the portfolio resulting from the trade that took place at time $t - 1$, and that the bank shares H_t^B must be multiplied with the money market account price $B_{t-1} = (1 + r)^{t-1}$ to obtain the bank account balance at that time. Throughout this chapter on the multiperiod binomial model we write for $t = 1, 2, \dots, T$

- $x_t \;:=\; H_t^B \cdot (1+r)^{t-1}$ = bank money at time $t-1$ after the trade,
- $y_t \;:=\; H_t^S$ = stock shares at time $t-1$ after the trade.

Actually, this formulation is correct only for $t > 1$. For $t = 0$, we should replace the phrase "at time 0 after the trade" with "after the initial setup", since trade of an old portfolio for a new one did not take place at $t - 1 = 0$. □

We recall from Definition 7.7 on p.150 that an arbitrage portfolio is a self–financing portfolio H with the properties

$$V_0^H \;= 0, \qquad P\{V_T^H \geq 0\} \;=\; 1, \qquad P\{V_T^H > 0\} \;>\; 0. \;\square$$

We will see that the condition $d < 1 + r < u$ is both necessary and sufficient for the multiperiod binomial asset model. The proof that this condition is sufficient will be given in Theorem 8.3, but the proof of sufficiency will be done now.

Proposition 8.10.

If the multiperiod model is free of arbitrage, then it satisfies the condition

$$d \;<\; (1+r) \;<\; u. \tag{8.32}$$

PROOF: Similar to the one period case (Proposition 8.5 on p.165).

We prove the contrapositive. We assume that $1 + r \leq d < u$ or $d < u \leq 1 + r$ and construct an arbitrage portfolio. We only handle the case $1 + r \leq d < u$. The proof for $d < u \leq 1 + r$ is similar.

- At $t = 0$ we borrow x dollars from the bank and use it to buy stock. The portfolio value is zero since what we own in stock is what we owe the bank.
- At each trading time $t = 1, 2, 3, \ldots$ we do nothing.
- Since $1 + r \leq d < u$, the following is true for each period: The increase in stock value is at least as high as the interest penalty that is added to the bank loan.
- There is positive probability p_u that $Z_t = u$ for one or more t. In such a case we will not just break even but make a profit since $u > 1 + r$.
- Thus the probability is at least p_u, thus strictly positive, for the following event: When we sell the stock at time T the proceeds will exceed $(1 + r)^T x$, the amount we owe to the bank. We have constructed an arbitrage portfolio. ■

We remind the reader of Assumption 7.1 on p.150 about efficient market behavior.

- The binomial model is free of arbitrage. We thus assume that

$$d \;<\; (1+r) \;<\; u. \;\square$$

We next adapt Definition 8.1 (Martingale Measure) on p.167 to the multiperiod model, remembering from Proposition 8.6 which precedes it, that a martingale measure was characterized by making the discounted stock price a martingale.

Definition 8.2 (Martingale Measure)**.**

We call a probability measure $\widetilde{P}$ that satisfies for all trading times $t = 0, 1, 2, \dots, T-1$ and for all possible values s' of S_t the relation

$$
\begin{aligned}
s' &= \frac{1}{1+r} \cdot \widetilde{E}[S_{t+1} | S_t = s'], \\
\text{i.e., } S_t &= \frac{1}{1+r} \cdot \widetilde{E}[S_{t+1} | S_t],
\end{aligned} \tag{8.33}
$$

a **martingale measure** or also a **risk–neutral measure**. □

Proposition 8.11.

The multiperiod model (which does not admit arbitrage by assumption) possesses a unique martingale measure $\widetilde{P}$. As in the one period model it is defined by the two "martingale probabilities"

$$
\begin{aligned}
\tilde{p}_u &= \frac{(1+r)-d}{u-d}, \\
\tilde{p}_d &= \frac{u-(1+r)}{u-d}.
\end{aligned}
$$

PROOF:
It follows from the definition of $\tilde{p}_u$ and $\tilde{p}_d$ that $u\tilde{p}_u + d\tilde{p}_d = 1 + r$. Thus the discount process is

$$
D_t = \frac{1}{(1+r)^t} = \frac{1}{(u\pi_u + d\pi_d)^t}.
$$

We conclude from Theorem 8.1 on p.164 that the process $D_t S_t$ is a martingale. ■

Proposition 8.12.

Let $\widetilde{P}$ be a probability measure in the multiperiod model. We have the following.

- ***(a)*** *$\widetilde{P}$ is a martingale measure $\Leftrightarrow$ Discounted stock price $D_t S_t$ is a $\widetilde{P}$–martingale.*
- ***(b)*** *In particular, $D_t S_t$ is a martingale with respect to the risk–neutral probability measure $\widetilde{P}$, defined by $\tilde{p}_u u + \tilde{p}_d d = 1 + r$.*

PROOF: of **(a):** First some preparatory work.

S_t is clearly Markov, since either $S_{t+1} = uS_t$, or $S_{t+1} = dS_t$. Thus S_{t+1} does not depend on stock price before t.

It follows from the alternate characterization of the Markov property in Proposition 6.2 on p.124 that if Y is a random variable that only depends on stock price information $S_t, S_{t+1}, S_{t+2}, \dots$, then

$$
\widetilde{E}[Y \mid \mathfrak{F}^S_{t'}] = \widetilde{E}[Y \mid S_{t'}], \quad \text{for all } t' < t.
$$

In particular, since $Y := S_t$ only depends on such information, it follows that

$$(\star\mathbf{1}) \qquad \widetilde{E}[S_t \mid \mathfrak{F}^S_{t'}] = \widetilde{E}[S_t \mid S_{t'}], \quad \text{for all } t' < t\,.$$

$$(\star\mathbf{2}) \qquad \text{Further,} \quad (1+r)D_{t+1} = D_t, \qquad \text{i.e.,} \quad \frac{1}{1+r} = \frac{D_{t+1}}{D_t}\,.$$

$$\text{Thus,} \qquad \frac{1}{1+r}\cdot\widetilde{E}[S_{t+1} \mid S_t] \overset{(\star 1)}{=} \frac{1}{1+r}\cdot\widetilde{E}[S_{t+1} \mid \mathfrak{F}_t]$$

$$(\star\mathbf{3}) \qquad \overset{(\star 2)}{=} \frac{D_{t+1}}{D_t}\cdot\widetilde{E}[S_{t+1} \mid \mathfrak{F}_t] = \frac{1}{D_t}\cdot\widetilde{E}[D_{t+1}S_{t+1} \mid \mathfrak{F}_t]\,.$$

PROOF: of **(a)**, $\Rightarrow$): We show that $\widetilde{E}[D_{t+1}S_{t+1} \mid \mathfrak{F}_t] = S_t$ as follows.

$$\begin{aligned}\widetilde{E}[D_{t+1}S_{t+1} \mid \mathfrak{F}_t] &= D_{t+1}\,\widetilde{E}[S_{t+1} \mid \mathfrak{F}_t] \\ &\overset{(\star\star)}{=} \frac{D_t}{1+r}\cdot\widetilde{E}[S_{t+1} \mid \mathfrak{F}_t] \overset{(\star)}{=} \frac{D_t}{1+r}\cdot\widetilde{E}[S_{t+1} \mid S_t] \overset{(8.33)}{=} S_t\,.\end{aligned}$$

PROOF: of **(a)**, $\Leftarrow$): Since D_t is an $\mathfrak{F}_t$–martingale for $\widetilde{P}$,

$$(\star 4) \qquad \widetilde{E}[D_{t+1}S_{t+1} \mid \mathfrak{F}_t] = S_t\,.$$

Thus,

$$\frac{1}{1+r}\cdot\widetilde{E}[S_{t+1} \mid S_t] \overset{(\star 3)}{=} \frac{1}{D_t}\cdot\widetilde{E}[D_{t+1}S_{t+1} \mid \mathfrak{F}_t] \overset{(\star 4)}{=} \frac{1}{D_t}\cdot(D_tS_t) = S_t\,.$$

PROOF: of **(b):** This follows from **(a)** and Proposition 8.11. ■

Proposition 8.13.

In the multiperiod model, assume that

(a) *$\widetilde{P}$ is a martingale measure,*

(b) *$\vec{H}_t = (H^B_t, H^S_t)$ is a self–financing portfolio.*

Then discounted portfolio value $D_tV^{\vec{H}}_t$ is a $\mathfrak{F}^S_t$–martingale with respect to $\widetilde{P}$.

PROOF: $\vec{H}_t$ is self–financing, thus we have the budget equation

$$(\mathbf{A}) \qquad V_t = x_t(1+r) + y_tS_t = x_{t+1} + y_{t+1}S_t\,.$$

We also know that discounted stock price is a martingale, thus

$$(1+r)^{-1}\cdot\widetilde{E}[S_{t+1} \mid \mathfrak{F}_t] = S_t\,.$$

We recall that x_{t+1} and y_{t+1} were established during the trade at time t and thus are $\mathfrak{F}_t$–measurable. We write as usual $D_1 = (1+r)^{-1}$ and obtain

$$\begin{aligned}\widetilde{E}\left[D_1V_{t+1} \mid \mathfrak{F}_t\right] &= \widetilde{E}\left[D_1x_{t+1}(1+r) + D_1y_{t+1}S_{t+1} \mid \mathfrak{F}_t\right] \\ &= \widetilde{E}\left[x_{t+1} \mid \mathfrak{F}_t\right] + \widetilde{E}\left[y_{t+1}D_1S_{t+1} \mid \mathfrak{F}_t\right] \\ (\mathbf{B}) \qquad &= x_{t+1} + y_{t+1}\cdot\widetilde{E}\left[D_1S_{t+1} \mid \mathfrak{F}_t\right] \\ (\mathbf{C}) \qquad &= x_{t+1} + y_{t+1}\cdot S_t = V_t.\end{aligned}$$

Here we obtained **(B)** by moving the $\mathfrak{F}_t$–measurable variables x_{t+1} and x_{t+1} out of the conditional expectation. The first equation of **(C)** follows from the fact that D_tS_t is a $\widetilde{P}$–martingale, and the second equation of **(C)** follows from the budget equation **(A)**. Thus

$$\widetilde{E}\left[D_{t+1}V_{t+1} \mid \mathfrak{F}_t\right] \;=\; D_t\,\widetilde{E}\left[D_1V_{t+1} \mid \mathfrak{F}_t\right] \;=\; D_tV_t \quad \blacksquare$$

The fact that the discounted portfolio value of a self–financing portfolio is a $\widetilde{P}$–martingale (thus, by the pricing principle, the discounted price $\Pi_t(\mathcal{X})$ of a reachable claim $\mathcal{X}$) also is a $\widetilde{P}$–martingale), will be employed in the next example.

Example 8.1. Consider a market which follows the multiperiod binomial model with the following parameters.

- Time of expiry is $T = 4$.
- The interest rate is $R = 0.5$ (per unit of time). That's not very realistic, but it makes this example computationally simple.
- We denote the "true" probability with P, and the martingale probability with $\widetilde{P}$. The corresponding expectations are E^P and $\widetilde{E}$. Note that nothing is said about $p_u, p_d, \tilde{p}_u, \tilde{p}_d$.
- Assume that a hedge portfolio must be created for a simple claim with contract value $\Phi(S_4)$

(a) If it is known today that $E^P[\Phi(S_4)] = \$240$, is $V_0 = \$50$ possible as the setup value of this hedge?

(b) If it is known today that $\widetilde{E}[\Phi(S_4)] = \$180$, is $V_0 = \$50$ possible as the setup value of this hedge?

We answer the questions above as follows.

(a) Given the real world probabilities, everything is possible. That's about all that can be said with the information at hand.

(b) The situation is different under risk–neutral probability measure $\widetilde{P}$, even if we do not know the values of $\tilde{p}_u$ and $\tilde{p}_d$.

Since D_tV_t is a $\widetilde{P}$–martingale, the expected value is constant for all t, thus,

$\widetilde{E}[D_4V_4] = \widetilde{E}[D_0V_0] = \widetilde{E}[V_0]$.

Since $D_t = (1+r)^{-t}$ and V_0 are deterministic and $B_t = (1+r)^{-t}$, and $V_4 = \Phi(S_4)$ by the pricing principle, we obtain $V_0 = \widetilde{E}[V_0] = (1+r)^{-4}\widetilde{E}[V_4](1+r)^{-4}\widetilde{E}[\Phi(S_4)] = 180 \cdot 1.5^{-4}$.

Since $1.5^4 = 2.25^2$ and $2 \leq 2.25 \leq 3$, we obtain $180/9 \leq V_0 \leq 180/4$, i.e., $20 \leq V_0 \leq 45$.

Thus, $\$50$ is too big a value for the value V_0 of the hedge at time 0. $\square$

Example 8.2. We have a financial market with one bond and one stock which follows the one period model. We assume the interest rate is $R = 0$, so the bond price is $B_0 = B_1 = 1$. We also assume that

$$S_0 \;=\; s \;=\; 100; \quad S_1 \;=\; \begin{cases} \frac{5}{4} \cdot S_0 = 125 & \text{with probability } 0.8, \\ \frac{3}{4} \cdot S_0 = 75 & \text{with probability } 0.2. \end{cases}$$

(1) How do you price a European call at a strike price of 115 at $t = 0$?

(2) If x = the money in the bank and y = number of shares in the stock in the hedge you establish for this contract, what are x and y at $t = 0$?

This problem is solved as follows. The risk–neutral probabilities are $\tilde{p}_u = \tilde{p}_d = \frac{1}{2}$, since

$$1 + r \;=\; 1 \;=\; \frac{1}{2} \cdot \frac{5}{4} + \frac{1}{2} \cdot \frac{3}{4}.$$

Contract values are $\Phi(su) \;=\; 125 - 115 \;=\; 10$ and $\Phi(sd) \;=\; 0$.
Thus, the options price at time zero is

$$\Pi_0(\mathcal{X}) \;=\widetilde{E}[\mathcal{X}] \;=\; \tilde{p}_d \cdot \Phi(sd) + \tilde{p}_u \cdot \Phi(su) \;=\; \frac{1}{2} \cdot 10 \;=\; \frac{10}{2} \;=\; 5.$$

The quantities involved for setting up the hedge are (see Proposition 8.8 on p.168)

$$x \;=\; \frac{1}{1+r} \cdot \frac{u\Phi(sd) - d\Phi(su)}{u-d} \;=\; 1 \cdot \frac{1.25 \cdot 0 \;-\; 0.75 \cdot 10}{0.5} \;=\; -15,$$
$$y \;=\; \frac{1}{s} \cdot \frac{\Phi(su) - \Phi(sd)}{u-d} \;=\; \frac{1}{100} \cdot \frac{10-0}{0.5} \;=\; \frac{20}{100} \;=\; 0.2.$$

Thus the hedging portfolio consists of 0.2 shares of the stock and a short position (loan) of 15 bond units (worth \$15.00 at the time of setup $t = 0$).

For a sanity check, we validate that in fact $V_0^H \;=\; 5 \;=\; \Pi_0(\mathcal{X})$, as must be true according to the definition of a hedge for the claim.

$V_0^H \;=\; x + ys \;=\; -15 + 0.2 \cdot 100 \;=\; 5.$ $\square$

In the one period model absence of arbitrage was sufficient to yield completeness of the market, i.e., every claim can be hedged. In the multiperiod model we can still show that every simple claim, i.e., a claim for which the payoff $\mathcal{X}$ is a function $\Phi(S_T)$ of stock price at time T, can be hedged.

Theorem 8.2. *Let $\mathcal{X}$ be a simple claim with expiration date T and contract function $\Phi(x)$, i.e., $\mathcal{X} = \Phi(S_T)$. Let $\Pi_t(\mathcal{X})$ denote the arbitrage free price of that option at time $t \leq T$.*

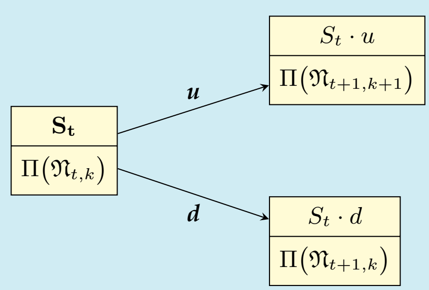
(1) *The discounted option price $\frac{1}{(1+r)^t}\Pi_t(\mathcal{X})$ is a $\widetilde{P}$–$\mathfrak{F}_t^S$–martingale.*

(2) *The option price is computed at time $0 \leq t \leq T$ for a stock price of $S_t(\omega) = su^k d^{t-k}$, attained by k upward moves and $t-k$ downward moves, as*

(8.34) $$\Pi_t(\mathcal{X}) \;=\; \frac{1}{(1+r)^{T-t}}\, \widetilde{E}\big[\Phi(S_T) \mid S_t = su^k d^{t-k}\big].$$

(3) *$\mathcal{X}$ can be hedged. The portfolio quantities H_{t+1}^B and H_{t+1}^S are $H_{t+1}^B = (1+r)^{-t} x_{t+1}$ and $H_{t+1}^S = y_{t+1}$, where x_{t+1}, y_{t+1} for the node $\mathfrak{N}_{t,k}$ (remember: $\vec{H}_t$ = purchases at time $t-1$!) in the tree excerpt shown below are as follows.*

(8.35)
$$x_{t+1} \;=\; \frac{1}{1+r} \cdot \frac{u\Pi\big(\mathfrak{N}_{t+1,k}\big) - d\Pi\big(\mathfrak{N}_{t+1,k+1}\big)}{u-d},$$
$$y_{t+1} \;=\; \frac{1}{s} \cdot \frac{\Pi\big(\mathfrak{N}_{t+1,k+1}\big) - \Pi\big(\mathfrak{N}_{t+1,k}\big)}{u-d}.$$



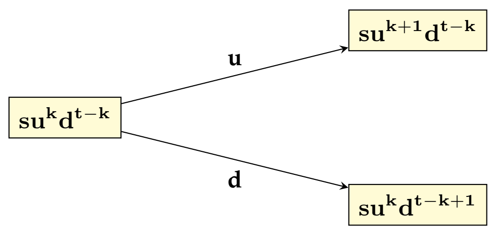
PROOF: (outline): ⋆ For the following all indices, including $t, T, T', \dots$, are assumed to be trading times in the binomial model, hence non–negative integers. Also recall the notation we introduced for the nodes of the binomial tree displayed in Figure 8.1 (Stock price dynamics) on p.170. Fix a time $0 \le t < T$ and assume that the arbitrage free claim price are known for all nodes at time $t+1$. We can consider those prices as the contract function $\Phi^{(t+1)}$ of a new contingent claim

$$\mathcal{X}^{(t+1)} = \Phi^{(t+1)}(s'), \quad \text{where} \quad s' = sd^{t+1},\, sud^t,\, su^2d^{t-1}, \dots, su^td,\, su^{t+1}$$

runs through the stock prices that can be attained at time $t+1$.

Fix $0 \le k \le t$ and consider the node $\mathfrak{N}_{t,k}$ in the tree. That node was reached by a combination of k upward movements and $t-k$ downward movements in stock price. The two nodes at time $t+1$ that can be reached from $\mathfrak{N}_{t,k}$ by either an upward move or a downward move in stock price are $\mathfrak{N}_{t+1,k+1}$ and $\mathfrak{N}_{t+1,k}$. In particular, if $t = T-1$, we obtain $\mathcal{X}^{(t+1)} = \mathcal{X}$ and $\Phi^{(t+1)}(s') = \Phi(s')$ for each $s' = sd^T,\, sud^{T-1}, \dots, su^T$.

We now condition on $S_t = su^kd^{t-k}$. Since such conditioning makes stock price constant at t, we can apply our findings from the one period model to the tree which consists of the nodes $\mathfrak{N}_{t,k}, \mathfrak{N}_{t+1,k+1}$ and $\mathfrak{N}_{t+1,k}$.



With the symbols introduced in Notations 8.1 on p.170 we have

$$\Pi\big(\mathfrak{N}_{t+1,k+1}\big) = \Phi^{(t+1)}(su^{k+1}d^{t-k}), \quad \text{and} \quad \Pi\big(\mathfrak{N}_{t+1,k}\big) = \Phi^{(t+1)}(su^kd^{t-k+1}).$$

We apply the risk–neutral valuation formula (8.31) of Proposition 8.9 on p.169 to this one–period tree with the new contract function $\Phi^{(t+1)}$. We must adjust the notation as follows:

- Times 0 and 1 in Prop.8.9 correspond to times t and $t+1$ here.
- Stockprice $S_0 = s$ in Prop.8.9 corresponds to $S_t = su^kd^{t-k}$.
- Stockprices $S_1 = su$ and $S_1 = sd$ in Prop.8.9 correspond to $S_{t+1} = su^{k+1}d^{t-k}$ and $S_{t+1} = su^kd^{t-k-1}$.
- Option values $\Pi_0(\mathcal{X})$ at time 0 and $\mathcal{X}$ at time 1 in Prop.8.9 correspond to $\Pi\big(\mathfrak{N}_{t,k}\big)$ at time t, and to $\Pi\big(\mathfrak{N}_{t+1,k+1}\big) = \Phi^{(t+1)}(su^{k+1}d^{t-k})$ and $\Pi\big(\mathfrak{N}_{t+1,k}\big) = \Phi^{(t+1)}(su^kd^{t-k-1})$ at time $t+1$.

Thus we obtain the arbitrage free price of $\mathcal{X}$ for the node $\mathfrak{N}_{t,k}$, which we denote by $\Pi\big(\mathfrak{N}_{t,k}\big)$, as

$$\begin{aligned}
\Pi\big(\mathfrak{N}_{t,k}\big) &= \frac{1}{1+r} \cdot \widetilde{E}[\mathcal{X}^{(t+1)}] \\
&= \frac{1}{1+r}\big(\tilde{p}_u \cdot \Phi^{(t+1)}(su^{k+1}d^{t-k}) + \tilde{p}_d \cdot \Phi^{(t+1)}(su^kd^{t-k+1})\big) \qquad (8.36)\\
&= \frac{1}{1+r}\big(\tilde{p}_u \cdot \Pi\big(\mathfrak{N}_{t+1,k+1}\big) + \tilde{p}_d \cdot \Pi\big(\mathfrak{N}_{t+1,k}\big)\big).
\end{aligned}$$

Since $\widetilde{E}[\mathcal{X}^{(t+1)}]$ is just a real number, $\Pi_t(\mathcal{X})(\omega)$ is constant for all ω such that $S_t(\omega)$ belongs to $\mathfrak{N}_{t,k}$, i.e., for all ω such that $S_t(\omega) = su^kd^{t-k}$. In other words, $\Pi\big(\mathfrak{N}_{t,k}\big)$ is a function of stock price at time t. Thus there is a function $\Phi^{(t)}(x)$ of $x > 0$ such that

$$\Pi\big(\mathfrak{N}_{t,k}\big) = \Phi^{(t)}(S_t)\,.$$

We have managed to express the arbitrage free option price at t as a simple contract at time t.

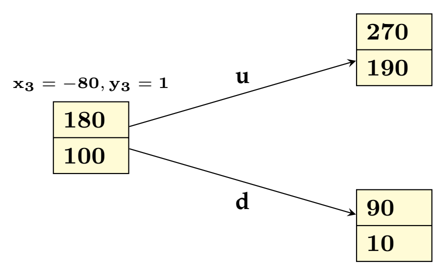
The above procedure tells us how to recursively compute today's ($t = 0$) arbitrage free option price $\Pi_0(\mathcal{X})$ from the contract values $\Phi(x)$ at time T:

We compute $\Phi^{(T-1)}(x)$ from $\Phi^{(T)}(x) = \Phi(x)$, then $\Phi^{(T-2)}(x)$ from $\Phi^{(T-1)}(x)$, ..., then $\Phi^{(1)}(x)$ from $\Phi^{(2)}(x)$, then $\Phi^{(0)}(x)$ from $\Phi^{(1)}(x)$. We now obtain from those contract functions $\Phi(t)(x)$ the corresponding options prices $\Pi_t(\mathcal{X}) = \Phi^{(t)}(S_t)$, in particular, $\Pi_0(\mathcal{X})$.

Working our way backward in time also is how we find the arbitrage free option price at time zero from its contract values at expiration time in practice. See Example 8.3 which follows this proof. But a correct proof is done best by using strong induction in the forward direction.

This proof is very complicated and omitted. Be sure to carefully study instead Example 8.3 on p.177 which follows the "proof" given above. It shows you how to apply this theorem in practical computations! ■

In the following we will draw trees which look like the one to the right. (We did so already in the proof of Theorem 8.2.) The nodes have an upper half which denotes stock price and a lower half which denotes the arbitrage free price of a claim. If there is a label above such a node, then it denotes the quantities x_t and y_t of the corresponding replicating portfolio, evaluated at that node. Note that $u = 1.5$ and $d = 0.5$ since the stock price of 180 increases to 270 and decreases to 90.



The following example is taken from chapter 2 of [7] Björk, Thomas: Arbitrage Theory in Continuous Time.

Example 8.3. We set $T = 3, s := S_0 = 80, u = 1.5, d = 0.5, p_u = 0.6, p_d = 0.4$ and $R = 0$.

These numbers have been chosen to make computations as simple as possible. Since there is no interest, $1 = 1 + r$ is the midpoint between $u = 1.5$ and $d = 0.5$, thus $\tilde{p}_u = \tilde{p}_d = 0.5$.

Figure 8.1 shows the binomial tree for this example. There are no values in the lower halfs of the nodes for the claims prices since we did not yet decide on a claim.

The claim we want to price is a European call with a strike price of $K = \$80.00$, and an expiration date of $T = 3$.

This is a simple claim $\mathcal{X} = \Phi(S_T)$ with contract function $\Phi(s) = (s - 80)^+ = \max(s - 80, 0)$. We immediately compute $\Pi_3(\mathcal{X})$ for the stock prices S_3 as follows.

$$\Phi(270) = (270 - 80)^+ = 190; \qquad \Phi(90) = (90 - 80)^+ = 10,$$
$$\Phi(30) = (30 - 80)^+ = 0, \qquad \Phi(10) = (10 - 80)^+ = 0,$$

Figure 8.2 shows the updated tree.

We know from formula (8.36) on p.176 how to compute a claims price from those of the two child nodes to the right. With the notations introduced in Notations 8.1 on p.170,

$$\Pi\big(\mathfrak{N}_{t,k}\big) = \frac{1}{1+r}\big(\tilde{p}_u \cdot \Pi(\mathfrak{N}_{t+1,k+1}) + \tilde{p}_d \cdot \Pi(\mathfrak{N}_{t+1,k})\big).$$

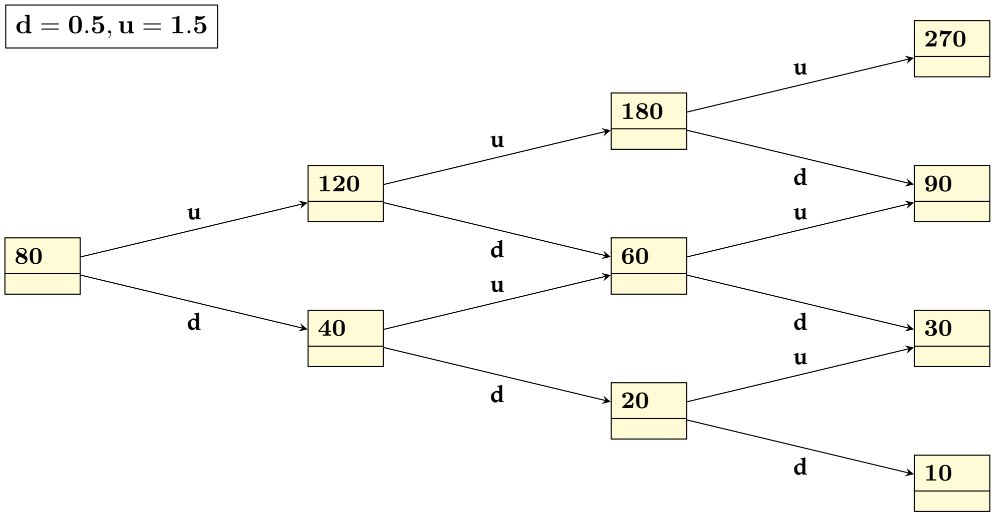


Figure 8.1: Stock prices.

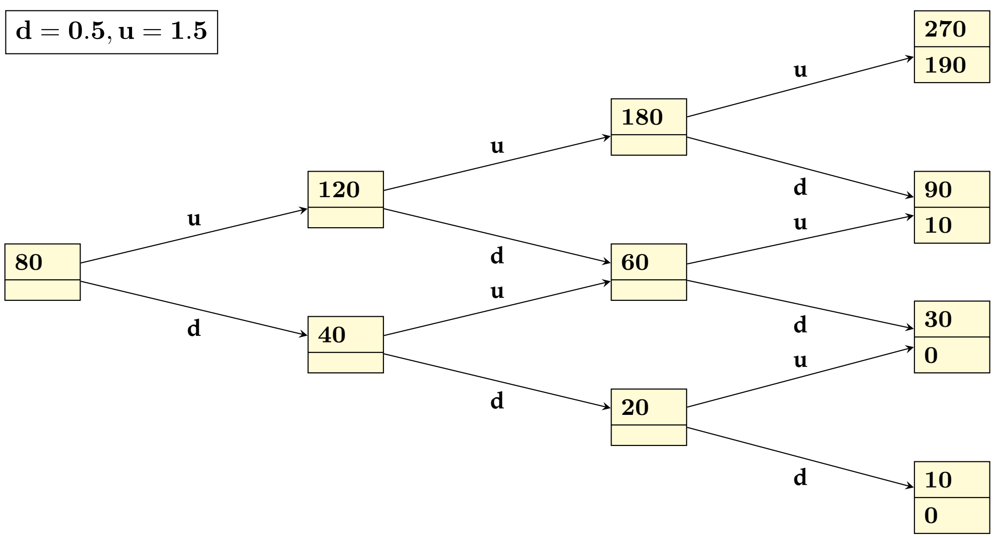


Figure 8.2: Stock prices and contract function values.

For example, for node $\mathfrak{N}_{2,2}$ we obtain $S_2 = 180$, $\Pi(\mathfrak{N}_{3,3}) = 190$, $\Pi(\mathfrak{N}_{3,2}) = 10$. Thus

$$\Pi\big(\mathfrak{N}_{2,2}\big) \;=\; \frac{1}{1+0}\big(0.5 \cdot 190\big) \;+\; 0.5 \cdot 10\big) \;=\; 100.$$

Likewise, for node $\mathfrak{N}_{2,1}$ we obtain $S_1 = 60$, $\Pi(\mathfrak{N}_{3,2}) = 10$, $\Pi(\mathfrak{N}_{3,1}) = 0$. Thus

$$\Pi\big(\mathfrak{N}_{2,1}\big) \;=\; \frac{1}{1+0}\big(0.5 \cdot 10\big) \;+\; 0.5 \cdot 0\big) \;=\; 5.$$

We just computed the two options prices for the descendants of node $\mathfrak{N}_{1,1}$, the one with stock price $S_1 = 120$. Its associated price for the European call is

$$\Pi\big(\mathfrak{N}_{1,1}\big) \;=\; \frac{1}{1+0}\big(0.5 \cdot 100\big) \;+\; 0.5 \cdot 0.5\big) \;=\; 52.5.$$

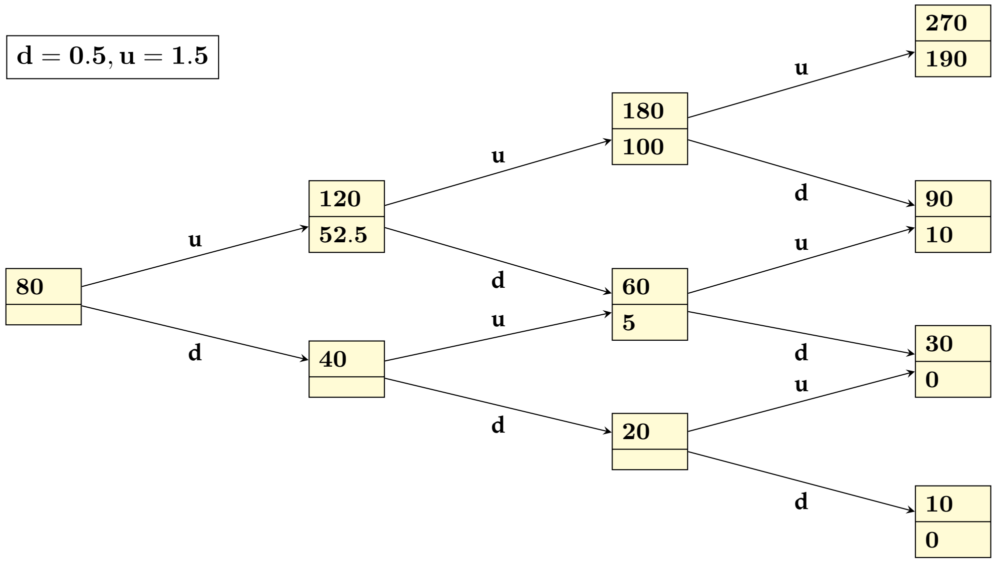


Figure 8.3: Stock prices and contract function values.

Figure 8.3 shows the tree with those additional values.
We compute the arbitrage free option prices for the remaining three nodes in this order:

$$\Pi(\mathfrak{N}_{2,0}),\ \Pi(\mathfrak{N}_{1,0}),\ \Pi(\mathfrak{N}_{0,0}).$$

The completed tree is shown in Figure 8.4.

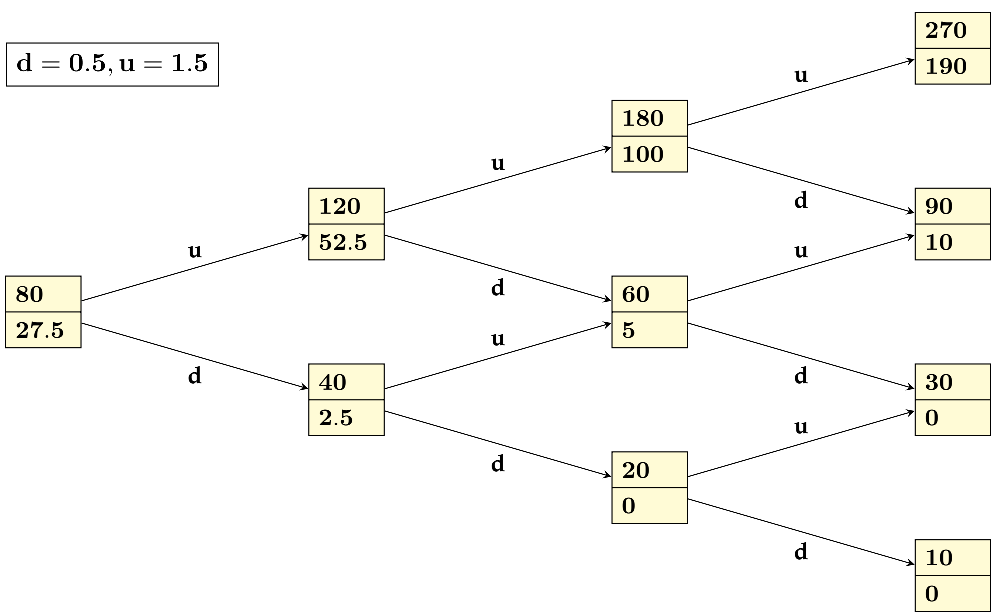


Figure 8.4: Completed tree with all option prices.

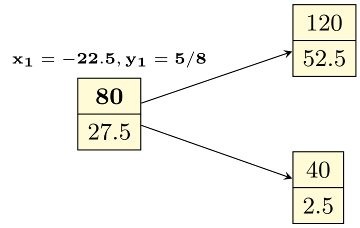
The result of all the above: We have managed to compute the arbitrage free prices of the simple claim with contract function $\mathcal{X} = \Phi(S_3) = (S_3 - K)^+$ for all possible stock prices $S_t,\ t = 0, 1, 2, 3$. In particular we found that the correct price for the option at time zero is 27.5.

We are not finished yet. Next we compute the quantities x_t and y_t of the replication portfolio for this claim.

We start at $t = 0$, and since we want to reproduce the claim $(52.5, 2.5)$ at $t = 1$, we can use formulas (8.35) of Theorem 8.2 on p.175 and obtain $x_1 = -22.5, y_1 = \frac{5}{8}$ since

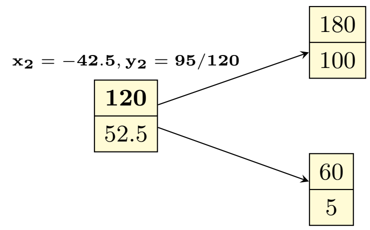
$x_1 = \frac{1}{1+0} \cdot \frac{1.5 \cdot 2.5 - 0.5 \cdot 52.5}{1.5 - 0.5} = \frac{3 \cdot 5 - 1 \cdot 105}{4} = -\frac{90}{4} = -22.5,$

$y_1 = \frac{1}{s} \cdot \frac{\Phi(u) - \Phi(d)}{u - d} = \frac{1}{80} \cdot \frac{52.5 - 2.5}{1.5 - 0.5} = \frac{50}{80} = \frac{5}{8}.$

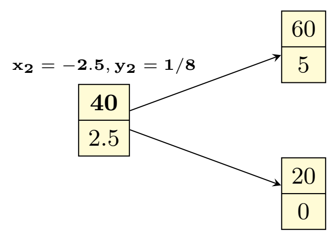


You are encouraged to verify that the cost of this portfolio is indeed 27.5.

If an upward move takes place and $S_1 = 120$ then the value of our hedging portfolio at time 1 is computed from $x_1 = -22.5$ and $y_1 = \frac{5}{8}$ as $-22.5 \cdot (1 + 0) + \frac{5}{8} \cdot 120 = 52.5$.
To reproduce the claim claim $(100, 5)$ at $t = 2$ we again use the formulas (8.35) and obtain $x_2 = -42.5, y_2 = \frac{95}{120}$.
Again you should check that the cost of those holdings, valued at a stock price of $S_1 = 120$,
equals the value 52.5 of the previous holdings x_1 and y_1.



If instead of an upward move a downward move had taken place and $S_1 = 40$ then the value of our hedging portfolio at time 1 is computed from the same holdings $x_1 = -22.5$ and $y_1 = \frac{5}{8}$ as $-22.5 \cdot (1 + 0) + \frac{5}{8} \cdot 40 = 2.5$.
To reproduce the claim claim $(100, 5)$ at $t = 2$ we again use the formulas (8.35) and obtain $x_2 = -2.5, y_2 = 1.8$.
Again you should check that the cost of those holdings, valued at a stock price of $S_1 = 120$,
equals the value 52.5 of the holdings x_1 and y_1 established at time zero.



We can continue in this manner with the nodes at time $t = 2$ and afterwards at expiration time $T = 3$ and in this way compute the hedging portfolio holdings at each node of the tree. The resulting tree is shown in figure 8.5.

This concludes the example. □

Remark 8.4. The following is a cookbook recipe for computing the prices of a simple claim using the risk–neutral validation method.

Step 1: Compute the martingale probabilities!
Note that the martingale probabilities $\tilde{p}_u, \tilde{p}_d$ are constant for the entire tree since they only depend on u, d, and R. In this example they are

$$\tilde{p}_u = \frac{(1+r) - d}{u - d} = \frac{\frac{3}{2} - 1}{\frac{3}{2}, - \frac{1}{2}} = \frac{\frac{1}{2}}{1} = \frac{1}{2}, \qquad \tilde{p}_d = 1 - \tilde{p}_u = \frac{1}{2}$$

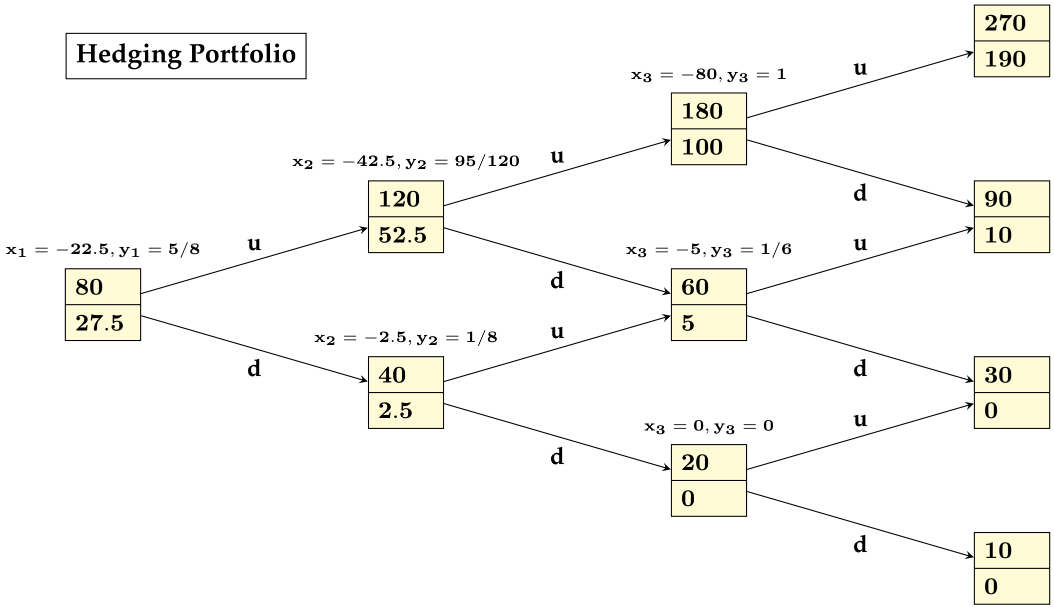


Figure 8.5: Hedging portfolio holdings.

Step 2: Use the risk–neutral valuation formula from the one–period model to compute for each of the three $t = 2$ nodes in the tree its option price $\Pi(2; \mathcal{X})$ from the option prices $\Pi(3; \mathcal{X})$ of the two $t = 3$ nodes that can be reached from this $t = 2$ node. We then compute

$$\Pi(2; \mathcal{X}) \;=\; \frac{1}{1+r}\Big[\tilde{p}_u \cdot \Pi(3; \mathcal{X}) \text{ of upward node } + \tilde{p}_d \cdot \Pi(3; \mathcal{X}) \text{ of downward node}\Big].$$

This method can be employed for any binomial tree, for arbitrarily many periods.

Step t-1: Let $\boldsymbol{N}$ be a $t-1$ node in the binomial tree. We denote the reachable node to the upper left by $\boldsymbol{N}_u$ and the reachable node to the lower left by $\boldsymbol{N}_d$. We write $\Pi_{t-1}(\boldsymbol{N})$ for the option price of node $\boldsymbol{N}$ and we write $\Pi_t(\boldsymbol{N}_u)$ and $\Pi_t(\boldsymbol{N}_d)$ for the option prices of $\boldsymbol{N}_u$ and $\boldsymbol{N}_d$.
If $\Pi_t(\boldsymbol{N}_u)$ and $\Pi_t(\boldsymbol{N}_d)$ have already been computed then we use the risk–neutral valuation formula from the one–period model to compute $\Pi_{t-1}(\boldsymbol{N})$:

$$\Pi_{t-1}(\boldsymbol{N}) \;=\; \frac{1}{1+r}\Big[\tilde{p}_u \cdot \Pi_t(\boldsymbol{N}_u) \;+\; \tilde{p}_d \cdot \Pi_t(\boldsymbol{N}_u)\Big]. \quad \square$$

We mention again that this entire chapter 8 (Financial Models - Part 1) closely follows the book [7] Björk, Thomas: Arbitrage Theory in Continuous Time.

Notation 8.2. We will write

$$V\big(\mathfrak{N}_{t,k}\big) \;\; (0 \leq t \leq T),$$

for the value process of the replicating portfolio strategy, determined in Theorem 8.2 on p.175 by the formulas (8.35), when computed for the node $\mathfrak{N}_{t,k}$ of the binomial tree. $\square$

Proposition 8.14. *Given are a simple claim $\mathcal{X} = \Phi(S_T)$, its associated pricing process $\Pi_t(\mathcal{X})$, and its hedging portfolio $\vec{H}_t$ with value process V_t^H. If we replace the symbols $\Pi_t(\mathcal{X})$ and $V_t^{\vec{H}}$ with their tree node equivalents, $\Pi(\mathfrak{N}_{t,k})$ and $V(\mathfrak{N}_{t,k})$, we have the following.*

The replicating portfolio is determined by the recursive formulas

$$\begin{aligned} V(\mathfrak{N}_{t,k}) &= \frac{1}{1+r}\big(\tilde{p}_u V(\mathfrak{N}_{t+1,k+1}) + \tilde{p}_d V(\mathfrak{N}_{t+1,k})\big),\\ V(\mathfrak{N}_{T,k}) &= \Phi(su^k d^{T-k}). \end{aligned} \tag{8.37}$$

Here, $\tilde{p}_u$ and $\tilde{p}_d$ are the martingale probabilities from Proposition 8.11 on p.172, given by

$$\tilde{p}_u = \frac{(1+r)-d}{u-d}, \qquad \tilde{p}_d = \frac{u-(1+r)}{u-d}. \tag{8.38}$$

Further, the hedging portfolio quantities x_{t+1}, y_{t+1} for the node $\mathfrak{N}_{t,k}$ are

$$\begin{aligned} x_{t+1} &= \frac{1}{1+r}\cdot\frac{uV(\mathfrak{N}_{t+1,k}) - dV(\mathfrak{N}_{t+1,k+1})}{u-d},\\ y_{t+1} &= \frac{1}{s}\cdot\frac{V(\mathfrak{N}_{t+1,k+1}) - V(\mathfrak{N}_{t+1,k})}{u-d}, \end{aligned}$$

and the arbitrage free option prices are given by $\Pi(\mathfrak{N}_{t,k}) = V(\mathfrak{N}_{t,k})$, for all trading times $0 \le t \le T$ and number of upward moves $0 \le k \le t$. In particular, the arbitrage free price of the claim at $t = 0$ is given by $V(\mathfrak{N}_{0,0}) = x_1 + y_1 S_0$.

PROOF: This is just a rehash of Proposition 8.11 and Theorem 8.2 together with the pricing principle, Theorem 7.1 on p.151, which states that

$$V(\mathfrak{N}_{t,k}) = \Pi(\mathfrak{N}_{t,k}) \text{ for all nodes } \mathfrak{N}_{t,k} \text{ in the binomial tree.} \quad \blacksquare$$

Considering that stock price S_t develops according to an iid sequence of Bernoulli variables Z_t (with success probability p_u under the "real world" measure P and success probability $\tilde{p}_u$ under the risk–neutral measure (martingale measure) $\widetilde{P}$ it should not come as a surprise that the options price process $\Pi_T(\mathcal{X})$ for a simple claim $\mathcal{X}$, and thus also the identical portfolio value process V_t^H for a replicating portfolio $\vec{H}_t$, have a close connection with the binomial distribution.

Proposition 8.15 (Arbitrage free price at time zero)**.**

The arbitrage free price at $t = 0$ of a simple claim $\mathcal{X}$ at time T is

$$\Pi_0(\mathcal{X}) = \frac{1}{(1+r)^T}\cdot \widetilde{E}[\mathcal{X}], \tag{8.39}$$

where $\widetilde{P}$ denotes the martingale measure. Further,

$$\Pi_0(\mathcal{X}) = \frac{1}{(1+r)^T}\cdot\sum_{k=0}^{T}\binom{T}{k}\tilde{p}_u^k\tilde{p}_d^{T-k}\Phi(su^k d^{T-k}). \tag{8.40}$$

PROOF: According to Theorem 8.2 on p.175, discounted option price $(1+r)^{-t}\Pi_t(\mathcal{X})$ is a $\widetilde{P}$–martingale and thus has constant expectations in t. Hence, since $\Pi_T(\mathcal{X}) = \mathcal{X}$,

$$\frac{1}{(1+r)^T}\,\widetilde{E}[\mathcal{X}] \;=\; \widetilde{E}\left[\frac{1}{(1+r)^T}\,\Pi_T(\mathcal{X})\right] \;=\; \widetilde{E}[\Pi_0(\mathcal{X})]\,.$$

This proves (8.39). Formula (8.40) is immediate from Corollary 8.2 (Expectation of a simple claim in the binomial tree model) on p.165 ■

We end this section by proving absence of arbitrage.

Theorem 8.3.

The binomial asset model is free of arbitrage $\Leftrightarrow$ $d < 1+r < u$.

PROOF: We already proved the "⇒" direction in Proposition 8.10 (see p.171).

For the other direction, we assume that $d < 1+r < u$ and that $\vec{H}_t$ is a self–financing portfolio such that $P\{V_0^{\vec{H}} \geq 0\} = 1$ and $P\{V_T^{\vec{H}} > 0\} > 0$. We now show that $P\{V_0^{\vec{H}} > 0\} > 0$.

It follows from Proposition 8.13 on p.173, that $D_t V_t^{\vec{H}}$ is a $\widetilde{P}$–martingale for the martingale measure $\widetilde{P}$ determined by $\tilde{p}_u$ and $\tilde{p}_d$ such that $u\tilde{p}_u + d\tilde{p}_d = 1+r$ and $\tilde{p}_u + \tilde{p}_d = 1$. We recall that P and $\widetilde{P}$ are equivalent measures, i.e., the P-Null sets coincide with the $\widetilde{P}$-Null sets, thus $P(A) > 0 \Leftrightarrow \widetilde{P}(A) > 0$ for any event A.

It follows from $P\{V_T^{\vec{H}} > 0\} > 0$ that $\widetilde{P}\{V_T^{\vec{H}} > 0\} > 0$, hence, $\widetilde{E}\left[V_T^{\vec{H}}\right] > 0$. Since the $\widetilde{P}$–martingale $D_t V_t^{\vec{H}}$ has constant expectations in t,

$$\widetilde{E}\left[V_0^{\vec{H}}\right] \;=\; \widetilde{E}\left[D_T V_T^{\vec{H}}\right] > 0\,.$$

It follows from $V_0^{\vec{H}} \geq 0$ $\widetilde{P}$–a.s. that $\widetilde{P}\{V_0^{\vec{H}} > 0\} > 0$. Thus $P\{V_0^{\vec{H}} > 0\} > 0$, hence, $\vec{H}$ is not an arbitrage portfolio. Since $\vec{H}$ is an arbitrary self–financing portfolio such that $P\{V_0^{\vec{H}} \geq 0\} = 1$ and $P\{V_T^{\vec{H}} > 0\} > 0$, we have shown that arbitrage portfolio do not exist. ■

8.3 Exercises for Ch.8

Exercise 8.1. Prove the following part of Proposition 8.5 on p.165 of this document: If

$$d < (1+r) < u. \;\square$$

then the one period binomial asset model is free of arbitrage.

Hint: Show that

$$V_1^h \;=\; ys\,(u-(1+r)),\ \text{if } Z=u, \quad ys(d-(1+r)),\ \text{if } Z=d,$$

and examine this separately for $y > 0$ and $y < 0$. □

Exercise 8.2. We asserted that the probability measure $\widetilde{P}$ defined by (8.24) on p.166 is equivalent to P on $\sigma(S_0, S_1)$. Prove it. □

8.4 Blank Page after Ch.8

This page is intentionally left blank!

9 One dimensional Stochastic Calculus

9.1 Riemann–Stieltjes Integrals

In stochastic finance one would like to work with "stochastic integrals" where one integrates a process $Z = Z_t$ not simply with respect to time t, but rather with respect to the "density" $W'_t = \frac{dW_t}{dt}$ of Brownian motion, i.e., we sould like to form integrals

$$\int_{t_1}^{t_2} Z_t(\omega) W'_t(\omega)\, dt\,.$$

Unfortunately, this is not possible, since the paths of W_t are nowhere differentiable almost surely. See Theorem 6.1 on p.127. Riemann–Stieltjes integrals provide a way out of this dilemma. We will discuss this topics briefly in this subchapter.

Remark 9.1. Let $a, b \in \mathbb{R}$ such that $a < b$, let $f, g : [a, b] \to \mathbb{R}$ be such that the derivative $g'(t)$ exists for all $a < t < b$. By definition of the Riemann integral as the limit of Riemann sums,

$$\int_a^b f(t)g'(t)dt \;=\; \lim_{\|\Pi\|\to 0} \sum_{j=0}^{n-1} f(u_j)g'(u_j)\,(t_{j+1} \,-\, t_j) \quad (t_j \le u_j \le t_{j+1} \text{ for all } j)\,,$$

where the limit is taken over partitions $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ in such a way that mesh $\|\Pi\| = \max_j(t_{j+1} - t_j)$ converges to zero. See Definition 6.9 (Quadratic Variation) on p.134. Of course we must assume that this limit exists.

For small differences $t_{j+1} - t_j$ we obtain approximately $g'(u_j) \;\approx\; \dfrac{g(t_{j+1}) - g(t_j)}{t_{j+1} - t_j}$, hence

$$\begin{aligned}
\int_a^b f(t)g'(t)dt \;&\approx\; \sum_{j=0}^{n-1} f(u_j)g'(u_j)\,(t_{j+1} \,-\, t_j) \\
&\approx\; \sum_{j=0}^{n-1} f(u_j)\frac{g(t_{j+1}) - g(t_j)}{t_{j+1} - t_j}\,(t_{j+1} \,-\, t_j) \\
&=\; \sum_{j=0}^{n-1} f(u_j)\,(g(t_{j+1}) \,-\, g(t_j))\;.
\end{aligned}$$

Thus, if the right–hand limit for $\|\Pi\| = \max_j(t_{j+1} - t_j) \to 0$ exists, it will be a generalization of $\int_a^b f(t)g'(t)dt$, in case that g is not differentiable. □

This leads to the next definition.

Definition 9.1 (Riemann–Stieljes Integral)**.**

Let $a, b \in \mathbb{R}$ such that $a < b$ and $f, g : [a, b] \to \mathbb{R}$. If

$$\int_a^b f(t)dg(t) \ := \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} f(u_j)\,(g(t_{j+1}) \,-\, g(t_j))$$

exists as limit over all partitions Π of the interval $[a, b]$, then we call $\int_a^b f(t)dg(t)$ the **Riemann–Stieltjes integral** of f with respect to g over $[a, b]$ with **integrand** f and **integrator** g. ☐

The above definition will become the starting points for stochastic integrals $\int_a^b Z_t dW_t$ with respect to Brownian motion.

9.2 The Itô Integral for Simple Processes

This chapter is very sketchy as far as proofs are concerned since the material follows extremely closely that of SCF2 Chapter 4.

Unless explicitly stated otherwise $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ is a filtered probability space and $W = W_t$ is a Brownian motion on Ω with respect to $\mathfrak{F}_t$.
Often we assume a fixed expiration time $T > 0$ and W and all other stochastic processes have index set $[0, T]$, but occasionally we also consider other index sets. Usually this would be the interval $[0, \infty[$ of all times, or it would be the interval $[t_0, T]$ in which $0 \le t_0 < T$ asumes the role of a start time.

The following definitions are from SCF2 ch.4.2.1.

Definition 9.2 (Simple Process)**.**

Let $T > 0$ be fixed, and let $\Pi := \{\, t_0,\, t_1,\, \dots,\, t_n \,\}$ be a partition of $[0, T]$. In other words,

$$0 \,=\, t_0, < \, t_1, < \, \cdots \, < \, t_n \,=\, T.$$

An adapted process $Z = Z_t$ is called a **simple process** if $t \mapsto Z_t(\omega)$ is constant on each interval $[t_j, t_{j+1}[$ almost surely. ☐

Definition 9.3 (Itô Integral of a Simple Process)**.**

Let $\Pi := \{t_0, t_1, \dots, t_n\}$, where $0 = t_0, < t_1, < \cdots < t_n = T$ be a partition of $[0, T]$, and let Z_t be a simple process on Ω which has constant trajectories on each partitioning interval $[t_j, t_{j+1}[$. Let

$$\text{(9.1)} \qquad \int_0^t Z_u dW_u := \begin{cases} \sum_{j=0}^{k-1} Z(t_j)[W(t_{j+1}) - W(t_j)] + Z(t_k)[W_t - W(t_k)] & \text{if } 0 \le t < T\,, \\ \sum_{j=0}^{n-1} Z(t_j)[W(t_{j+1}) - W(t_j)] & \text{if } t = T\,, \end{cases}$$

where the index k is chosen such that $t_k \le t < t_{k+1}$. We call $\int_0^t Z_u dW_u$ the **Itô integral** of Z with respect to W. □

Theorem 9.1 (SCF2 Theorem 4.2.1)**.**

The Itô integral $\int_0^t Z_u dW_u$ is an $\mathfrak{F}_t$–martingale.

PROOF: See SCF2. ■

Because $I_t = \int_0^t Z_u dW_u$ is a martingale and $I(0) = 0$, it follows that

$$E[I_t] = 0 \text{ for all } t \ge 0. \quad \text{Thus } \; Var[I_t] = E[I_t^2].$$

The next theorem shows how to evaluate $E[I_t^2]$.

Theorem 9.2 (SCF2 Theorem 4.2.2 - Itô isometry)**.**

The Itô integral defined by (9.1) *on p.187 satisfies*

$$\text{(9.2)} \qquad E[I_t^2] \;=\; E\left[\int_0^t Z_u^2 du\right].$$

PROOF: See SCF2. ■

Theorem 9.3 (SCF2 Theorem 4.2.3)**.**

The quadratic variation $[I, I]_t$ up to time t of the Itô integral $I_t = \int_0^t Z_u dW_u$ is

$$\text{(9.3)} \qquad [I, I]_t \;=\; \int_0^t Z_u^2 du.$$

PROOF: See SCF2. ■

Remark 9.2. If we think of integration and differentiation as operations that cancel each other when we look at $\int_0^t Z_u dW_u$ as a function of the upper limit of integration then we obtain

$$\textbf{(A)} \qquad d\int_0^t Z_u\,dW_u \;=\; Z_t\,dW_t$$

Strictly speaking the above is the <u>definition</u> of the **differential** $d\int\limits_0^t Z_u dW_u$ in terms of the right hand side. The above makes a lot of sense for $Z_t = 1$: If we take the partition $\Pi = \{0, t\}$ then Definition 9.3 (Itô Integral of a Simple Process) yields

$$\int_0^t 1\,dW_u \;=\; 1(W_t - W_0) \;=\; W_t, \quad \text{thus applying } d \text{ on both sides should give } d\int_0^t 1\,dW_u \;=\; dW_t.$$

Formula (**A**) results in exactly that last equation when $Z_t = 1$. $\square$

Remark 9.3. We write the Itô integral $I_t = \int\limits_0^t Z_u dW_u$ as a differential

$$d\,I_t \;=\; d\int_0^t Z_u\,dW_u \;=\; Z_t\,dW_t.$$

We square both sides of this equation and obtain

$$dI_t\,dI_t \;=\; Z_t^2\,dW_t\,dW_t \;=\; Z_t^2\,dt.$$

See Remark 6.9 on p.135 for the last equation. $\square$

9.3 The Itô Integral for General Processes

We apply Example 6.2**(f)** on 130 and formula (6.20) of that example to the following.

Definition 9.4 (L^2 convergence of random variables). ⋆

Given is a probability space $(\Omega, \mathfrak{F}, P)$, $T > 0$. Let Z and Z' be random variables which are **square integrable**, i.e., $E[Z^2] < \infty$ and $E[Z'^2] < \infty$. Then

$$(9.4) \qquad \|Z\|_{L^2} \;=\; \sqrt{\int Z^2 dP} \;=\; \sqrt{E[Z^2]} \;<\; \infty\,,$$

$$(9.5) \qquad d_{L^2}(Z, Z') \;=\|Z - Z'\|_{L^2} \;=\; \sqrt{E\,[Z - Z']^2} \;<\; \infty\,.$$

Let $Z^{(n)}$ and Z, where $n \in \mathbb{N}$, be square integrable random variables. We say that the sequence $Z^{(n)}$ **converges in $\mathbf{L^2}$** to Z, and we write

$$(9.6) \quad L^2\text{–}\lim_{n\to\infty} Z^{(n)} = Z, \;\text{ if }\; \lim_{n\to\infty} d_{L^2}(Z^{(n)}, Z) \;=\; 0, \;\text{ i.e., }\; \lim_{n\to\infty} E\left[(Z^{(n)} - X)^2\right] \;=\; 0. \;\square$$

Definition 9.5 (L^2 convergence of stochastic processes). ⋆

Given is a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_t, P)$, $T > 0$. Let $X = (X_u)_{0 \le u \le T}$ be an adapted process. We say that X_t is **square integrable**, if $E\left[\int\limits_0^T X_u^2 du\right] < \infty$.

Let $X_u, X_u^{(1)}, X_u^{(2)}, X_u^{(3)}, \cdots$ be adapted, square integrable, stochastic processes. We say that the sequence $X^{(n)}$ **converges in $\mathbf{L^2}$** to X, and we write

$$(9.7) \qquad L^2\text{–}\lim_{n\to\infty} X^{(n)} = X, \quad \text{if} \quad \lim_{n\to\infty} E\left[\int_0^T (X_u^{(n)} - X_u)^2\, du\right] = 0. \ \square$$

Fact 9.1.

Let $T > 0$. Let $Z_u, 0 \le t \le T$, be an adapted and square–integrable process. Then

(a) *One can find a sequence $Z(n)$ of* simple *processes, also square–integrable, such that $L^2\text{–}\lim\limits_{n\to\infty} Z^{(n)} = Z$ (see formula (9.7)).*

(b) *There exists an adapted process $\Phi = \Phi_t$ with* continuous *paths such that the Itô integrals $I_t^{(n)} := \int\limits_0^t Z_u^{(n)} dW_u$ converge in L^2 to Φ, i.e.,*

$$(9.8) \qquad \lim_{n\to\infty} E\left[\int_0^T (I_u - \Phi_u)^2\, du\right] = 0.$$

(c) *If $Z'(n)$ is another sequence of simple and square–integrable processes such that $L^2\text{–}\lim\limits_{n\to\infty} Z'^{(n)} = Z$, and if Φ'_t is another square–integrable process with continuous paths such that $L^2\text{–}\lim\limits_{n\to\infty} I_t^{(n)} := \int\limits_0^t Z_u'^{(n)} dW_u = \Phi'$, then there exists a set of probability zero which contains the set $\{\omega \in \Omega : \Phi(\cdot, \omega) \ne \Phi'(\cdot, \omega)\}$.* $\square$

Remark 9.4. ⋆ We would not be able to ascertain in Fact 9.1**(c)** that the trajectories $t \mapsto \Phi(t, \omega)$ and $t \mapsto \Phi'(t, \omega)$ are identical, except on a set of probability zero, without assuming that those trajectories are continuous. $\square$

Definition 9.6 (Itô integral for general integrands)**.** We write

$$(9.9) \qquad \int_0^t Z_u\, dW_u$$

for the process $\quad \Phi_t = L^2\text{–}\lim\limits_{n\to\infty} \int_0^t Z_u^{(n)}\, dW_u\,,$

described in **(b)** of Fact 9.1, and call it the **Itô integral** of Z_t with respect to W_t. $\square$

Remark 9.5. Chances are that you have overlooked the following dissimilarity between the sums $\sum_{j=0}^{n-1} f(u_j)\,(g(t_{j+1}) - g(t_j))$ which approximate the Riemann Stieltjes integral $\int f(s)dg(s)$ and the sums $\sum_{j=0}^{n-1} Z(t_j)[W(t_{j+1}) - W(t_j)]$ which approximate the Itô integral $\int Z_s dW_s$. In the first case we only require for the arguments u_j of the integrand that $t_j \leq u_+j \leq t_{j+1}$, in the second case we specifically demand that $u_j = t_j$, i.e., the arguments of the integrand must be the left endpoints of the partitioning intervals.

Why do we not allow the argument u_j to vary In the definition of the Itô integral? Because doing so would rule out even a nice, continuous process such as W_t as an integrand: Let

$$\Pi \;=\; \{\, t_0,\, t_1,\, \dots,\, t_n \,\}, \quad \text{where } \; 0 \;=\; t_0, < \, t_1, < \, \cdots \, < \, t_n \;=\; T,$$

be a partition of $[0, T]$ and let

$$X_T^\Pi \;:=\; \sum_{j=0}^{n-1} W_{t_j}(W_{t_{j+1}} - W_{t_j})\,, \qquad Y_T^\Pi \;:=\; \sum_{j=0}^{n-1} W_{t_{j+1}}(W_{t_{j+1}} - W_{t_j})\,.$$

According to Definition 9.3 (Itô Integral of a Simple Process) on p.186,

(A)
$$\int_0^T W_s dW_s \;=\; L^2\!\!-\lim_{\|\Pi\| \to 0} X_T^\Pi\,.$$

If the choice of u_j did not matter as long as $t_j \leq u_+j \leq t_{j+1}$, then it should also be true that

(B)
$$\int_0^T W_s dW_s \;=\; L^2\!\!-\lim_{\|\Pi\| \to 0} Y_T^\Pi\,.$$

However, these limits are fundamentally different since $E[X_T^\Pi] = 0$, and $E[Y_T^\Pi] = T$ for all partitions Π, [39] hence the expectation of **(A)** is zero and that of **(B)** is T.

So why then did we choose in formula (9.1) of Definition 9.2 above to pick the values Z_{t_j} which correspond to the left bounds of the intervals $[t_j, t_{j+1}[$ rather than, say. the values $Z_{(t_{j+1}-t_j)/2}$ taken at the midpoints or the values $Z_{t_{j+1}}$ taken at the right bounds?

There are some important technical reasons. For example Theorem 9.1 which follows this remark asserts that the Itô integral is a martingale when viewed as a process $t \mapsto \int_0^t Z_u dW_u$. If $u_j > t_j$ then this theorem will generally no longer be valid.

> But at least as important is the way we use Itô integrals when modeling financial markets. The Brownian motion increments $W_{t_{j+1}} - W_t$ represent uncertainty that happens in the future, whereas the history of the integrand Z_t up to the "present" t_j is known to us (since it is $\mathfrak{F}_{t_j}$–measurable for all times $t < t_J$ of the past.) $\square$

Theorem 9.4 (SCF2 Theorem 4.3.1 - Itô isometry)**.** *Given are the Itô integrals*

[39] You are asked to prove that $E[X_t^\Pi] = 0$, and $E[Y_t^\Pi] = t$ in Exercise 9.1 on p.199.

$$\int_0^t Z_u dW_u \;:=\; L^2\text{–}\lim_{n\to\infty}\int_0^t Z_u^{(n)} dW_u\,, \quad \int_0^t Z_u dW_u \;:=\; L^2\text{–}\lim_{n\to\infty}\int_0^t Z_u^{(n)} dW_u\,. \qquad \textit{Then}$$

a. ***(Continuity)*** *The paths of* $t \mapsto \int_0^t Z_u dW_u$ *are continuous.*

b. ***(Adaptivity)*** *For each* t, I_t *is* $\mathfrak{F}_t$*–measurable.*

c. ***(Linearity)*** *If* $\alpha, \beta \in \mathbb{R}$, *then* $\int_0^t \big(\alpha Y_u + \beta Z_u\big) dW_u \;=\; \alpha \int_0^t Y_u dW_u \;+\; \beta \int_0^t Z_u dW_u$.

In particular, for every constant α, $\int_0^t \alpha Z_u dW_u = \alpha \int_0^t Z_u dW_u$.

d. ***(Martingale)*** $\int_0^t Z_u dW_u$ *is a martingale.*

e. ***(Itô isometry)*** $E\left[\int_0^t Z_u dW_u\right] \;=\; E\left[\int_0^t Z_u^2 du\right]$.

f. ***(Quadratic variation)*** *If* $I_t \;=\; \int_0^t Z_u dW_u$, *then* $[I,I]_t \;=\; \int_0^t Z_u^2 du$.

PROOF: Not given. ■

9.4 The Itô Formula for Functions of Brownian Motion

Theorem 9.5 (SCF2 Theorem 4.4.1 - Itô–Doeblin formula for Brownian motion)**.**

Let $f(t,x)$ *be a function for which the partial derivatives* $f_t(t,x)$, $f_x(t,x)$, *and* $f_{xx}(t,x)$ *are defined and continuous, and let* W_t *be a Brownian motion. Then, for every* $T \geq 0$,

$$\begin{aligned} &f\big(T, W_T\big) - f\big(0, W(0)\big) \\ &\quad= \int_0^T f_t\big(t, W_t\big)dt \;+\; \int_0^T f_x\big(t, W_t\big)dW_t \;+\; \frac{1}{2}\int_0^T f_{xx}\big(t, W_t\big)dt. \end{aligned} \tag{9.10}$$

PROOF: See SCF2 for a sketch. ■

9.5 The Itô Formula for Functions of an Itô Process

Definition 9.7 (SCF2 Definition 4.4.3 - Itô process)**.** Let $W_t, t \geq 0$, be a Brownian motion, and let $\mathfrak{F}_t, t \geq 0$, be an associated filtration.

An **Itô process** on $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ is a stochastic process

(9.11) $$X_t = x + \int_0^t \Delta_u dW_u + \int_0^t \Theta_u du,$$

which we also equivalently express as

(**A**) $$dX_t = \Delta_t dW_t + \Theta_t dt,$$

(**B**) $$X_0 = x\,.$$

Here Δ_t and Θ_t are $\mathfrak{F}_t$–adapted processes, and $x \in \mathbb{R}$. We call (**A**) the **stochastic differential**, also just the **dynamics**, and (**B**) the **initial condition** of (9.11). Furthermore we say that (**A**) and **(B)** express (9.11) in differential notation, and that (9.11) expresses (**A**) and (**B**) as an **integral equation**. □

Remark 9.6.

(1). The phrase ".... which we also equivalently express as ...." is to be taken literally: We do not mathematically distinguish between the integral equation (**B**) and the associated set of stochastic differential (**A**) plus initial condition (**B**). They mean exactly the same thing.

(2). We bury into this footnote [40] a technical remark taken literally from SCF2. □

Lemma 9.1 (SCF2 Lemma 4.4.4)**.** *The quadratic variation of the Itô process* (9.11) *is*

(9.12) $$[X, X]_t = \int_0^t \Delta_u^2 du.$$

PROOF: See SCF2 for a sketch. ■

Definition 9.8 (SCF2 Definition 4.4.5)**.**

Given are an Itô process

$$X_t = X_0 + \int_0^t \Delta_u dW_u + \int_0^t \Theta_u du,$$

on $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ and an adapted process $\Gamma_t, t \geq 0$.

We define [41]

(9.13) $$\int_0^t \Gamma_u dX_u := \int_0^t \Gamma_u \Delta_u dW_u + \int_0^t \Gamma_u \Theta_u du.$$ □

[40]**This note literally from SCF2:** We assume that $\int_0^t \Delta_u dW_u$ and $\int_0^t \Theta_u du$ are finite for every $t > 0$ so that the integrals on the right–hand side of formula (9.11) are defined and the Itô integral is a martingale. We shall always make such integrability assumptions, but we do not always explicitly state them.

[41]We assume that $E\left[\int_0^t \Gamma_u^2 \Delta_u^2 du\right]$ and $\int_0^t |\Gamma_u \Theta_u| du$ are finite for each $t > 0$ so that the integrals on the right-hand side of (9.13) are defined.

Theorem 9.5 (Itô–Doeblin formula for Brownian motion) on p.191. which was stated for functions $f(t, W_t)$ can be generalized to functions $f(t, X_t)$ where the second argument is an Itô process. This will be done here.

Theorem 9.6 (SCF2 Theorem 4.4.6 - Itô–Doeblin formula for an Itô process)**.**

Let $X_t, t \geq 0$ be an Itô process as described in Definition 9.7 on p.191, and let $(t,x) \mapsto f(t,x)$ be a function with continuous partial derivatives $f_t(t,x)$, $f_x(t,x)$, and $f_{xx}(t,x)$. Then, for every $T \geq 0$,

$$\begin{aligned} f(T, X_T) &= f(0, X_0) + \int_0^T f_t(t, X_t)dt + \int_0^T f_x(t, X_t)dX_t \\ &\quad + \frac{1}{2}\int_0^T f_{xx}(t, X_t)d[X,X]_t \\ &= f(0, X_0) + \int_0^T f_t(t, X_t)dt + \int_0^T f_x(t, X_t)\Delta_t dW_t \\ &\quad + \int_0^T f_x(t, X_t)\Theta_t dt + \frac{1}{2}\int_0^T f_{xx}(t, X_t)\Delta_t^2 dt. \end{aligned} \tag{9.14}$$

PROOF: See SCF2. ■

Remark 9.7. The reader may wonder about the meaning of the term "$d[X,X]_t$". We claim that

$$d[X,X]_t \;=\; dX_t\,dX_t\,.$$

This is seen as follows. According to Lemma 9.1 on p.192, $[X,X]_t = \int_0^t \Delta_u^2 du$. This means that $[X,X]_t$ is an Itô process. (Set $\Delta_u = 0$ in Definition 9.7 of an Itô process which precedes that lemma.) The differential form of this Itô process is, according to **(A)** of that definition, $d[X,X]_t = \Delta_t^2 dt$.
We will see in ($\star$), which occurs further down in this remark, that $dX_t dX_t = \Delta_t^2 dt$. A comparison of those two equation yields $d[X,X]_t = dX_t dX_t$.

Itô formula for an Itô process in differential notation:

$$df(t, X_t) = f_t(t, X_t)\,dt + f_x(t, X_t)\,dX_t + \frac{1}{2} f_{xx}(t, X_t)\,dX_t dX_t. \tag{9.15}$$

The differential form of $X_t = X_0 + \int_0^t \Delta_u dW_u + \int_0^t \Theta_u du$ is

$$dX_t \;=\; \Delta_t dW_t + \Theta_t dt$$

from this we compute $dX_t dX_t$ using the multiplication table as follows.

$$\begin{aligned} dX_t\,dX_t &= (\Delta_t dW_t + \Theta_t dt)\,(\Delta_t dW_t + \Theta_t dt) \\ &= \Delta_t^2 dW_t dW_t + 2\Delta_t\Theta_t dW_t dt + \Theta_t^2 dt dt = \Delta_t^2 dt \end{aligned} \tag{$\star$}$$

We make these substitutions in (9.15) and group the dt terms:

$$
(9.16) \qquad \begin{aligned} df(t, X_t) &= f_x(t, X_t)\Delta_t \, dW_t \\ &+ \left(f_t(t, X_t) + f_x(t, X_t)\Theta_t + \frac{1}{2} f_{xx}(t, X_t)\Delta_t^2 \right) dt. \ \square \end{aligned}
$$

Example 9.1 (Generalized Geometric Brownian Motion)**.** Definition 6.11 on p.137 gave the definition of geometric Brownian Motion as the process

$$
S_t = S_0 \exp\left[\sigma W_t + \left(\alpha - \frac{1}{2}\sigma^2 \right) t \right],
$$

defined on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ with a Brownian motion $W = W_t$.
We will obtain this process in a more general setting as the solution of a stochastic differential equation. Let

$$
(9.17) \qquad X_t = \int_0^t \sigma_u dW_u + \int_0^t \left(\alpha_u - \frac{1}{2}\sigma_u^2 \right) du,
$$

where α_t and σ_t are adapted processes. Then X is an Itô process with differential

$$
(9.18) \qquad dX_t = \sigma_t dW_t + \left(\alpha_t - \frac{1}{2}\sigma_t^2 \right) dt, \quad X_0 = 0.
$$

From the multiplication table we obtain its squared differential

$$
(9.19) \qquad dX_t dX_t = \sigma_t^2 dW_t dW_t = \sigma_t^2 dt.
$$

Let $S_0 \in]0, \infty[$ (i.e., S_0 is deterministic), and $f(x) := S_0 e^x$. Since f does not have t as an argument it is constant in t, thus $f_t = 0$. There also is no need for using partial derivatives notation and we can write $f'(x)$ for $f_x(x)$ and $f''(x)$ for $f_{xx}(x)$. Note that

$$
f'(x) = f''(x) = f(x) = S_0 e^x.
$$

We define **generalized geometric Brownian motion** as the process

$$
(9.20) \qquad S_t := S_0 e^{X_t} = S_0 \exp\left[\int_0^t \sigma_s dW_s + \int_0^t \left(\alpha_s - \frac{1}{2}\sigma_s^2 \right) ds \right],
$$

Since $S_t = f(X_t)$ an application of the Itô formula yields

$$
(9.21) \qquad \begin{aligned} dS_t &= df(X_t) = f'(X_t)\, dX_t + \frac{1}{2} f''(X_t)\, dX_t dX_t \\ &= S_0 e^{X_t}\, dX_t + \frac{1}{2} S_0 e^{X_t}\, dX_t dX_t = S_t\, dX_t + \frac{1}{2} S_t dX_t dX_t. \end{aligned}
$$

This last formula describes a **stochastic differential equation**. It defines the random process S_t via a formula for its differential dS_t, and this formula involves, besides the random process S_t itself, also the differential dX_t of an Itô process X_t. $\square$

Remark 9.8. It follows from formulas (9.18) and (9.19) that

$$\begin{aligned} S_t dX_t &\overset{(9.18)}{=} \sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt \,-\, \frac{1}{2}\sigma_t^2 S_t\, dt \\ &\overset{(9.19)}{=} \sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt \,-\, \frac{1}{2} S_t\, dX_t dX_t, \end{aligned}$$

We plug this expression for $S_t dX_t$ into the last equation of (9.21) and obtain

$$\begin{aligned} dS_t &= \left(\sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt \,-\, \frac{1}{2} S_t\, dX_t dX_t\right) \,+\, \frac{1}{2}\, S_t dX_t dX_t \\ &= \sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt. \end{aligned}$$

This last formula is another example of a stochastic differential equation. It improves on the one given at the end of Example 9.1, since the differential dW_t of a Brownian motion replaces that of the more general Itô process X_t.

Here is a Financial market interpretation of this formula

(9.22) $$dS_t \,= \alpha_t S_t\, dt \,+\, \sigma_t S_t\, dW_t$$

which describes the dynamics of S_t. If this process denotes the price of a stock, then (9.22) expresses that this asset has an **instantaneous mean rate of return** α_t and **volatility** σ_t. "Instantaneous" indicates that $t \mapsto \alpha_t(\omega)$ depends on the paricular time (and the sample path ω) where the price is observed.

Generalized GBM is a good model for the price evolution of a stock for the following reasons.

- It is always positive.
- The fluctuations introduced by the random term $\sigma_t dW_t$ express the risk inherent in investing in such an asset.

The drawback: The trajectories of S_t are continuous at all points in time. To consider asset prices with jumps a different model is needed.

In the Black–Scholes market we specialize to constant α and σ. Then (9.20) becomes ordinary GBM

(9.23) $$S_t \,= S_0\, \exp\left\{\sigma W_t \,+\, \left(\alpha - \frac{1}{2}\sigma^2\right)\, t\right\}.$$

If we further assume that the instantaneous mean rate of return α is zero then the asset price and its dynamics are

$$S_t \,= S_0\, \exp\left\{\sigma W_t \,-\, \frac{1}{2}\sigma^2\, t\right\}, \qquad dS_t \,=\, \sigma S_t\, dW_t.$$

We recognize S_t as the level σ exponential martingale of Definition 6.12 on p.137. We obtain a new proof that S_t is a martingale from the fact that $dS_t = \sigma S_t\, dW_t$ reveals this process as a stochastic integral with respect to Brownian motion,

$$S_t \,= S_0 \,+\, \int_0^t \sigma_u S_u\, dW_u. \;\square$$

Theorem 9.7 (SCF2 Theorem 4.4.9 - Itô integral of a deterministic integrand)**.**

Let $W_s, s \geq 0$, be a Brownian motion and let Δ_s be a nonrandom function of time. Let

$$I_t := \int_0^t \Delta_s \, dW_s \,.$$

Then, for each $t \geq 0$, the random variable I_t is normally distributed with

$$E[I_t] = 0 \,, \qquad Var[I_t] = \int_0^t \Delta_s^2 \, ds \,.$$

PROOF: See SCF2. ■

Here are some examples of the Itô firmula.

Example 9.2. Source: [7] Björk, Thomas: Arbitrage Theory in Continuous Time.
Assume that Z is a normal variable with expectation zero. Compute Compute $E[Z^4]$.
We will solve this problem with stochastic calculus by transforming it into one concerning Brownian motion W_t. We accomplish this by writing $t := Var[Z]$. Then Z and W_t have the same distribution. Hence, $E[Z^4] = E[W_t^4]$. Let $X_t = W_t^4$. Then $X_t = f\big(t, W_t\big)$, where f is given by $f(t,x) = x^4$. The partial derivatives are

$$\frac{\partial f}{\partial t} = 0 \,, \quad \frac{\partial f}{\partial x} = 4x^3 \,, \quad \frac{\partial^2 f}{\partial x^2} = 12x^2 \,.$$

The Itô formula plus the equation $W_0^4 = 0$ yield

$$dX_t = df(t, W_t) = f_t dt + f_x dW_t + \frac{1}{2} f_{xx} dt = 0 + 4W_t^3 dW_t + 6W_t^2 dt \,; \quad X_0 = 0 \,.$$

The equivalent integral form is $X_t = 0 + 6\int_0^t W_s^2 ds + 4\int_0^t W_s^3 dW_s$. We take expected values of all members of this equation. Since Itô integrals $\int \dots dW$ are martingales,

$$E\left[\int_0^t W_s^3 dW_s\right] = E\left[\int_0^0 W_s^3 dW_s\right] = 0 \,.$$

Since $E[\dots]$ is an abstract integral $\int \dots dP$, Fubini allows us to move the expectation inside the ds–integral. We obtain

$$E[X_t] = 6\int_0^t E[W_s^2] ds = 6\int_0^t s ds = 3t^2 . \ \square$$

Example 9.3. Let W be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$.
Let the processes A_t and B_t be defined as follows.

$$\begin{aligned} dA_t &= 5A_t \, dt - A_t \, dW_t, \quad A_0 = 0 \,, \\ B_t &= e^{-5t} A_t \,. \end{aligned}$$

Apply Itô's formula to the function $f(t,x) = e^{-5t}x$ to

(a) compute dB_t so it has the form $dB_t = U_t dt + V_t dW_t$ where U_t and V_t are adapted stochastic processes.

(b) Prove that V_t is a martingale. This is easy once you have computed part **(a)**

We solve this problem as follows.

The partial derivatives of f are

$$f_t(t,x) \;=\; -5e^{-5t}x\,, \quad f_x(t,x) \;=\; e^{-5t}\,, \quad f_{xx}(t,x) \;= 0\,.$$

Further, it follows from $dtdt = dtdW_t = dW_t dt = 0$ and $dW_t dW_t = dt$, that

$$dA_t dA_t \;=\; (-A_t)^2 dt \;=\; A_t^2 dt\,.$$

Observe that we won't need this, since $f_{xx} = 0$. Since $B_t = f(t, A_t)$, Itô's formula yields

$$\begin{aligned} dB_t &= df(t,A_t) \;=\; f_t dt \,+\, f_x dA_t \,+\, \frac{1}{2} f_{xx} dA_t\, dA_t \\ &= (-5)e^{-5t}A_t dt \,+\, e^{-5t}\big(5A_t\, dt \,-\, A_t\, dW_t\big) \,+0 \\ &= (-5)e^{-5t}A_t dt \,+\, 5e^{-5t}A_t\, dt \,-\, e^{-5t}A_t\, dW_t \;=\; -e^{-5t}A_t\, dW_t\,. \end{aligned}$$

We have solved **(a)** (with $U_t = 0$ and $V_t = -e^{-5t}A_t$) and also **(b)**, since the integrated form of the above is

$$B_t \;=\; B_0 \,-\, \int_0^t e^{-5u}A_u dW_u \;=\; -\int_0^t e^{-5u}A_u dW_u\,,$$

and integrals with respect to Brownian motion are martingales.

As an aside, we also note that B_t is a generalized geometric Brownian motion: Since $B_t = e^{-5t}A_t$, $dB_t \;=\; -e^{-5t}A_t dW_t$ can be rewritten as

$$dB_t \;=\; -B_t dW_t\,.$$

Thus the differential B_t is if the form (9.22) when we set $\alpha_t = 0$ and $\sigma_t = -1$. Since α and σ are constant in t and ω, B_t actually is a (non–generalzied) geometric Brownian motion. □

Example 9.4. Let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ be a filtered probability space with a Brownian motion W_t. Let

$$X_t \;=\; 5 \,+\, \int_0^t W_u du \,+\, 2\int_0^t W_u^2 dW_u\,.$$

What is $d\big(t^2 X_t^2\big)$? We will apply the Itô formula to compute this differential as follows.

Since $dX_t = W_t dt + 2\,W_t^2 dW_t$, and $dtdt = dtdW_t = dW_t dt = 0$, and $dW_t dW_t = dt$,

$$\begin{aligned} dX_t\, dX_t &= (W_t dt \,+\, 2\,W_t^2 dW_t)(W_t dt \,+\, 2\,W_t^2 dW_t) \\ &= W_t^2\, dt\, dt \,+\, 2\big(2W_t^3\, dt\, dW_t\big) \,+\, 2^2\, W_t^2\, dW_t dW_t\,. \end{aligned}$$

We aim to compute $df(t, X_t)$ for the function $f(t,x) = t^2x^2$. Since

$$f_t = 2tx^2; \quad f_x = 2t^2x; \quad f_{xx} = 2t^2\,,$$

Itô's formula yields

$$\begin{aligned} d\big(f(t,X_t)\big) &= 2tX_t^2dt \,+\, 2t^2X_tdX_t \,+\, \frac{2}{2}t^2dX_tdX_t \\ &= 2tX_t^2dt \,+\, 2t^2X_td[W_tdt \,+\, 2W_t^2dW_t] \,+\, t^2(4W_t^4)dt \\ &= 2tX_t^2dt \,+\, 2t^2X_tW_tdt \,+\, 2t^2X_t2W_t^2dW_t \,+\, t^2(4W_t^4)dt \\ &= \big[2tX_t^2 \,+\, 2t^2X_tW_t \,+\, 4t^2W_t^4\big]dt \,+\, 4t^2X_tW_t^2dW_t\,. \ \square \end{aligned}$$

The following propositions are applications of the Itô formula to interest rate models.

Proposition 9.1 (SCF2 Example 4.4.10 - Vasicek interest rate model). ⋆

Given is a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ with a Brownian motion $W = W_t$. The ***Vasicek model*** *is a financial market in which the interest rate $R = R_t(\omega)$ has dynamics*

(9.24) $$dR_t \;=\; \big(\alpha - \beta R_t\big)\,dt \,+\, \sigma\,dW_t\,.$$

Here we assume that $\alpha, \beta, \sigma \in]0, \infty[$, i.e., they are positive and deterministic constants.
The solution to this SDE is

(9.25) $$R_t \;=\; e^{-\beta t}R_0 \,+\, \frac{\alpha}{\beta}(1 - e^{-\beta t}) \,+\, \sigma e^{-\beta t}\int_0^t e^{\beta s}dW_s.$$

For a proof see SCF2. $\square$

Remark 9.9. ⋆

The following results from that last proposition. Since the normal density is strictly positive for all arguments, there is positive probability that R_t is negative, no matter how one choses $\alpha > 0$, $\beta > 0$, and $\sigma > 0$. This is not desirable for an interest rate model.

On the other hand, the Vasicek model has the desirable property that the interest rate is mean–reverting:

- When $R_t = \frac{\alpha}{\beta}$, the drift term (the dt term) in (9.24) is zero.
- When $R_t > \frac{\alpha}{\beta}$, this term is negative, which pushes R_t back toward $\frac{\alpha}{\beta}$.
- When $R_t < \frac{\alpha}{\beta}$, this term is positive, which pushes R_t back toward $\frac{\alpha}{\beta}$.

Moreover, we have the following:

- if $R_0 = \frac{\alpha}{\beta}$, then $E[R_t] = \frac{\alpha}{\beta}$ for all $t \geq 0$,
- if $R_0 \neq \frac{\alpha}{\beta}$, then $\lim\limits_{t\to\infty} E[R_t] = \frac{\alpha}{\beta}$. $\square$

Proposition 9.2 (SCF2 Example 4.4.11 - Cox–Ingersoll–Ross (CIR) interest rate model). ⋆

Given is a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ *with a Brownian motion* $W = W_t$*. Assume that the interest rate* $R = R_t(\omega)$ *in a market economy is modeled by the SDE*

$$xx \tag{9.26}$$

$$dR_t = (\alpha - \beta R_t)dt + \sigma\sqrt{R_t}\,dW_t, \tag{9.27}$$

$\alpha, \beta, \sigma \in]0, \infty[$ *are positive and deterministic constants. We call this the* ***Cox–Ingersoll–Ross model****, We also abbreviate this as the* ***CIR model****.*

The CIR model has the following properties:

$$E[R_t] = e^{-\beta t}R_0 + \frac{\alpha}{\beta}(1 - e^{-\beta t}). \tag{9.28}$$

Note that this is the same expectation as in the Vasicek model.

$$Var[R_t] = \frac{\sigma^2}{\beta}R_0(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha\sigma^2}{2\beta^2}(1 - 2e^{-\beta t} + e^{-2\beta t}). \tag{9.29}$$

In particular,

$$\lim_{t\to\infty} Var[R_t] = \frac{\alpha\sigma^2}{2\beta^2}. \ \square$$

For a proof see SCF2. ■

The next theorem will be proven later, when we have the multidimensional Itô formula at our disposal. We state it here since we use it in Chapter 10 (Black–Scholes Model Part I: The PDE)

Theorem 9.8.

If X_t *and* Y_t *are two Itô processes then*

$$d(X_t Y_t) = X_t\,dY_t + Y_t\,dX_t + dX_t\,dY_t. \tag{9.30}$$

PROOF: Will be given later, in Chapter 11 (Multidimensional Stochastic Calculus). See Corollary 11.1 (Itô product rule) on p.224. ■

9.6 Exercises for Ch.9

Exercise 9.1. Prove the following assertion which was made in Remark 9.5 on p.190 of this document: Let $\Pi = \{t_0, t_1, \dots, t_n\}$ $(0 = t_0 < t_1 < \cdots < t_n = T)$ and

$$X_T^{\Pi} := \sum_{j=0}^{n-1} W_{t_j}(W_{t_{j+1}} - W_{t_j}), \qquad Y_T^{\Pi} := \sum_{j=0}^{n-1} W_{t_{j+1}}(W_{t_{j+1}} - W_{t_j}).$$

Here W_t is a Brownian motion on $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$, $W_j := W_{t_j}$, and $I_j := [t_j, t_{j+1}[$. Then

$$E[X_t^{\Pi}] = 0, \quad \text{and} \quad E[Y_t^{\Pi}] = T. \ \square$$

Exercise 9.2. Let W_t be a Brownian motion, Y_t an adapted process on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Assume that the process X has dynamics

$$d\,X_t \;=\; Y_t^2\,dW_t; \qquad X_0 \;=\; 16.$$

Compute $E[X_{10}]$.
Hint: Stochastic integrals with respect to Brownian motion are martingales. □

Exercise 9.3 (Björk exc-4.2)**.** Let

$$Z(t) := \frac{1}{X_t}, \quad \text{where } X_t \text{ is an Itô process with differential} \quad dX(t) \;=\; \alpha X(t)dt + \sigma X(t)dW(t).$$

Prove that Z_t also is an Itô process by showing that this process has a differential of the form $dZ_t = \Phi_t dt + \Psi_t dW_t$ for suitable processes Φ_t and Ψ_t.
Hint: Apply the Itô formula with the function $f(x) = x^{-1}$. □

Exercise 9.4. Let $\alpha \in \mathbb{R}$. Compute $E[e^{\alpha W_t}]$ by doing the following.

(1). Let $Y_t := e^{\alpha W_t}$. Use Itô's formula with $f(x) := e^{\alpha x}$ to obtain

(**A**) $$Y_t \;= 1 + \tfrac{1}{2}\alpha^2 \int_0^t Y_u du + \alpha \int_0^t Y_u dW_u.$$

(2). Define $m(t) := E[Y_t]$. Apply Fubini to (**A**) and then differentiate $\frac{d}{dt}$ to show that $t \mapsto m(t)$ satisfies the ODE (ordinary differential equation)

(**B**) $$m'(t) \;= \tfrac{\alpha^2}{2} m(t), \qquad m(0) \;=\; 1.$$

(3). (**B**) shows that $m(t)$ satisfy a relation of the kind $y' = cy, y(0) = 1$. Convince yourself that this means that $y(x) = e^{cx}$ and show that $m(t) = e^{\alpha^2 t/2}$

(4). Now it is easy to compute $m(t) = E[e^{\alpha W_t}]$ and thus finish the problem. □

9.7 Blank Page after Ch.9

This page is intentionally left blank!

10 Black–Scholes Model Part I: The PDE

Introduction 10.1. This chapter is based on the finance application oriented aspects of GBM (geometric Brownian motion) that were briefly mentioned in Remark 9.8 about generalized GBM (p.195) and replicating portfolios for a contingent claim given in Chapter 8 (The Binomial Asset Model). There the dynamics of price of the risky asset developed as a binomial tree: price either was multiplied by an upward factor u with probability p_u, or it was multiplied by a downward factor d with probability p_d.

The Black–Scholes market model has in common with the Binomial Asset Model that there is a single risky asset (a stock) in addition to a single risk free asset (bond). In this chapter, we study the dynamics of the discounted asset price and build a hedging portfolio based on the idea that its value must match, at each point in time, the price of the contingent claim it replicates. From this condition we will derive a (deterministic) partial differential equation for the pricing function of the claim. □

10.1 Prologue: The Budget Equation in Continuous Time Markets

This subchapter closely follows [7] Björk, Thomas: Arbitrage Theory in Continuous Time.

To derive the continuous time budget equation of a self–financing portfolio at a fixed time t, we discretize the trading times and assume, for some small $h > 0$, that trading takes place only at

$$\dots,\ t-2h,\ t-h,\ t,\ t+h,\ t+2h,\ \dots$$

Then we examine what happens in the limit as $h \to 0$.

Since we will deal quite extensively with differences $X_{t+h} - X_t$, it is convenient to introduce some special notation for such differences.

Notation 10.1. We assume for the remainder of this subchapter 10.1 that $h > 0$ is fixed.

Given is an arbitrary real–valued stochastic process $X = X_t = X_t(\omega)$. We define

$$\Delta X_t \ :=\ \Delta X(t)\ :=\ \Delta X(t,\omega)\ :=\ X_{t+h} - X_t\,.$$

For a vector–valued process $\vec{Y}_t = \big(Y_t^{(1)},\dots,Y_t^{(n)}\big)$, we write

$$\Delta\vec{Y}_t \ :=\ \vec{Y}_{t+h} - \vec{Y}_t\,.$$

The Δ operation binds stronger than arithmetic operations. Thus,

$$\Delta X_t + Y_t \ =\ \big(\Delta X_t\big) + Y_t\,,\quad \Delta X_t\, Y_t \ =\ \big(\Delta X_t\big)\, Y_t\,,\quad \Delta\vec{X}_t \bullet \vec{Y}_t \ =\ \big(\Delta\vec{X}_t\big)\bullet\vec{Y}_t\,.$$

Here are some examples.

- $\Delta X_{t-h} = X_t - X_{t-h}$.
- $\big(\Delta\vec{Y}_t\big)^{(j)} = \Delta\big(Y_t^{(j)}\big) = Y_{t+h}^{(j)} - Y_t^{(j)}$. In other words, we take the Δ differences separately for each coordinate. □

Let us review portfolios in discrete time financial markets. We recall from Remark **??** that the holdings $\vec{H}_t$ were created at time $t-h$. They will be traded at time t for new holdings $\vec{H}_{t+h}$, which will be traded at time $t+h$ for new holdings $\vec{H}_{t+2h}$, which will be traded at time $t+2h$ ...

A self–financing portfolio is one which satisfies the budget equation

$$\sum_{j=0}^{n} H_t^{(j)} S_t^{(j)} = V_t^{\vec{H}} = \sum_{j=0}^{n} H_{t+h}^{(j)} S_t^{(j)} . \tag{10.1}$$

In other words, the previously established holdings $\vec{H}_t$, valued at time t, are worth the same amount $V_t^{\vec{H}}$ as the newly established holdings $\vec{H}_{t+h}$, also valued at time t. We apply $\bullet$ and Δ notation to (10.1) and obtain $\vec{H}_t \bullet \vec{S}_t = \vec{H}_{t+h} \bullet \vec{S}_t$. Hence, the budget equation becomes

$$\vec{S}_t \bullet \Delta\vec{H}_t = 0 . \tag{10.2}$$

We remember the following from calculus. The derivative

$$f'(x) = \frac{df}{dx}, \qquad \text{written in differential form as} \quad df(x) = f'(x)dx ,$$

was obtained from the difference quotient as a limit

$$\frac{df}{dx} = \lim_{h\to 0} \frac{f(x+h)-f(x)}{(x+h)-x} = \lim_{h\to 0} \frac{\Delta f(x)}{\Delta x} .$$

Thus, letting $h \to 0$ in (10.2) should give us the budget equation $\vec{S}_t \bullet d\vec{H}_t = 0$. But **this approach has a fatal flaw** and gives an incorrect result. To understand the nature of the problem, we examine the j-th term $S_t^{(j)}\, dH_t^{(j)}$ of $\vec{S}_t \bullet d\vec{H}_t = \sum\limits_{j=0}^{n} S_t^{(j)}\, dH_t^{(j)}$.

1. $S_t^{(j)}\, dH_t^{(j)}$ represents $\int_0^t S_u^{(j)}\, dH_u^{(j)}$, just as $Z_t\, dW_t$ represents $\int_0^t Z_u\, dW_u$.
2. The Itô integral $\int_0^t Z_u\, dW_u$ is a limit of $\sum_k Z_{t_k}\big(W_{t_{k+1}} - W_{t_k}\big)$, as $\max_k(t_{k+1}-t_k) \to 0$.
3. It is crucial that a forward difference $W_{t_{k+1}} - W_{t_k}$ of the integrator process W was taken: Neither t_{k+1} nor t_k is in the past of the integrands time, t_k. [42] Intuitively, this means that the value of the integrand must be known by the times t_k and t_{k+1} when the integrator values $W_{t_{k+1}} - W_{t_k}$ are used.
4. Likewise, $\int_0^t S_u^{(j)}\, dH_u^{(j)}$ is a limit of $\sum_k S_{t_k}^{(j)}\big(H_{t_{k+1}}^{(j)} - H_{t_k}^{(j)}\big)$, as $\max_k(t_{k+1}-t_k) \to 0$.
5. Again, forward differences $H_{t_{k+1}}^{(j)} - H_{t_k}^{(j)}$ of the integrator process $H^{(j)}$ must be taken.
6. The problem: The integrator value $H_t^{(j)}$ is the portfolio holding for the period $[t-h, t[$. It is established at time $t-h$, before the integrand, the asset price S_t is known.

Note that the problem goes away if we can work in **(4)** with $\sum_k S_{t_{k-1}}^{(j)}\big(H_{t_{k+1}}^{(j)} - H_{t_k}^{(j)}\big)$ instead of $\sum_k S_{t_k}^{(j)}\big(H_{t_{k+1}}^{(j)} - H_{t_k}^{(j)}\big)$, since $S_{t_{k-1}}^{(j)}$ is known at t_{k-1}, the time where $H_{t_k}^{(j)}$ is established.
We achieve this by subtracting and re–adding $\vec{S}_{t-h} \bullet \Delta\vec{H}_t$ to (10.2) as follows.

$$0 = \big(\vec{S}_t \bullet \Delta\vec{H}_t - \vec{S}_{t-h} \bullet \Delta\vec{H}_t\big) + \vec{S}_{t-h} \bullet \Delta\vec{H}_t = \Delta\vec{S}_{t-h} \bullet \Delta\vec{H}_t + \vec{S}_{t-h} \bullet \Delta\vec{H}_t \tag{10.3}$$

[42] For example, taking forward differences is necessary so that stochastic integrals with respect to Brownian motion are martingales.

Now we may take limits $h \to 0$ for $\vec{S}_{t-h} \bullet \Delta \vec{H}_t$, since $\Delta \vec{H}_t = \vec{H}_{t+h} - \vec{H}_t$, and both portfolio holdings are known at $t-h$. It follows from (10.3) that

$$d\vec{S}_t \bullet d\vec{H}_t \,+\, \vec{S}_t \bullet d\vec{H}_t \;=\; 0\,. \tag{10.4}$$

We fix a coordinate $0 \le j \le n$. By Itô's product rule,

$$d\big(H_t^{(j)} S_t^{(j)}\big) \;= H_t^{(j)}\, dS_t^{(j)} \,+\, \big(S_t^{(j)}\, dH_t^{(j)} \,+\, dS_t^{(j)}\, dH_t^{(j)}\big)\,. \tag{10.5}$$

Since, by (10.1), $V_t^{\vec{H}} \;=\; \sum\limits_{j=0}^{n} H_t^{(j)} S_t^{(j)} \;=\; \vec{S}_t \bullet \vec{H}_t$,

$$\begin{aligned} dV_t^{\vec{H}} \;&=\; \sum_{j=0}^{n} d\big(H_t^{(j)} S_t^{(j)}\big) \;\stackrel{(10.5)}{=}\; \sum_{j=0}^{n} H_t^{(j)}\, dS_t^{(j)} \,+\, \left(\sum_{j=0}^{n} S_t^{(j)}\, dH_t^{(j)} \,+\, \sum_{j=0}^{n} dS_t^{(j)}\, dH_t^{(j)}\right) \\ &= \vec{H}_t \bullet d\vec{S}_t \,+\, \big(\vec{S}_t \bullet d\vec{H}_t \,+\, d\vec{S}_t \bullet d\vec{H}_t\big) \;\stackrel{(10.4)}{=}\; \vec{H}_t \bullet d\vec{S}_t\,. \end{aligned}$$

Those observations are of a heuristic nature because taking the limit $h \to 0$ was involved to bridge the gap from discrete trading times to continuous trading times. Nevertheless, it suggests how to define the continuous time budget equation and give mathematical precision to Definition 7.5 of a self–financing portfolio (see p.149). for a continuous market portfolio $\vec{H}_t$.

The following definition also provides a solid mathematical foundation for Definition 7.7 on p.150 of an arbitrage portfolio, and for Definition 7.9 on p.151 of a hedging portfolio.

Definition 10.1 (Continuous time budget equation and self–financing portfolios).

(A.) The **budget equation** for a portfolio $\vec{H}_t$ in a continuous time financial market is

$$dV_t^{\vec{H}} \;=\; \sum_{j=0}^{n} H_t^{(j)}\, dS_t^{(j)} \;=\; \vec{H}_t \bullet d\vec{S}_t\,, \quad \text{for } 0 \le t \le T. \tag{10.6}$$

(B.) We call $\vec{H}_t$ a **self–financing portfolio strategy** aka **self–financing portfolio**, if it satisfies this budget equation. □

10.2 Formulation of the Black–Scholes Model

Notation 10.2. I will stay in this chapter close to SCF2 Chapter 4.5 (Black–Scholes–Merton Equation). I often will just copy the theorems and propositions presented there and refer to the text as far as the proofs are concerned.

I also will mostly use that book's notation and doing so make it easier for you to relate the material presented here to the SCF2 text even though I much prefer the notation of [7] Björk, Thomas: Arbitrage Theory in Continuous Time which I used in Chapter 8 (The Binomial Asset Model) of these lecture notes. The following table summarizes the most important differences.

Björk	**Shreve**	
S_t	S_t	price of the risky asset (stock, the underlying).
B_t	N/A	unit price of the riskless asset (money market account price).
$\vec{H}_t$	N/A	portfolio (# of shares) vector for all assets.
$x_t = H_t^B B_t$	N/A	dollar value of the riskless asset.
$y_t = H_t^S$	Δ_t	# of shares of the stock.
V_t	X_t	value process of the portfolio.
$\Pi_t(\mathcal{X})$	N/A	price process of a contingent claim $\mathcal{X}$.
N/A	$c(t,x)$	pricing function of a European call. $c(t, S_t)$ equals $\Pi_t(\mathcal{X})$.
N/A	$p(t,x)$	pricing function of a European put. $p(t, S_t)$ equals $\Pi_t(\mathcal{X})$.

The most likely exception to me trying to stick with SCF2 notation will occur with respect to portfolio holdings and values, but since only two assets are involved, including the bank account, I will use a modified Björk notation and write

- H_t^B for the number of bank account shares (with a money value of B_t dollars per share),
- V_t^B rather than $H_t^B B_t$ for the value (dollars) invested in the bank account,
- H_t^S (S = **S**tock) for the number of shares in the stock.
- either V_t or $V_t^{\vec{H}}$ for value of the portfolio $\vec{H}_t$.
- X_t and Y_t for x_t and y_t, since those are stochastic processes.

The portfolio value process thus will be written in any of the following ways.

(10.7) $$V_t^{\vec{H}} \;=\; V_t \;=\; H_t^B B_t \,+\, H_t^S S_t \;=\; V_t^B \,+\, H_t^S S_t \;=\; X_t \,+\, Y_t S_t\,.$$

Also note that $X_t = V_t^B$, the money value of the bank account holdings, satisfies

(10.8) $$X_t \;=\; V_t \,-\, Y_t S_t\,, \qquad \text{and} \qquad H_t^B \;=\; \frac{X_t}{B_t} \;=\; D_t X_t\,. \;\square$$

Definition 10.2 (Black–Scholes Market Model)**.**

The **Black–Scholes market model** consists of a time $T > 0$, a risk free asset (bond) with price process $B = B_t, 0 \leq t \leq T$, a risky asset (stock) with price process $S = S_t, 0 \leq t \leq T$, a simple contingent claim $\mathcal{X} = \Phi(S_T)$ with expiration date T, contract function $\Phi(x)$, and price process $\Pi_t(\mathcal{X})$, such that the following conditions hold.

(10.9) $$dB_t \;= rB_t\,dt;\;\; B_0 = 1;$$

(10.10) $$dS_t \;= \alpha S_t\,dt \,+\, \sigma S_t\,dW_t;\;\; S_0 \in [0,\infty[;\, \alpha,\sigma \in]0,\infty[\,,$$

(10.11) $$\mathcal{X} \;= \Phi(S_T) \quad \text{(simple contingent claim)},$$

- $c : [0,T] \times [0,\infty[\quad (t,x) \mapsto c(t,x)$ twice continuously differentiable such that

(10.12) $$\Pi_t(\mathcal{X}) \;= c(t, S_t) \quad \text{(price process of } \mathcal{X}\text{)}$$

- The market is efficient: No arbitrage portfolios. $\square$

Remark 10.1.

(1) $dB_t = rB_t dt;\ B_0 = 1$ is equivalent to $B_t = e^{rt}$, i.e., an account which pays continuously compounded interest at the constant and deterministic rate r per unit time.

(2) Formula (10.10) states that S_t is GBM with constant, instantaneous mean rate of return α and constant volatility σ. See Remark 9.8 on p.195. There are more general models (Definition 13.1 on p.237) in which the constants α and σ are replaced by measurable functions $\alpha(t,x), \sigma(t,x)$ of time. The price of the stock then is given by

$$dS_t \;=\; \alpha(t,S_t)S_t\,dt \;+\; \sigma(t,S_t)S_t\,dW_t;\quad S_0 \in [0,\infty[.$$

(3) The symbol c was chosen for the function $c(t,x)$ to remain in sync with the SCF2 text where only the example of a (European) c̲all is used when deriving the corresponding PDE. Note that this function must satisfy the terminal condition

$$c(T,S_T) \;=\; \Pi(T;\mathcal{X}) \;=\; \Phi(S_T). \tag{10.13}$$

(4) Smoothness (the existence of partial derivatives of any order) is not really necessary for $c(t,x)$. It suffices that this be a C^2 **function**, , i.e., all partial derivatives of order 2 exist and are continuous.

(5) Recall that Assumption 7.1 on p.150 includes that the market is free of arbitrage, in addition to other assumptions such as complete liquidity, no transaction costs and no bid–ask spread. □

10.3 Discounted Values of Option Price and Hedging Portfolio

Proposition 10.1.

The budget equation for a self–financing portfolio in a Black–Scholes market evolves according to the following dynamics.

$$dV_t = Y_t dS_t \;+\; rX_t\,dt \tag{10.14}$$

$$= rV_t dt \;+\; (\alpha - r)Y_t S_t dt \;+\; Y_t \sigma S_t\,dW_t\,. \tag{10.15}$$

PROOF: See SCF2, Chapter 4.5.1 (Evolution of Portfolio Value). ■

Remark 10.2. Formula (10.15) signifies that a portfolio value change dV_t is composed of

a. An average underlying rate of return r on the bond value $V_t - Y_t S_t$,

b. An average underlying rate of return $r + (\alpha - r) = \alpha$ on the stock investment in height of $Y_t S_t$. Since people will not take a greater risk investing in a stock than putting money in the bank we should expect that $\alpha \geq r$, thus $(\alpha - r)$ is a risk premium for investing in the stock.

c. A volatility term $Y_t \sigma S_t dW_t$. It is proportional to the size $Y_t \sigma S_t$ of the stock investment. □

Remark 10.3. We already mentioned that Formula (10.14) which asserts that $dV_t = Y_t dS_t + rX_t dt$, is the budget equation of a self–financing portfolio in the Black–Scholes market.

You obtain from it the discrete time analogue by replacing dV_t with $V_{n+1} - V_n$, replacing dS_t with $S_{n+1} - S_n$, and replacing dt with $(n+1) - n = 1$. Then

$$\begin{aligned} V_{n+1} - V_n &= Y_n S_{n+1} - Y_n S_n + r X_n \cdot 1 \\ &= Y_n S_{n+1} - Y_n S_n + r(V_n - Y_n S_n) \end{aligned}$$

Thus

$$\begin{aligned} V_{n+1} &= V_n + Y_n S_{n+1} - Y_n S_n + r V_n - r Y_n S_n \\ &= (1+r) V_n - (1+r) Y_n S_n + Y_n S_{n+1} \\ &= (1+r)(V_n - Y_n S_n) + Y_n S_{n+1} = (1+r) X_n + Y_n S_{n+1}\,, \end{aligned}$$

just as the budget equation demands it: The portfolio value at the new trading time must be the old bank account value X_n, increased by interest rX_n, plus the value of the stock holdings Y_n, valued at the new price S_{n+1} per unit, i.e., valued at $Y_n S_{n+1}$. □

Proposition 10.2.

Discounted stock price $e^{-rt}S_t$ and discounted portfolio value $e^{-rt}V_t$ satisfy

$$d\big(e^{-rt}S_t\big) = (\alpha - r)\, e^{-rt} S_t\, dt + \sigma e^{-rt} S_t\, dW_t\,, \tag{10.16}$$

$$\begin{aligned} d\big(e^{-rt}V_t\big) &= (\alpha - r) Y_t\, e^{-rt} S_t\, dt + \sigma Y_t e^{-rt} S_t\, dW_t \\ &= Y_t\, d\big(e^{-rt}S_t\big)\,. \end{aligned} \tag{10.17}$$

PROOF: See SCF2, Chapter 4.5.1 (Evolution of Portfolio Value). ■

Remark 10.4.

(a) It follows from (10.16), that discounting stock price has the following effect: Whereas S_t has a mean rate of return of α, it has dropped to $\alpha - r$ for $e^{-rt}S_t$.

(b) Formula (10.17) shows that change in the discounted portfolio value has nothing to do with a change in the bank account. It entirely depends on the change in the discounted stock price. □

We now investigate the ramifications of the existence of a deterministic function $c(t,x)$ in the definition 10.2 of the Black–Scholes Market Model such that $\Pi_t(\mathcal{X}) = c(t, S_t)$.

Proposition 10.3.

The price dynamics of the contingent claim are

$$dc\big(t,S_t\big) = \Big[c_t\big(t,S_t\big) + \alpha S_t\, c_x\big(t,S_t\big) + \frac{1}{2}\sigma^2 S_t^2\, c_{xx}\big(t,S_t\big)\Big]\, dt + \sigma S_t\, c_x\big(t,S_t\big)\, dW_t. \tag{10.18}$$

Those of the discounted option price $e^{-rt}c\big(t,S_t\big)$ are

(10.19)

$$\begin{aligned} d\big(e^{-rt}c(S_t)\big) &= e^{-rt}\Big[-rc\big(t,S_t\big) + c_t\big(t,S_t\big) + \alpha S_t\, c_x\big(t,S_t\big) + \frac{1}{2}\sigma^2 S_t^2\, c_{xx}\big(t,S_t\big)\Big]\, dt \\ &\quad + e^{-rt}\sigma S_t\, c_x\big(t,S_t\big)\, dW_t. \end{aligned}$$

PROOF: See SCF2, Chapter 4.5.2 (Evolution of Option Value). ■

10.4 The Pricing Principle in the Black–Scholes Market

According to the pricing principle (Theorem 7.1 on p.151) an arbitrage free price $\Pi_t(\mathcal{X}) = c(t, S_t)$ of the contingent claim $\mathcal{X}$ requires that a replicating portfolio with value process V_t satisfies

$$c(t, S_t) \;=\; V_t, \text{ for all trading times } t.$$

This is equivalent to $e^{-rt}V_t \;=\; e^{-rt}c(t, S_t)$ for all t. In terms of differentials:

$$\begin{aligned} d\big(e^{-rt}V_t\big) &= d\big(e^{-rt}c(t, S_t)\big) \text{ for all } t, \\ V_0 &= c\big(0, S_0\big) \end{aligned} \tag{10.20}$$

We apply (10.16) and (10.19) to the first part of (10.20). We cancel the factor e^{-rt} everywhere and omit the argument (t, S_t) of the function $c(t, x)$ and its derivatives $c_t(t, x), c_x(t, x), c_{xx}(t, x)$, and obtain

$$\begin{aligned} &Y_t\sigma S_t\, dW_t + Y_t(\alpha - r)S_t\, dt \\ &= \sigma S_t\, c_x\, dW_t + \Big[- rc \,+\, c_t \,+\, \alpha S_t\, c_x \,+\, \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx} \Big]\, dt. \end{aligned} \tag{10.21}$$

Since evolution with respect to dt is fundamentally different of that with respect to $dW t$ it is allowed to separately equate first the dW_t terms and then the dt terms of formula (10.21). We first equate the dW_t terms and obtain after canceling $\sigma e^{-rt}S_t$ the

delta–hedging rule:

$$Y_t \;= c_x\big(t, S_t\big) \text{ for all } t \in [0, T[. \tag{10.22}$$

At each time t prior to expiration, the number of shares Δ_t held by the hedging portfolio of the short option position is the delta of the option price $c(t, S_t)$ at that time.

Definition 10.3 (Delta (Greek))**.** Let $\mathcal{X}$ be a simple contingent claim in the Black–Scholes market, and let $(t, x) \mapsto c(t, x)$ be the twice continuously differentiable function which yields the price process $\Pi_t(\mathcal{X}) = c(t, S_t)$ [43] and thus, in particular, the contract function $\Phi(S_T) = c(T, S_T)$. We call the partial derivative of $c(t, x)$ with respect to stock price x,

$$\text{delta} \;:= \frac{\partial c}{\partial x}\,, \tag{10.23}$$

the **delta** of the claim. Delta is one of the so called **greeks** of the claim. □

We just proved that $Y_t = c_x(t, S_t)$. Equating the dt terms of formula (10.21) thus yields

$$c_x(\alpha - r)S_t \;=\; -rc \,+\, c_t \,+\, \alpha S_t\, c_x \,+\, \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx}.$$

[43] See Definition 10.2 of the Black–Scholes Market Model on p.205

We cancel the term $\alpha S_t c_x$ on both sides:

$$-rc_x S_t \;=\; -rc \,+\, c_t \,+\, \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx}.$$

We reorder those terms and obtain

$$rc \;=\; c_t \,+\, rc_x S_t \,+\, \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx}. \tag{10.24}$$

We bring back the arguments (t, S_t) and recall that the pricing principle asks that all equations we have encountered must hold for all t:

$$r\,c\big(t,S_t\big) \;=\; c_t\big(t,S_t\big) \,+\, rS_t\,c_x\big(t,S_t\big) \,+\, \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx}\big(t,S_t\big) \;\text{ for all } t \in [0,T[,$$

together with the expiration time condition $c(T, S_T) = \Phi(S_T)$ of formula (10.13).

We summarize our findings. The pricing principle lets us demand that the pricing function of a simple claim $\mathcal{X} = \Phi(S_T)$ be function $c(t, x)$ of time t and stock price x that solves the

Black–Scholes partial differential equation

$$c_t(t,x) + rx\,c_x\big(t,x\big) \,+\, \frac{1}{2}\,\sigma^2 x^2 c_{xx}(t,x) \;= r\,c\big(t,x\big),\; x \geq 0, \tag{10.25}$$

subject to the terminal condition

$$c(T,x) \;= \Phi(S_T). \tag{10.26}$$

The equations $V_t = c(t, S_t) = V_t^B + V_t^S$, $V_t^B = H_t^B e^r t = X_t$, $V_t^S = H_t^S S_t = Y_t S_t = c_x(t, S_t)S_t$, allow us to express the hedging portfolio for the claim $\mathcal{X}$ purely in terms of the pricing function $c(t, x)$ for the claim and the discount factor e^{-rt} as follows.

$$\vec{H}_t \;= (H_t^B, H_t^S) \;=\; \big(\, e^-rt\big[c(t,S_t) \,-\, c_x(t,S_t)S_t\big],\, c_x(t,S_t)\,\big)\,. \tag{10.27}$$

In other words, at time t this portfolio invests $c(t, S_t) \,-\, c_x(t, S_t)$ in the bank and holds $c_x(t, S_t)$ shares of the stock.

Remark 10.5. Observe that we only are concerned with stock price parameter $x > 0$ since $S_t > 0$ is a GBM. Thus, if we can prove that the solution $c(t, x)$ is continuous for all $0 \leq t \leq T$ satisfies the PDE just for $0 \leq t \leq T$ and $x \geq 0$ then we are fine, since continuity of $t \mapsto c(t, S_t)$ and $t \mapsto V_t$ for $0 \leq t \leq T$ implies that the hedge equation $V_t = c(t, S_t)$ extends from $0 \leq t < T$ to $t = T$, and the boundary condition $c(T, x) = \Phi(x)$ yields $V_T = \Phi(X_T)$.

To summarize, it is enough to show that the Black–Scholes PDE holds for all $x \geq 0$ and $t \in [0, T[$ □

10.5 The Black–Scholes PDE for a European Call

The Black–Scholes PDE (10.25) on p.209 is a purely deterministic PDE, and it can be solved by exclusively using tools from the theory of partial differential equations which do not rely on probability theory.

We need more knowledge of Itô calculus, in particular, the construction of martingale measures, before we will solve this PDE. Obviously probability theory plays a heavy role there. Here we simply present the solution for the special case of a European call, i.e., a simple contingent claim $\mathcal{X}$ with contract function

$$\Phi(x) = c(T,x) = (x-K)^+.$$

Remark 10.6. Here are two conditions specific to the European call.

a. In the case of a European call the solution of the Black–Scholes PDE must satisfy the following boundary condition for stock price $x=0$.

$$c(t,0) = 0 \text{ for all } t \in [0,T]. \tag{10.28}$$

This is true for the following reason. Formula (10.25) states that $y(t) := c(t,0)$ satisfies the ODE

$$y' = ry; \qquad \text{thus} \quad y(t) = \text{const} \cdot e^{rt}.$$

We obtain const by setting $t=0$: $y(0) = \text{const} \cdot 1$, i.e., const $= y(0) = c(0,0)$. Thus

$$c(t,0) = c(0,0)\, e^{rt} \quad \text{for all } 0 \le t \le T. \tag{A}$$

$K \ge 0$, thus $c(T,0) = \Phi(0) = (0-K)^+ = 0$. From (**A**): $0 = c(T,0) = c(0,0)e^{rT}$.
But expiration $T > 0$, thus $e^{rT} > 0$, thus $c(0,0) = 0$.
We use (**A**) once more: $c(0,0) = 0 \;\Rightarrow\; c(t,0 = 0 \cdot e^{rt} = 0$ for all t.
In summary: $c(t,0 = 0$ for all t.

B. This solution not only satisfies the **initial condition** $c(t,0) = 0$ for all t which we had deduced in Remark 10.6 above but also the growth condition

$$\lim_{x\to\infty} \left(c(t,x) - (x - e^{r(T-t)}K) \right) = 0 \text{ for all } t \in [0,T]. \tag{10.29}$$

Since $e^{r(T-t)}K$ is constant in x this condition implies that the value $c(t,x)$ of the call option grows at the same rate as x as $x \to \infty$. It will thus exceed the strike price K by a significant amount for large x and it is very likely that this will remain true as t approaches T. Since it is very unlikely for large x that $S_T - K < 0$, i.e.,

$$(S_T - K)^+ \neq S_t - K,$$

(the holder of the option will almost certainly be **in the money**, i.e., make a profit), it should not come as a surprise that the price for a European call approaches that of a claim with contract function $\Phi(x) = x - K$. You may recall from Definition 7.3 on p.146 that this was the contract function for a forward contract with strike price K. $\square$

Without proof for now:

Theorem 10.1. *The solution to the Black–Scholes partial differential equation* (10.25) *with terminal condition* (10.26)*, zero stock price condition* (10.28)*, and growth condition* (10.29) *is*

$$c(t,x) \;= xN\big(d_+(T-t,x)\big) \;-\; Ke^{-r(T-t)}N\big(d_-(T-t,x)\big),\;\; 0 \le t < T,\, x > 0, \tag{10.30}$$

where

$$d_\pm(\tau,x) \;= \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} \;+\; \left(r \pm \frac{\sigma^2}{2}\right)\tau\right], \tag{10.31}$$

and N is the cumulative standard normal distribution

$$N(y) \;= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{z^2}{2}}\,dz \;=\; \frac{1}{\sqrt{2\pi}}\int_{-y}^{\infty} e^{-\frac{z^2}{2}}\,dz. \tag{10.32}$$

PROOF: Will be given later: The entire subchapter 13.4 (Risk–Neutral Pricing of a European Call) is devoted to that proof. ■

Remark 10.7. We will sometimes write $\text{BSM}(\tau,x;K,r,\sigma)$ for $c(t,x)$ (where $\tau = T-t$, i.e., $t = T-\tau$).

We call $\text{BSM}(\tau,x;K,r,\sigma)$ the **Black–Scholes–Merton function**. Then (10.30) becomes

$$\text{BSM}(\tau,x;K,r,\sigma) \;= xN\big(d_+(\tau,x)\big) \;-\; Ke^{-r\,\tau}N\big(d_-(\tau,x)\big), \tag{10.33}$$

In this formula, τ and x denote the time to expiration and the current stock price, respectively. The parameters K, r, and σ are the strike price, the interest rate, and the stock volatility, respectively. □

Remark 10.8. There is various software to calculate the parameters for Black–Scholes contract functions Here are some links that were active as of April 16, 2021.

- **a.** Magnimetrics Excel implementation:
  https://magnimetrics.com/black-scholes-model-first-steps/
- **b.** Drexel U Finance calculator:
  https://www.math.drexel.edu/~pg/fin/VanillaCalculator.html
- **b.** EasyCalculation.com:
  https://www.easycalculation.com/statistics/black-scholes-mode.php □

Remark 10.9. Formula (10.30) does not define $c(t,x)$ when $t = T$ (because then $\tau = T - t = 0$ and this appears in the denominator in (10.31)), nor does it define $c(t,x)$ when $x = 0$ (because $\log x$ appears in (10.31)), and $\log 0$ is not a real number). However, (10.30) defines $c(t,x)$ in such a way that

$$\lim_{t\to T} c(t,x) \;=\; (x-K)^+ \quad \text{and} \quad \lim_{x\downarrow 0} c(t,x) \;=\; 0.$$

You will be asked to prove those claims in Exercise 4.9 of SCF2. □

10.6 The Greeks and Put–Call Parity

This chapter is largely a summary of SCF2 ch.4.5.5 and 4.5.6.

We assume for all of this chapter that we have a Black–Scholes market with interest rate r, instantaneous mean rate of return α, and volatility σ. All those are asumed to be constant. We further assume that $r \geq 0$ and $\sigma > 0$.

We denote by $F(t, x)$ the pricing function for a simple claim $\mathcal{X}$ with contract function $\Phi(x)$:

$$F(t, S_t) \;=\; \Pi_t(\mathcal{X}).$$

For people working in finance it often matters greatly how stable or volatile the function this pricing function is with respect to

1. changes in the price S_t of the underlying asset, i.e., changes in x,
2. changes in the interest rate r and the volatility σ.

Those changes are given by the derivatives of F. As far as derivatives with respect to r and σ are concerned we can examine F with respect to a variety of values of r and σ, i.e., we can think of F as a function

$$\widetilde{F} \;:\; (t, x, r, \sigma) \;\mapsto\; \widetilde{F}(t, x, r, \sigma).$$

So we really mean, e.g., $\frac{\partial \widetilde{F}}{\partial r}$ when we write $\frac{\partial F}{\partial r}$.

Definition 10.4 (Björk Def.9.4: Greeks)**.**

The following derivatives are part of what is known as the **Greeks** of the function F.

(10.34) $$\Delta = \frac{\partial F}{\partial x}$$ **delta**

(10.35) $$\Gamma = \frac{\partial^2 F}{\partial x^2}$$ **gamma**

(10.36) $$\rho = \frac{\partial F}{\partial r}$$ **rho**

(10.37) $$\Theta = \frac{\partial F}{\partial t}$$ **theta**

(10.38) $$\nu = \frac{\partial F}{\partial \sigma}$$ **vega** □

Remark 10.10. When reading SCF2 you might get the impression that those Greeks only exist for the pricing function $c(t, x)$ of a European call but that is not so.

- One can replace $c(t, x)$ with the pricing function $F(t, x)$ of any simple contingent claim in the Black–Scholes market where the underlying asset has a geometric Brownian motion as price process.
- In particular the Greeks exist for puts and forward contracts. □

Having stated that the Greeks are defined for all simple claims, we emphasize that the following formulas are specific for the pricing function $c(t, x)$ of a European call.

Proposition 10.4.

The following is true for the Greeks of a European call.

$$\text{(10.39)} \quad \textbf{delta} = c_x(t,x) = N\big(d_+(T-t,x)\big),$$

$$\text{(10.40)} \quad \textbf{gamma} = c_{xx}(t,x) = \frac{1}{\sigma x\sqrt{T-t}} N'\big(d_+(T-t,x)\big),$$

$$\text{(10.41)} \quad \textbf{theta} = c_t(t,x) = -rK\, e^{-r(T-t)} N\big(d_-(T-t,x)\big) - \frac{\sigma x}{2\sqrt{T-t}} N'\big(d_+(T-t,x)\big).$$

Because both the cumulative distribution function $N(x)$ of a standard normal random variable and its density $N'(x)$ are always strictly positive, Delta and Gamma are strictly positive, and Theta is strictly negative.

PROOF: Not given here. Those proofs are just an exercise in differentiation. ■

The delta hedging rule allows us to compute the replicating portfolio for a simple contract in the Black–Scholes market.

Proposition 10.5.

Let $\vec{H}_t = \big(H_t^B, H_t^S\big)$ be the hedging portfolio for a simple claim with pricing function $F(t,x)$. Thus H_t^B denotes the number of shares, i.e., dollars, in the bond, and H_t^S denotes the number of shares held in the stock. Take note that this one incident where we do not use SCF2 notation (he writes X_t for H_t^S)!
The following is true if it is known (or hypothesized) that $S_t = x$.

$$\text{(10.42)} \quad V_t^H = F(t,x),$$

$$\text{(10.43)} \quad e^{rt} H_t^B = F(t,x) - x \cdot F_x(t,x),$$

$$\text{(10.44)} \quad H_t^S = F_x(t,x).$$

PROOF: Formula (10.42) is just the pricing principle which says that the value of a replicating portfolio must always match the price of the option it replicates.

Formula (10.44) is the delta hedging rule which states the number of shares in the underlying stock is the derivative of the pricing function F with respect to stock price, evaluated at $x = S_t$.

Formula (10.43) just reflects the simple fact that, since the hedge $\vec{H}$ is self–financing, whatever is not invested in the underlying is in the bank.

$$e^{rt} H_t^B = V_t^B = V_t^H - S_t \cdot H_t^S, \qquad \text{i.e.,} \quad e^{rt} H_t^B = F(t,x) - x \cdot F_x(t,x)). \ ■$$

Remark 10.11. The hedging portfolio tells us what amounts must be invested in bank account and the underying by someone who holds a **short position in the claim**, i.e., someone who sold the claim at $t = 0$ and wants to be able to have the funds available at $t = T$ to deliver the derivative to the buyer.

In the specific case of a European call option, $H_t^S = c_x(t, S_t)$ is positive. See Proposition 10.4. We thus have the following.

- To hedge a short position in a European call, one needs to hold shares in the underlying and must borrow money from the bank to buy those shares.
- To hedge a long position in a Eoropean call, one must do the opposite, hold a position of minus $c_x(t, S_t$ shares of stock (i.e., have a short position in stock) and invest, assuming $S_t = x$, $V_t^B = c(t,x) - xc_x(t,x) = Ke^{-r(T-t)}N(d_-)$ in the money market account. See formula (10.39). □

Proposition 10.6.

Let $f(t, x)$ be the pricing function of a forward contract, i.e., simple claim with contract function $\Phi(x) = x - K$. [44] *Then*

$$f(t,x) = x - e^{-r(T-t)}K. \tag{10.45}$$

PROOF: Assume that this forward contract is sold at time zero for a price of $f\big(0, S_0\big) = S_0 - e^{-rT}K$. Then a bank loan of $e^{-rT}K$ will allow the seller to buy a share of the underlying. We look at the portfolio strategy $\vec{H} = (H^B, H^S$ which thus has been established at $t = 0$ by the short sale of the foward contract, i.e.,

$$H_0^B = -e^{-rT}K, \qquad H_0^S = 1.$$

We make this a **static hedge**, i.e., there will be no further trades until time of expiration T. Note though that the amount owed to the bank will increase due to compounded interest owed on the loan. At time t the interest factor will be e^{rt}. Thus portfolio and portfolio value are

$$\begin{aligned} H_t^B &= -H_0^B = -e^{-rT}K, \quad \text{and} \quad H_t^S = H_0^S = 1 \quad \text{for } 0 \leq t \leq T, \\ V_t &= -e^{rt}H_t^B + H_t^S S_t = -e^{-r(T-t)}K + 1 \cdot S_t = S_t - e^{-r(T-t)}K\,. \end{aligned}$$

In particular, at expiration time T, the portfolio value is

$$V_T^H = S_T - e^{-r(T-T)}K = S_T - K = \Phi(S_T)\,.$$

This static hedge thus is a replicating portfolio for the forward contract. It follows from the pricing principle that

$$f(t, S_t) = V_t^H = S_t - e^{-r(T-t)}K \ \text{ for all } 0 \leq t \leq T\,. \ \blacksquare$$

We associate with such a forward contract its fair strike price, if it had been set at time $0 \leq t \leq T$ and not at time zero. We call this the forward price For_t of the forward contract at time t:

Definition 10.5 (Forward price).

The **forward price** For_t of the underlying asset at time t is that value of K for which the forward contract has value zero at time t. □

Remark 10.12. By definition, For_t is that value K, for which $\Pi_t(\mathcal{X}) = 0$, i.e.,

$$0 = f(t, S_t) = S_t - e^{-r(T-t)}\text{For}_t\,.$$

This is the basis for the following.

A. The forward price satisfies the equation

$$S_t - e^{-r(T-t)}\,\mathrm{For}_t = 0. \tag{10.46}$$

B. Note that $\mathrm{For}_0 = K$. This should not come as a surprise. Both parties in the contract will agree at $t = 0$ to a strike price which does not give one of them an advantage over the other.

C. We solve formula (10.46) for For_t and obtain

$$\mathrm{For}_t = e^{r(T-t)} S_t. \tag{10.47}$$

D. Note that, for a given time t,

> the forward price For_t is NOT the price (or value) $f(t, S_t)$ of the forward contract. □

We recall from Definition 7.3 on p.146 that a European put with strike price K is a simple claim with contract function $\Phi(x) = (K - x)^+$. It is an option to sell, rather than buy, a share of the underlying at price K. Thus such an option generates a profit $K - S_T)$ if share price at expiration is below K, and it is worthless otherwise.

In the following we will write $p(t, x)$ rather than $F(t, x)$ for the price of a European put option.

We relate puts and calls by mean of the following simple identity.

Lemma 10.1. *For any real number α,*

$$\alpha = \alpha^+ - (-\alpha)^+. \tag{10.48}$$

PROOF:

$$\textbf{Case 1}: \ \alpha \geq 0 \Rightarrow \alpha^+ = \alpha,\ (-\alpha)^+ = 0 \Rightarrow \alpha^+ - (-\alpha)^+ = \alpha - 0 = \alpha.$$
$$\textbf{Case 2}: \ \alpha < 0 \Rightarrow \alpha^+ = 0,\ (-\alpha)^+ = -\alpha \Rightarrow \alpha^+ - (-\alpha)^+ = 0 - (-\alpha) = \alpha. \ \blacksquare$$

Corollary 10.1.

$$f\big(T, S_T\big) = S_T - K = (S_T - K)^+ - (K - S_T)^+ = c\big(T, S_T\big) - p\big(T, S_T\big).$$

> *the contract function of a forward contract with strike price K coincides with that of a portfolio that is long one European call and short one European put.*

PROOF: This is an immediate consequence of Lemma 10.1. ■

Proposition 10.7 (Put–call parity)**.** *We write, for one and the same strike price K,*

- *$c(t, x)$ for the pricing function of a European call,*
- *$p(t, x)$ for the pricing function of a European put,*
- *$f(t, x)$ for the pricing function of a forward contract.*

Then the following formula is satisfied:

Put–call parity:

(10.49) $$f(t,x) = c(t,x) - p(t,x), \quad \textit{for all } x \geq 0,\ 0 \leq t \leq T.$$

PROOF: We apply the pricing principle to the formula $p(T,S_T) = c(T,S_T) - f(T,S_T)$. This is valid according to Corollary 10.1. We obtain

$$p(t,x) = c(t,x) - f(t,x), \quad \text{for all } x \geq 0,\ 0 \leq t \leq T. \blacksquare$$

Proposition 10.8.

The pricing function $p(t,x)$ of a European put with strike price K satisfies

(10.50) $$\begin{aligned} p(t,x) &= x\big(N(d_+(T-t,x)) - 1\big) - Ke^{-r(T-t)}\big(N(d_-(T-t,x)) - 1\big) \\ &= Ke^{-r(T-t)}N\big(-d_-(T-t,x)\big) - x\big(N\big(-d_+(T-t,x)\big)\big), \end{aligned}$$

PROOF: We abbreviate $\tau = T - t,\ N(d_+) = N\big(d_+(T-t,x),\ N(d_-) = N\big(d_-(T-t,x)\big)$.
Put–call parity yields $f(t,x) = c(t,x) - p(t,x)$, thus $p(t,x) = c(t,x) - f(t,x)$.
The BSM formula yields $f(t,x) = xN(d_+) - Ke^{-r\tau}N(d_-)$. Thus,

$$\begin{aligned} p(t,x) &= xN(d_+) - Ke^{-r\tau}N(d_-) - \big(x - e^{-r\tau}K\big) \\ &= x\big(N(d_+) - 1\big) + Ke^{-r\tau}\big(1 - N(d_-)\big) \\ &= x\big(N(d_+) - 1\big) - Ke^{-r\tau}\big(N(d_-) - 1\big) \end{aligned}$$

This proves the first equation of (10.50).
Symmetry of the normal density yields $N(-\alpha) = 1 - N(\alpha)$ for any $\alpha \in \mathbb{R}$. Thus,

$$\begin{aligned} N(d_+) - 1 &= -\big(1 - N(d_+)\big) = -N(-d_+)\,, \\ N(d_-) - 1 &= -\big(1 - N(d_-)\big) = -N(-d_-)\,. \end{aligned}$$

We substitute those expressions into the already proven first equation of (10.50) and obtain the second equation. $\blacksquare$

10.7 American Call Options

Recall the following from Definition 7.3 on p.146.

- An **American call** option is a contract written at some time t_0. It specifies that, at any time up to the time of expiration $T > t_0$, the holder of this option has the right, but not the obligation, to buy a share of an underlying security stock for the price of K (strike price).
- An **American put** option is a contract written at some time t_0. It specifies that, at any time up to the time of expiration $T > t_0$, the holder of this option has the right, but not the obligation, to sell a share of an underlying security for the price of K (strike price).

Let $\mathcal{X}$ denote an American call or an American put. The freedom of the holder of such an American option to exercise it at any time τ between the present time t and the time of expiration T obviously implies the following. Its value $\Pi_t(\mathcal{X})$ is at least as big as that of the corresponding European option. How big? This is a complicated question since τ need not be deterministic. Rather, we assume that τ can be any random time

$$\tau = \tau(\omega),$$

which satisfies the following. Each σ–algebra $\mathfrak{F}_t$ contains enough information to determine whether τ has already happened at time t. This is expressed by the condition

$$\{\tau \leq t\} \in \mathfrak{F}_t, \quad \text{whenever } 0 \leq t \leq T.$$

Such a random time is called a **stopping time** (for the filtration $\left(\mathfrak{F}_t\right)_t$).

You will find more information in SCF2 Chapter 8 (American Derivative Securities). For us this material is outside the scope of our course. However, an answer can be obtained with elementary reasoning in the case of an American call option.

We assume the following.

- **(a)** A risk free asset with a constant interest rate $r > 0$.
- **(b)** A stock which pays no dividends and has price dynamics $dS_t = \alpha S_t\,dt + \sigma S_t\,dW_t$, where $\alpha, \sigma > 0$ are constant.
- **(c)** No arbitrage.

Compare the above market assumptions to those of Definition 10.2 (Black–Scholes Market Model) on p.205.

Lemma 10.2. *Under the assumptions* ***(a)–(c)*** *we have the following for the price function $c(t,x)$ of a European call with expiration date T and strike price K.*

$$c(t, S_t) \geq S_t - Ke^{-r(T-t)}. \tag{10.51}$$

PROOF:

Let C_t be the value at time t of a portfolio which consists of one European call option. Then

$$C_t = c(t, S_t), \quad \text{thus,} \quad C_T = c(T, S_T) = (S_T - K)^+. \tag{A}$$

Let B_t be the value at time t of a portfolio which consists of one share of the stock and a bank loan in height of K, due at time T. Today we only need the discounted value $e^{-r(T-t)}K$ to pay back that loan at time K. it follows that

$$B_t = S_t - e^{-r(T-t)}K, \quad \text{thus,} \quad B_T = S_T - K. \tag{B}$$

Since $\alpha^+ \geq \alpha$ for all $\alpha \in \mathbb{R}$, we obtain $C_T \geq B_T$. We employ risk–neutral validation to reason as follows.

$$\begin{aligned} C_T \geq B_T &\Rightarrow e^{-r(T-t)}C_T \geq e^{-r(T-t)}B_T \\ &\Rightarrow C_t = \widetilde{E}[e^{-r(T-t)}C_T \mid \mathfrak{F}_t] \geq \widetilde{E}[e^{-r(T-t)}B_T \mid \mathfrak{F}_t] = B_t\,. \end{aligned}$$

We use **(A)** and **(B)** to conclude that $c(t, S_t) \geq S_t - e^{-r(T-t)}K$. ■

Proposition 10.9.

> *Under the assumptions **(a)**–**(c)** we have the following.*
> *The optimal (stopping) time τ to exercise an American call option on that stock in **(b)** with expiration time T and strike price $K > 0$, is $\tau = T$. Accordingly, the price $\Pi_t(\mathcal{X})$ of that option equals the price $c(t, S_t)$ of the corresponding European call option.*

PROOF: Let $0 \le t \le T$. Then

(A) $\Pi_t(\mathcal{X}) \ge c(t, S_t)$, since exercising the American call at T guarantees a profit of $c(T, S_T)$.

(B) $c(t, S_t) \ge S_t - e^{-r(T-t)}K$, according to Lemma 10.2.

(C) $S_t - e^{-r(T-t)}K > S_t - K$, for $0 \le t < T$, since $0 < e^{-r(T-t)} < 1$.

It follows from **(A)**, **(B)**, **(C)**, that

$$\Pi_t(\mathcal{X}) > S_t - K \quad \text{for } 0 \le t < T.$$

$S_t - K$ is the profit we stand to make if we exercise the option now [45] The larger amount of $\Pi_t(\mathcal{X})$ is what we make if we sell the option to another party, or what we expect to make under risk–neutral validation, if we hold on to the option until expiration. Either way, selling the call before expiration is not an optimal strategy. ■

10.8 Miscellaneous Notes About Some Definitions in Finance

In this chapter we list some financial terms that are mentioned in SCF2 without ever having been formally defined. It will be continually in flow and its references thus are subject to change in newer editions of these lecture notes.

Remark 10.13.

The following is based on the Investopedia link http://www.math.fsu.edu/~pkirby/mad2104/SlideShow/s2_1.pdf (Long Position vs. Short Position: What's the Difference?).

SCF2 will deal a ot with hedges of short and long positions. Here is my understanding:

> **(a)** A **"(short option) hedging portfolio"** is a portfolio $\vec{h} = (h^B, h^S)$ meant to hedge a short position in the (call) option. Note that I am **short an <u>option</u>** and **NOT** a share of the underlying: I have sold such an option and now use that portfolio to hedge that sale, i.e., $V_t^{\vec{h}}(\omega) = c\big(t, S_t(\omega)\big)$.
>
> **(b)** A **"long position in a call option"** is one where I have **bought** such an option, and I now want to create a portfolio $\vec{h} = (h^B, h^S)$ to hedge this long position. Note that I am hedging the **purchase of an <u>option</u>** and **NOT** of a share of the underlying, i.e., $V_t^{\vec{h}}(\omega) = -c\big(t, S_t(\omega)\big)$. □

10.9 Exercises for Ch.10

None at this time!

[45] Actually we stand to lose $K - S_t$ if $S_t < K$ and we are crazy enough to exercise the call anyway.

11 Multidimensional Stochastic Calculus

We generalize in this chapter the results of Chapter 9 (One dimensional Stochastic Calculus)

This chapter is very sketchy as far as proofs are concerned since the material follows extremely closely that of SCF2 Chapter 4.6.

11.1 Multidimensional Brownian Motion

Definition 11.1 (Multidimensional Brownian Motion)**.** Given are a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ and $d \in \mathbb{N}$.

A d**–dimensional Brownian motion** is a vector–valued stochastic process

$$\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)} \dots, W_t^{(d)}\big)$$

with the following properties.

(1) Each $W_t^{(j)}$ is a one dimensional Brownian motion.

(2) If $i \neq j$, then the processes $W_t^{(i)}$ and $W_t^{(j)}$ are independent, i.e., the σ–algebras $\sigma\big(W_t^{(i)} : t \geq 0\big)$ and $\sigma\big(W_t^{(j)} : t \geq 0\big)$ are independent.

(3) The process $\vec{W}_t$ is $\mathfrak{F}_t$–adapted, i.e., the random vector $\vec{W}_t$ is $\mathfrak{F}_t$–measurable for each $t \geq 0$.

(4) Future increments are independent of the past: If $t \geq 0$ and $h > 0$, then the vector $\vec{W}_{t+h} - \vec{W}_t$ is independent of $\mathfrak{F}_t$. $\square$

Remark 11.1. Since $W^{(j)}$ is a Brownian motion for each $j = 1, \dots, d$, all results derived for Brownian motion apply to each one of those coordinate processes. In particular,

(1) $[W^{(j)}, W^{(j)}]_t = t$,

(2) $dW_t^{(j)} dt = dt\, W_t^{(j)} = 0$ and $dW_t^{(j)} dW_t^{(j)} = t$, $\square$

Definition 11.2 (Cross variation). ⋆

Given are two adapted processes X_t and Y_t on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Let $T > 0$ and $\Pi := 0 = t_0 < t_1 < \cdots < t_k = T$ a partition of $[0, T]$. The random variable

$$C_\Pi[X, Y]_T := \sum_{k=0}^{n-1} (X_{t_{k+1}} - X_{t_k})(Y_{t_{k+1}} - Y_{t_k})$$

is called the **sampled cross variation** of X and Y on $[0, T]$ with respect to Π.
If there is a stochastic process $Z = Z_t$ such that

$$\lim_{\|\Pi\| \to 0} E\left[(C_\Pi[X, Y]_T - Z_T)^2\right] = 0$$

for all $T > 0$ then we write $[X, Y]_t$ for Z_t, and we refer to the process $[X, Y]_t$ the **cross variation** of X and Y. □

Remark 11.2. Note that if $X = Y$ then the process $[X, X]_t$ is the quadratic variation of X. □

Theorem 11.1.

Let $\vec{W}_t = \left(W_t^{(1)}, W_t^{(2)} \ldots, W_t^{(d)}\right)$ be a d–dimensional Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ ($d \in \mathbb{N}$). Let i and j be two integers such that $1 \leq i < j \leq d$. Then

$$\left[W^{(i)}, W^{(j)}\right]_t = 0.$$

PROOF: See SCF2 ch.4.6.1. ■

Theorem 11.2.

Let $\vec{W}_t = \left(W_t^{(1)}, W_t^{(2)} \ldots, W_t^{(d)}\right)$ be a d–dimensional Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ ($d \in \mathbb{N}$). Let i and j be two integers such that $1 \leq i, j \leq d$ and $i \neq j$. Then

$$dW^{(i)}\, dW^{(j)} = 0.$$

PROOF: This can be shown with help of Theorem 11.1 on p.220. See SCF2 ch.4.6. for details. ■

11.2 The Multidimensional Itô Formula

One can generalize The Itô formula which computes the differential $f(t, X_t)$, to processes X_t which are driven by a d–dimensional Brownian motion in the sense of the next definition.

Definition 11.3. ⋆ Let $\vec{W}_t = \left(W_t^{(1)}, W_t^{(2)} \ldots, W_t^{(d)}\right)$ be a d–dimensional Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ ($d \in \mathbb{N}$).

We call a process X_t an **Itô process driven by** $\vec{W}$, if its dynamics are

$$(11.1)\qquad \begin{aligned} dX_t &= \Theta_t\,dt + \sum_{j=1}^{d} \sigma_j(t)\,dW_t^{(j)} = \Theta_t\,dt + \sigma_1(t)\,dW_t^{(1)} + \cdots + \sigma_d(t)\,dW_t^{(d)},\\ X_0 &= x, \end{aligned}$$

for suitable adapted and sufficiently integrable processes Θ_t and $\vec{\sigma}(t) = \big(\sigma_1(t)\ldots,\sigma_d(t)\big)$.

In integrated form (11.1) is equivalent to

$$(11.2)\qquad X_t = x + \int_0^t \Theta_u\,du + \sum_{j=1}^{d}\int_0^t \sigma_j(u)\,dW_u^{(j)}.\ \square$$

All this can be written more compactly if we extend the "bullet notation" $\vec{x}\bullet\vec{y}$ from vectors to differentials and integrals as follows.

Notation 11.1. Let $n\in\mathbb{N}$. If $\vec{\Gamma}_t = \big(\Gamma_t^{(1)},\ldots,\Gamma_t^{(n)}\big)$ and $\vec{A}_t = \big(A_t^{(1)},\ldots,A_t^{(n)}\big)$ are vector valued stochastic processes for which the expressions $\int\limits_0^t \Gamma_u^{(j)}dA_u^{(j)}$ exist, then we define

$$(11.3)\qquad \begin{aligned} \vec{\Gamma}_t\bullet d\vec{A}_t &:= \sum_{j=1}^{n}\Gamma_t^{(j)}\,dA_t^{(j)},\\ \int_0^t \vec{\Gamma}_u\bullet d\vec{A}_u &:= \sum_{j=1}^{n}\int_0^t \Gamma_u^{(j)}\,dA_u^{(j)},\ \square \end{aligned}$$

With this notation we can rewrite (11.1) and (11.2) as follows.

$$\begin{aligned} dX_t &= \Theta_t\,dt + \vec{\sigma}(t)\bullet d\vec{W}_t;\quad X_0 = x,\\ X_t &= x + \int_0^t \Theta_u\,du + \int_0^t \vec{\sigma}(u)\bullet d\vec{W}_u.\ \square \end{aligned}$$

Remark 11.3. It should be mentioned that Itô's Lemma not only generalizes to d–dimensional Brownian motions for $d>2$ but also to functions

$$f(t,\vec{x}) = f(t,x_1,x_2,\ldots,x_n)$$

in which each dummy argument x_k can be replaced by an Itô process

$$\begin{aligned} dX_t^{(k)} &= \Theta_t^{(k)}\,dt + \sum_{j=0}^{d}\sigma_{kj}(t)\,dW_t^{(j)};\\ X_0^{(k)} &= x_0^{(k)}. \end{aligned}$$

We will not strive for such generality. Instead, we follow SCF2 and limit ourselves to $d=n=2$. Thus there will be two Itô processes, each one driven by a two dimensional Brownian motion. $\square$

Notation 11.2. From now on we assume that $\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)}\big)$ is a two dimensional Brownian motion and that X_t and Y_t are the following Itô processes, driven by $\vec{W}_t$.

$$\begin{aligned} dX_t &= \Theta_1(t)\,dt + \sigma_{11}(t)\,dW_t^{(1)} + \sigma_{12}(t)\,dW_t^{(2)}, \\ dY_t &= \Theta_2(t)\,dt + \sigma_{21}(t)\,dW_t^{(1)} + \sigma_{22}(t)\,dW_t^{(2)}. \end{aligned} \tag{11.4}$$

The integrands $\Theta_i(u)$ and $\sigma_{ij}(u)$ are adapted processes. We integrate and get

$$\begin{aligned} X_t &= x_0 + \int_0^t \Theta_1(u)\,du + \int_0^t \sigma_{11}(u)\,dW_u^{(1)} + \int_0^t \sigma_{12}(u)\,dW_u^{(2)}, \\ Y_t &= y_0 + \int_0^t \Theta_2(u)\,du + \int_0^t \sigma_{21}(u)\,dW_u^{(1)} + \int_0^t \sigma_{22}(u)\,dW_u^{(2)}. \ \square \end{aligned} \tag{11.5}$$

Theorem 11.3. *The multiplication rules for the multidimensional Itô calculus are*

- $dt\,dt = 0$,
- $dt\,dW_t^{(i)} = dW_t^{(i)}\,dt = 0$,
- $dW_t^{(i)}\,dW_t^{(i)} = t$,
- $dW_t^{(i)}\,dW_t^{(j)} = 0$ *for* $i \neq j$.

PROOF: This follows from the one dimensional case (see Remark 6.9 on p.135), together with Theorem 11.1 on p.220. ■

Remark 11.4. The multiplication tables make computation of the differential $dX_t dY_t$ of two Itô processes X_t and Y_t a trivial affair. For example, if those processes are given by (11.4), then

$$\begin{aligned} dX_t\,dX_t &= \big[d\big(\Theta_1(t)\,dt + \sigma_{11}(t)\,dW_t^{(1)} + \sigma_{12}(t)\,dW_t^{(2)}\big)\big]^2 \\ &= \Theta_1(t)^2 dt\,dt + \Theta_1(t)dt\,\sigma_{11}(t)\,dW_t^{(1)} + \Theta_1(t)dt\,\sigma_{12}(t)\,dW_t^{(2)} \\ &\quad + \cdots + \sigma_{12}(t)^2\,dW_t^{(2)}\,dW_t^{(2)}\,. \end{aligned}$$

Only two of those nine terms survice, those with differentials
$dW_t^{(1)}\,dW_t^{(1)} = dt$ and $dW_t^{(2)}\,dW_t^{(2)} = dt$. Thus

$$dX_t\,dX_t = \sigma_{11}(t)^2\,dt + \sigma_{12}(t)^2\,dt = \big(\sigma_{11}(t)^2 + \sigma_{12}(t)^2\big)\,dt,$$

and similarly,

$$dY_t\,dY_t = \sigma_{21}(t)^2\,dt + \sigma_{22}(t)^2\,dt = \big(\sigma_{21}(t)^2 + \sigma_{22}(t)^2\big)\,dt.$$

Further,

$$\begin{aligned} dX_t\,dY_t &= \Theta_1(t)\Theta_2(t)dt\,dt + \Theta_1(t)dt\,\sigma_{21}(t)\,dW_t^{(1)} + \Theta_1(t)dt\,\sigma_{22}(t)\,dW_t^{(2)} \\ &\quad + \cdots + \sigma_{12}(t)\sigma_{22}(t)\,dW_t^{(2)}\,dW_t^{(2)} \end{aligned}$$

Again only the two terms with differentials $dW_t^{(1)}\,dW_t^{(1)}$ and $dW_t^{(2)}\,dW_t^{(2)}$ are not zero. Thus,

$$dX_t\,dY_t = \sigma_{11}(t)\sigma_{21}(t)\,dt + \sigma_{12}\sigma_{22}\,dt. \ \square$$

Here is the Itô formula for a sufficiently smooth function $f(t,x,y)$ of time t and two more parameters which will accept two Itô processes driven by a two dimensional Brownian motion. This is SCF2 Theorem 4.6.2

Theorem 11.4 (Two dimensional Itô formula)**.**

Let $f(t,x,y)$ be a function whose partial derivatives $f_t, f_x, f_y, f_{xx}, f_{xy}, f_{yx}$, and f_{yy} exist and are continuous. Let X_t and Y_t be Itô processes driven by a two dimensional Brownian motion. The process $(t,\omega) \mapsto f\big(t, X_t(\omega), Y_t(\omega)\big)$ then has the dynamics

$$\begin{aligned} df\big(t,X_t,Y_t\big) \;&=\; f_t\big(t,X_t,Y_t\big)\,dt \,+\, f_x\big(t,X_t,Y_t\big)\,dX_t \,+\, f_y\big(t,X_t,Y_t\big)\,dY_t \\ &+\, \frac{1}{2}\,f_{xx}\big(t,X_t,Y_t\big)\,dX_tdX_t \,+\, f_{xy}\big(t,X_t,Y_t\big)\,dX_tdY_t \\ &+\, \frac{1}{2}\,f_{yy}\big(t,X_t,Y_t\big)\,dY_tdY_t. \end{aligned} \tag{11.6}$$

PROOF: Omitted, but we mention that the continuity of f_{xy}, f_{yx} gives us $f_{xy} = f_{yx}$, That fact together with $dX_tdY_t = dY_tdX_t$ is the reason that $\frac{1}{2}f_{xy}\big(t,X_t,Y_t\big)dX_tdY_t \,+\, \frac{1}{2}f_{yx}\big(t,X_t,Y_t\big)dX_tdY_t$ can be replaced by $f_{xy}\big(t,X_t,Y_t\big)dX_tdY_t$ instead of ■

Remark 11.5. We use for the differentials dX_t, dY_t, dX_tdX_t, dY_tdY_t and dX_tdY_t, the expressions found in Notations 11.2 and Remark 11.4. If we express the Itô formula with integrals rather than differentials, we obtain

$$\begin{aligned} &f\big(t,X_t,Y_t\big) \,-\, f\big(0,X_0,Y_0\big) \\ =\;& \int_0^t \big[\sigma_{11}(u)\,f_x\big(u,X_u,Y_u\big) \,+\, \sigma_{21}(u)\,f_y\big(u,X_u,Y_u\big)\big]\,dW_1(u) \\ &+ \int_0^t \big[\sigma_{12}(u)\,f_x\big(u,X_u,Y_u\big) \,+\, \sigma_{22}(u)\,f_y\big(u,X_u,Y_u\big)\big]\,dW_2(u) \\ &+ \int_0^t \Big[f_t\big(u,X_u,Y_u\big) \,+\, \Theta_1(u)\,f_x\big(u,X_u,Y_u\big) \,+\, \Theta_2(u)\,f_y\big(u,X_u,Y_u\big) \\ &\quad + \frac{1}{2}\big(\sigma_{11}^2(u) + \sigma_{12}^2(u)\big)\,f_{xx}\big(u,X_u,Y_u\big) \\ &\quad + \big(\sigma_{11}(u)\sigma_{21}(u) \,+\, \sigma_{12}(u)\sigma_{22}(u)\big)\,f_{xy}\big(u,X_u,Y_u\big) \\ &\quad + \frac{1}{2}\big(\sigma_{21}^2(u) + \sigma_{22}^2(u)\big)\,f_{yy}\big(u,X_u,Y_u\big)\Big]\,du \end{aligned} \tag{11.7}$$

You probably agree that this version of the Itô formula is much harder to remember and more cumbersome to use than (11.6). Here is the other extreme, with all arguments of the tunction $f(t,x,y)$ and its partial derivatives omitted.

$$\begin{aligned} df\big(t,X,Y\big) \;&=\; f_t\,dt \,+\, f_x\,dX \,+\, f_y\,dY \\ &+\, \frac{1}{2}\,f_{xx}\,dX_t\,dX_t \,+\, f_{xy}\,dX_t\,dY_t \,+\, \frac{1}{2}\,f_{yy}\,dY_t\,dY_t. \end{aligned} \tag{11.8}$$ □

The following is an extremely useful consequence of the multidimensional Itô formula.

Corollary 11.1 (Itô product rule)**.** *If X_t and Y_t are two Itô processes then*

$$d\big(X_t\,Y_t\big) \;= X_t\,dY_t \;+\; Y_t\,dX_t + dX_t\,dY_t. \tag{11.9}$$

PROOF: We apply formula (11.8) with $f(t,x,y) = xy$. Then $f_t = 0$, $f_x = y$, $f_y = x$, $f_{xx} = 0$, $f_{xy} = 1$, and $f_{yy} = 0$. The corollary follows easily. ■

Proposition 11.1.

Let $W_t^{(1)}, \dots, W_t^{(m)}$ be a collection of n onedimensional Brownian motions. No assumption is made that they are the coordinate processes of a multidimensional Brownian motion or that $W^{(i)} \neq W^{(j)}$ for $i \neq j$. Let X and Y be Itô processes with differentials

$$dX_t \;=\; \sum_{i=1}^m \big(\Delta_t^{(i)}\,dW_t^{(i)} \;+\; \Theta_t^{(i)}\,dt\big)\,; \qquad dY_t \;=\; \sum_{j=1}^n \Psi_t^{(j)}\,dt\,,$$

where $\Delta_t^{(i)}, \Theta_t^{(i)}$ and $\Psi_t^{(j)}$ are suitable adapted processes. Then $(dX_t)(dY_t) = 0$.

The proof is left as exercise 11.1 (see p.227). ■

Corollary 11.2. *Let X_t and Y_t be Itô processes such that dY_t is free of Brownian motion differentials, i.e., $dY_t \;=\; \sum\limits_{j=1}^n \Psi_t^{(j)}\,dt$ for suitable adapted processes $\Psi_t^{(j)}$. Then $d\big(X_t\,Y_t\big) \;=\; X_t\,dY_t \;+\; Y_t\,dX_t$.*

PROOF:

It follows from Proposition 11.1 that $\big(dX_t\big)\big(dY_t\big) = 0$. By Itô's product rule,

$$d\big(X_t\,Y_t\big) \;=\; X_t\,dY_t \;+\; Y_t\,dX_t \;+\; \big(dX_t\big)\big(dY_t\big) \;=\; X_t\,dY_t \;+\; Y_t\,dX_t\,. \;\blacksquare$$

11.3 Lévy's Characterization of Brownian Motion

Brownian motion W_t is characterized by the following.

- W_t is an $\mathfrak{F}_t$–martingale,
- $W_0 = 0$ a.s.,
- $t \mapsto W_t(\omega)$ is continuous a.s.,
- W_t has quadratic variation $[W,W]_t = t$ a.s.

A theorem by the french mathematician Paul Pierre Lévy (1886–1971) shows that a stochastic process M_t with those properties is in fact a Brownian motion, i.e., those properties guarantee that future increments $W_{t+h} - W_t$ are independent of $\mathfrak{F}_t$ and they have a normal distribution with mean zero and variance h.

d–dimensional Brownian motion $\vec{W}_t$ is characterized by the following.

- each coordinate $W_t^{(j)}$ is a (one dimensional) Brownian motion,
- Different coordinate processes $W^{(i)}$ and $W^{(j)}$ are independent, and they have cross variation zero.

The multidimensional version of Lévy's theorem proves that the reverse is true. Any process $\vec{M}_t$ with those two properties is a d–dimensional Brownian motion.

First, we state the one dimensional version. This is SCF2 Theorem 4.6.4

Theorem 11.5 (Lévy's characterization of one dimensional Brownian Motion)**.**

let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ be a filtered probability space. Assume that the process $M_t, t \geq 0$, satisfies

• $M_0 = 0$, • *M_t has continuous paths,* • *M_t is an $\mathfrak{F}_t$–martingale,* • $[M, M]_t = t$ *for all* $t \geq 0$.

Then M_t is an $\mathfrak{F}_t$–Brownian motion.

PROOF: ⋆ An outline of the proof can be found in SCF2. We summarize the major steps.

(1) The following can be defined and proven with a continuous martingale M_t such that $M_0 = 0$ in place of a Brownian motion W_t. One can define
• Itô integrals $\int_0^t Z_u dM_u$ which adhere to the multiplication rules

$$dt\,dt \;=\; dt\,dM_t \;=\; dM_t\,dt \;=\; 0, \quad dM_t\,dM_t \;=\; t\,.$$

The last rule is obtained from the assumption $[M, M]_t = t$.
• Itô processes $X_t = X_0 + \int_0^t \Delta_u dM_u + \int_0^t \Theta_u du$ driven by a continuous martingale M_t, and one can prove the following Itô formula for X_t: [46]

$$df(t, X_t) \;=\; f_x(t, X_t)\Delta_t\, dM_t \;+\; \left(f_t(t, X_t) \;+\; f_x(t, X_t)\Theta_t \;+\; \frac{1}{2} f_{xx}(t, X_t)\Delta_t^2 \right) dt.$$

(2) Fix $u \in \mathbb{R}$. We apply this Itô formula to the function

$$f(t, x) \;:=\; \exp\left[ux \;-\; \frac{1}{2} u^2 t \right].$$

This yields the following:

$$E\left[e^{uM_t} \right] \;=\; e^{\frac{1}{2}u^2 t}.$$

(3) Thus M_t has the same MGF as a Brownian motion W_t, i.e., it is Brownian motion.
(4) It remains to prove the independence of $M_{t+h} - M_t$ and $\mathfrak{F}_t$ for all $t, h \geq 0$. ■

There also is a multidimensional version of Lévy's theorem (SCF2 Theorem 4.6.5).

Theorem 11.6 (Lévy's characterization of multidimensional Brownian Motion)**.**

[46] Compare this to (9.16) on p.193.

Assume that the process $\vec{M}_t = \big(M_t^{(1)}, \dots, M_t^{(d)}\big)$ satisfies the following.

- *Each coordinate process $M_t^{(j)}$ is a continuous $\mathfrak{F}_t$–martingale,*
- *its initial value is $\vec{M}_0 \;=\; 0$,*
- *its quadratic variations are given by $[M^{(j)}, M^{(j)}]_t \;=\; t \;\; (j = 1, \dots, d)$,*
- *its cross variations are given by $[M^{(i)}, M^{(j)}]_t \;=\; 0 \;\; (i, j = 1, \dots, d;\, i \neq j)$.*

Then, $\vec{M}_t$ is a d–dimensional Brownian motion. In particular, the coordinate processes $M_t^{(1)}, \dots, M_t^{(d)}$ are independent Brownian motions.

PROOF: ⋆ An outline of the proof can be found in SCF2 for $d = 2$. The idea is similar to that of the one dimensional case. Make again use of the fact that the Itô formula extends to Itô processes driven by continuous martingales. Apply it, for fixed $\vec{u} = (u_1, \dots, u_d)$, to the function

$$f(t, x_1, \dots, x_d) \;:=\; \exp\left[\sum_{j=1}^{d} u_j x_j \;-\; \frac{1}{2}\, t \sum_{j=1}^{d} u_j^2\right].$$

Use this equation to prove that the joint moment–generating functions of $\vec{M}_t$ and $\vec{W}_t$ are identical. This not only implies that each coordinate process $M_t^{(j)}$ is a Brownian motion (it better be since that is part of our assumptions). This MGF factors , and thus those processes are independent. We again refer to SCF2 for further detail. ■

The next proposition is a reformulation of SCF2 Example 4.6.6 (Correlated stock prices).

Proposition 11.2. ⋆

Assume that $\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)}\big)$ is a two dimensional Brownian motion and that $S_t^{(1)}$ and $S_t^{(2)}$ are two stocks with dynamics

$$\begin{aligned} dS_t^{(1)} \;&= \alpha_1 S_t^{(1)}\, dt \;+\; \sigma_1 S_t^{(1)}\, dW_t^{(1)}, \\ dS_t^{(2)} \;&= \alpha_2 S_t^{(2)}\, dt \;+\; \sigma_2 S_t^{(2)} \big[\, \rho\, dW_t^{(1)} \;+\; \sqrt{1-\rho^2}\, dW_t^{(2)} \,\big], \end{aligned}$$

where $\sigma_1, \sigma_2 > 0$ and $-1 \leq \rho \leq 1$ are constant.

(1) *Then the process*

$$W_t^* \;:=\; \rho\, W_t^{(1)} \;+\; \sqrt{1-\rho^2}\, W_t^{(2)}.$$

is a Brownian motion.

(2)

$$dS_t^{(2)} \;=\; \alpha_2 S_t^{(2)}\, dt \;+\; \sigma_2 S_t^{(2)}\, dW_t^*,$$

i.e., not only $S_t^{(1)}$, but also $S_t^{(2)}$ is a GBM (with constants α_2 and σ_2).

(3) *$W_t^{(1)}$ and W_t^* have correlation ρ for all t. Since this implies that $W_t^{(1)}$ and W_t^* are not independent, $\big(W_t^{(1)}, W_t^*\big)$ is **not** a two dimensional Brownian motion.*

PROOF:
W_t^* is a continuous martingale as the sum of continuous martingales, and $W_0^* = 0$. Further,

$$\begin{aligned} dW_t^*\, dW_t^* &= \rho^2\, dW_t^{(1)}\, dW_t^{(1)} \,+\, 2\rho\sqrt{1-\rho^2}\, dW_t^{(1)}\, dW_t^{(2)} \,+\, (1-\rho^2)\, dW_t^{(2)}\, dW_t^{(2)} \\ &= \rho^2\, dt \,+\, 0 \,+\, (1-\rho^2)\, dt \,=\, dt. \end{aligned}$$

Thus $[W^*, W^*]_t = t$ and assertion **(1)** follows from Theorem 11.5 (Lévy's characterization of one dimensional Brownian Motion).
The equation of assertion **(2)** is true by definition of W_t^*. Since we just proved assertion **(3)**, W_t^* is a Brownian motion. Thus, $dS_t^{(2)} = \alpha_2 S_t^{(2)}\, dt + \sigma_2 S_t^{(2)} dW_t^*$ is the equation of a GBM with parameters α_2 and σ_2.
To prove assertion **(3)**, we compute $\text{Cov}[W_t^1, W_t^*]$. Since $dW_t^{(1)} dW_t^{(2)} = 0$ and $dW_t^{(1)} dW_t^{(1)} = t$,

$$\begin{aligned} dW_t^{(1)}\, dW_t^* &= dW_t^{(1)} \left(\rho\, dW_t^{(1)} \,+\, \sqrt{1-\rho^2}\, dW_t^{(2)}\right) \\ &= \rho\, dW_t^{(1)}\, dW_t^{(1)} \,+\, \sqrt{1-\rho^2}\, dW_t^{(1)}\, dW_t^{(2)} \,=\, \rho\, dt. \end{aligned}$$

By Itô's product rule, $d\big(W_t^{(1)} dW_t^*\big) = W_t^{(1)} dW_t^* + W_t^* dW_t^{(1)} + dW_t^{(1)} dW_t^*$. We integrate and obtain

$$\textbf{(A)} \qquad W_t^{(1)} W_t^* \,=\, \int_0^t W_u^{(1)}\, dW_u^* \,+\, \int_0^t W_u^*\, dW_u^{(1)} \,+\, \rho t.$$

Since the Itô integrals on the right–hand side are martingales,

$$E\left[\int_0^t W_u^{(1)}\, dW_u^*\right] = E\left[\int_0^0 W_u^{(1)}\, dW_u^*\right] \,=\, 0, \quad \text{and} \quad E\left[\int_0^t W_u^*\, dW_u^{(1)}\right] \,=\, E\left[\int_0^0 W_u^*\, dW_u^{(1)}\right] \,=\, 0.$$

Thus, taking expectations in **(A)** yields $E\big[W_t^{(1)} W_t^*\big] = \rho t$.
Since $E[W_t^1] = E[W_t^*] = 0$, we conclude that

$$\text{Cov}[W_t^{(1)}, W_t^*] \,=\, E[W_t^{(1)} W_t^*] \,-\, E[W_t^{(1)}]\, E[W_t^*] \,=\, E[W_t^{(1)} W_t^*] \,=\, \rho t.$$

Since $\text{Var}[W_t^{(1)}] = \text{Var}[W_t^*] = t$, the correlation of $W_t^{(1)}]$ and W_t^* is

$$\text{Cor}[W_t^{(1)}, W_t^*] \,=\, \frac{\text{Cov}[W_t^{(1)}, W_t^*]}{\sqrt{\text{Var}[W_t^{(1)}] \cdot \text{Var}[W_t^*]}} \,=\, \frac{\rho t}{\sqrt{t^2}} \,=\, \rho\,.$$

This proves assertion **(3)**. ■

11.4 Exercises for Ch.11

Exercise 11.1. Prove prop.11.1 on p.224 of this document: If

$$dX_t \,=\, \sum_{i=1}^m \left(\Delta_t^{(i)}\, dW_t^{(i)} \,+\, \Theta_t^{(i)}\, dt\right); \qquad dY_t \,=\, \sum_{j=1}^n \Psi_t^{(j)}\, dt\,,$$

then $(dX_t)(dY_t) = 0$. □

11.5 Blank Page after Ch.11

This page is intentionally left blank!

12 Girsanov's Theorem and the Martingale Representation Theorem

12.1 Conditional Expectations on a Filtered Probability Space

For all of this chapter let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ be a filtered probability space.

The following combines both SCF2 Lemma 5.2.1 and SCF2 Lemma 5.2.2.

Proposition 12.1.

Let Z be a nonnegative random variable on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ such that $E[Z] = 1$ and $P\{Z = 0\} = 0$. Let $\widetilde{P}$ be the measure with density Z w.r.t. P, i.e.,

$$\tilde{P}(A) \;=\; \int_A Z(\omega)\, dP(\omega).$$

In other words, Z is the Radon–Nikodým derivative $\dfrac{d\widetilde{P}}{dP}$. See Chapter 4.8 (Equivalent Measures and the Radon–Nikodým Theorem). Then $\widetilde{P}$ is a probability measure which is equivalent to P, i.e.,

$$P(A) \;=\; 0 \;\;\Leftrightarrow\;\; \widetilde{P}(A) \;=\; 0.$$

We write $\widetilde{E}$ for the expectation of a random variable Y w.r.t. $\widetilde{P}$, i.e.,

$$\widetilde{E}(Y) \;=\; \int_\Omega Y\, d\widetilde{P}\,.$$

For the following we assume that $t, h \in [\,0, \infty\,[$ and that Y is an $\mathfrak{F}_t$–measurable random variable. Let $Z_t := E[Z \mid \mathfrak{F}_t]$ Then the following relations hold.

$$\widetilde{E}[Y] \;= E[Y Z_t], \tag{12.1}$$

$$\widetilde{E}\big[Y \mid \mathfrak{F}_t\big] \;= \frac{1}{Z_t}\, E[Y Z_{t+h} \mid \mathfrak{F}_t] \tag{12.2}$$

PROOF: ⋆

A. We show that $\widetilde{P}$ is a probability measure.

$$\widetilde{P}(\Omega) \;=\; \int_\Omega Z\, dP \;=\; E[Z] \;=\; 1.$$

This proves that $\widetilde{P}$ is a probability measure.

B. We show that $\widetilde{P}$ is equivalent to P.

Let $A \in \mathfrak{F}$ such that $\widetilde{P}(A) = 0$. To show $\widetilde{P} \sim P$ we only must prove that $P(A) = 0$ since $\widetilde{P} \ll P$ on account of Proposition 4.20 on p.99.

Let $Z' := (1/Z)\mathbf{1}_{\{Z>0\}}$. Then

$$\begin{aligned} 0 \;&= \widetilde{P}(A) \;=\; \int_A 1\, dP \;=\; \int_A Z Z'\, dP \;+\; \int_A 1 \cdot \mathbf{1}_{Z=0}\, dP \;=\; \int_A Z Z'\, dP \,+\, 0 \\ &= \int (\mathbf{1}_A Z') Z\, dP \;=\; \int \mathbf{1}_A Z'\, d\widetilde{P} \;=\; \int_A Z'\, d\widetilde{P} \;=\; 0. \end{aligned}$$

The last equality follows from Proposition 4.20, applied to $\mu := \widetilde{P}$ and $f := Z'$. We have shown that all $\widetilde{P}$–null sets are P–null sets, thus $P \sim \widetilde{P}$.

C. Proof of (12.1). We use in sequence

- the definition of $\widetilde{P}$: $d\widetilde{P} = Z dP$,
- iterated conditioning
- the "taking out what is known" rule
- the definition of Z_t:

$$\widetilde{E}[Y] = E[YZ] = E\big[E[YZ \,|\, \mathfrak{F}_t] \,|\big] = E\big[Y\, E[Z \,|\, \mathfrak{F}_t]\big] = E[YZ_t]. \blacksquare$$

D. Proof of (12.2). To prove that $\frac{1}{Z_t} E[YZ_{t+h} \,|\, \mathfrak{F}_t]$ is the conditional expectation of Y w.r.t. $\mathfrak{F}_t$ and $\widetilde{P}$ (not P!) we must show that

(1) $\dfrac{1}{Z_t} E[YZ_{t+h} \,|\, \mathfrak{F}_t]$ is $\mathfrak{F}_t$–measurable,

(2) $\dfrac{1}{Z_t} E[YZ_{t+h} \,|\, \mathfrak{F}_t]$ satisfies the partial averaging property

$$\text{(A)} \qquad \int_A \frac{1}{Z_t} E[YZ_{t+h} \,|\, \mathfrak{F}_t] \, d\widetilde{P} = \int_A Y \, d\widetilde{P} \quad \text{for all } A \in \mathfrak{F}_t.$$

We see that **(1)** is trivially satisfied, since $E[\cdots \mid \mathfrak{F}_t]$ enforces $\mathfrak{F}_t$–measurability.

To prove **(2)**, we first note that formula (12.1) with $\mathbf{1}_A \cdot \dfrac{1}{Z_t} \cdot E[YZ_{t+h} \,|\, \mathfrak{F}_t]$ in place of Y yields

$$\text{(B)} \quad \widetilde{E}\left[\left(\mathbf{1}_A \frac{1}{Z_t}\right) E[YZ_{t+h} \,|\, \mathfrak{F}_t]\right] = E\left[\left(\mathbf{1}_A \frac{1}{Z_t}\right) E[YZ_{t+h} \,|\, \mathfrak{F}_t] \cdot Z_t\right] = E\big[\mathbf{1}_A E[YZ_{t+h} \,|\, \mathfrak{F}_t]\big].$$

Since (12.1) holds true for all nonnegative time indices, we can replace t with $t+h$. Moreover, since $\mathbf{1}_A Y$ is $\mathfrak{F}_t$–measurable, it follows from $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+h}$ that $\mathbf{1}_A Y$ $\mathfrak{F}_{t+h}$–measurable. Thus, we are allowed to also replace Y with $\mathbf{1}_A Y$ in (12.1). We obtain

$$\text{(C)} \qquad \widetilde{E}[\mathbf{1}_A Y] = E[\mathbf{1}_A Y Z_{t+h}].$$

Proving **(2)** means proving (**A**). We will accomplish this as follows.

$$\begin{aligned}\int_A \frac{1}{Z_t} E[YZ_{t+h} \,|\, \mathfrak{F}_t] \, d\widetilde{P} &= \widetilde{E}\left[\mathbf{1}_A \frac{1}{Z_t} E[YZ_{t+h} \,|\, \mathfrak{F}_t]\right] \overset{\text{(B)}}{=} E\big[\mathbf{1}_A E[YZ_{t+h} \,|\, \mathfrak{F}_t]\big] \\ &= E\big[E[\mathbf{1}_A Y Z_{t+h} \,|\, \mathfrak{F}_t]\big] = E\big[\mathbf{1}_A Y Z_{t+h}\big] \overset{\text{(C)}}{=} \widetilde{E}[\mathbf{1}_A Y] = \int_A Y \, d\widetilde{P}.\end{aligned}$$

Here we have used the "taking out what is known" rule tobtain the equation after (**B**) and the iterated conditioning rule for the equation that follows it. We have shown that (**A**) is satisfied. ■

12.2 One dimensional Girsanov and Martingale Representation Theorems

The following is SCF2 Theorem 5.2.3.

Theorem 12.1 (Girsanov's Theorem in one dimension)**.**

Let $T > 0$ and let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ be a filtered probability space where the filtration members $\mathfrak{F}_t$ and all stochastic processes that are used in this theorem only need to exist for $0 \le t \le T$.
Let W_t be a Brownian motion on this filtered space, and let Θ_t be an adapted process which satisfies the integrability condition

(12.3) ⋆ $$E\left[\int_0^T \Theta_u^2\, Z_u^2\, du\right] < \infty.$$

where the process Z_t is defined in terms of Θ_t by formula (12.4) below.

Let

(12.4) $$Z_t := \exp\left\{-\int_0^t \Theta_u\, dW_u - \frac{1}{2}\int_0^t \Theta_u^2\, du\right\},$$

(12.5) $$\widetilde{P}(A) := \int_A Z_T\, dP \text{ for all } A \in \mathfrak{F}_T \quad \text{i.e.,} \quad Z_T = \frac{d\widetilde{P}}{dP},$$

(12.6) $$\widetilde{W}_t = W_t + \int_0^t \Theta_u\, du, \quad \text{i.e.,} \quad d\widetilde{W}_t = dW_t + \Theta_t\, dt.$$

*Then **(a)** $\widetilde{P}$ is a probability equivalent to P. **(b)** $\widetilde{W}_t$, $0 \le t \le T$, is a Brownian motion w.r.t. $\widetilde{P}$.*

PROOF ⋆ : See the proof of SCF2 Theorem 5.2.3. ■

Remark 12.1. ⋆

Strictly speaking, it is not correct to write $Z_T = \frac{d\widetilde{P}}{dP}$ in (12.5), because the domain of the probability measure P is all of $\mathfrak{F}$ and $\widetilde{P}$ only has domain $\mathfrak{F}_T$. Rather, we have

$$Z_T = \frac{d\widetilde{P}}{dP\big|_{\mathfrak{F}_T}},$$

where $P\big|_{\mathfrak{F}_T}$ is the restriction of the function $P : \mathfrak{F} \to [0,1]$ to $\mathfrak{F}_T$. See the formulation of Theorem 5.3 (Existence Theorem for Conditional Expectations) on p.113. □

Remark 12.2. The importance of the Girsanov theorem with respect to mathematical finance lies in the following. We will see later that if stock price is a generalized GBM

(12.7) $$dS_t = \alpha_t S_t\, dt + \sigma_t S_t\, dW_t, \quad 0 \le t \le T,$$

and we have a discount process with an interest rate R_t which can be stochastic (adapted):

(12.8) $$D_t = \exp\left[-\int_0^t R_s\, ds\right],$$

(see Definition 7.27 on p.156), Let us define Θ_t to be the so called market price of risk [47] process,

$$\Theta_t \;=\; \frac{\alpha_t - R_t}{\sigma_t}\,. \tag{12.9}$$

Then the discounted stock price has the dynamics

$$d\,(D_tS_t) \;=\; \sigma_t(D_tS_t)\big[\Theta_t dt \;+\; dW_t\big]\,. \tag{12.10}$$

We apply formula (12.6) of Girsanov's theorem and replace $\Theta_t dt + dW_t$ with the differential of the $\widetilde{P}$–Brownian motion $\widetilde{W}_t$. We obtain

$$d\,(D_tS_t) \;=\; \sigma_t(D_tS_t)\,d\widetilde{W}_t\,. \tag{12.11}$$

Itô calculus is defined for **any** Brownian motion, and all its theorems are in force. Thus, the process D_tS_t is a martingale with respect to the probability $\widetilde{P}$. It follows that

$$D_tS_t \;=\; \widetilde{E}[D_TS_T \mid \mathfrak{F}_t]\,. \tag{12.12}$$

Now, let us switch to self–financing portfolios

$$\vec{H}_t \;=\; \big(H_t^B, H_t^S\big) \;=\; \big(D_t(X_t - \Delta_tS_t), \Delta_t\big)\,.$$

Here we have given both the notion of MF454 Chapter 8 (The Binomial Asset Model) and SCF2: Recall that SCF2 writes Δ_t for the shares H_t^S held in the stock and X_t for the portfolio value V_t^H.
From (12.12) it will follow that the discounted portfolio value process has dynamics

$$d\big(D_tX_t\big) \;= \Delta_t\sigma_t(D_tS_t)\,d\widetilde{W}_t. \tag{12.13}$$

Thus D_tX_t is a $\widetilde{P}$–martingale. We obtain

$$D_tX_t \;=\; \widetilde{E}[D_TX_T \mid \mathfrak{F}_t]\,. \tag{12.14}$$

Now we get to the really important part. Assume that we have a contingent claim $\mathcal{X}$ with pricing process $\Pi_t(\mathcal{X})$, and that $\vec{H}$ is a replicating (thus self–financing) portfolio, i.e., it is a hedge for that claim, i.e., $X_T = \mathcal{X}$. Then, of course, $D_TX_T = D_T\mathcal{X}$. By the pricing principle,

$$X_t \;=\; \Pi_t(\mathcal{X}); \qquad \text{hence,} \quad D_tX_t \;=\; D_t\Pi_t(\mathcal{X}) \;\; \text{for } 0 \le t \le T. \tag{12.15}$$

We have found the long sought after pricing formula for a contingent claim based on a stock with generalized GBM as its price process S_t. It follows from (12.14) and (12.15) that

$$\Pi_t(\mathcal{X}) \;=\; \frac{1}{D_t}\widetilde{E}[D_TX_T \mid \mathfrak{F}_t]. \tag{12.16}$$

This formula will be used, e.g., to prove formula (10.30) of Theorem 10.1 on p.210 which gives the explicit solution for the price process $c(t,x)$ of a European call.

Before we get to develop the program outlined here we need some more theory to close the following gap. Formulas (12.15) and (12.16) hold for hedging portfolios of a contingent claim. But what claims are reachable? The martingale representation theorem, which we will discuss next, can be used to prove that **all claims can be hedged** if the information for the stock price S_t is contained in that of the driving Brownian motion W_t. □

[47] The formal definition of the market price of risk process willb be given in Definition 13.2 on p.239.

We have seen that being a martingale represents a very strong condition concerning what such a process can look like. Lévy's characterization of one dimensional Brownian Motion (Theorem 11.5 on p.225) tells us that if a martingale has continuous paths, starts at zero and has the quadratic variation of Brownian motion, then it is in fact a Brownian motion. What we will see next is that any martingale M_t with initial condition $M_0 = 0$ which is adapted to the filtration $\mathfrak{F}_t^W$ of a Brownian motion W_t is an Itô integral $M_t = \int_0^t \Gamma_u dW_u$ for some suitable adapted process Γ_t.
The following is SCF2 Theorem 5.3.1.

Theorem 12.2 (Martingale representation, one dimension)**.**

Let $T > 0$. Assume that
- $W_t, 0 \le t \le T$ *is a Brownian motion on a probability space* $(\Omega, \mathfrak{F}, P)$,
- $\mathfrak{F}_t^W, 0 \le t \le T$ *is the filtration generated by this Brownian motion,*
- $M_t, 0 \le t \le T$, *is a martingale with respect to this filtration:*
  - *for every* t, M_t *is* $\mathfrak{F}_t^W$*–measurable,*
  - $E[M_t \,|\, \mathfrak{F}_s^W] = M_s$, *for all* $0 \le s \le t \le T$.

Then there exists an adapted process $\Gamma_u, 0 \le u \le T$, *such that*

$$M_t = M_0 + \int_0^t \Gamma_u \, dW_u, \quad 0 \le t \le T. \tag{12.17}$$

PROOF: Beyond the scope of thix course. To find it, you must consult mathematically more advanced literature, e.g., [13] Øksendal, Bernt: Stochastic Differential Equations: An Introduction With Applications. ■

Remark 12.3.

If the assumptions of the martingale representation hold then **all martingales are continuous** since they are Itô integrals. This has some undesirable consequences.
If we want to model stock prices S_t which can jump at certain times without losing the very important property that the disounted stock price DTS_t is a martingale and sufficiently many claims can be hedged, then we need to include stochastic information, i.e., uncertainty, different from or besides that of Brownian motion.

This course does not discuss (continuous time) financial markets with non–continuous asset prices. Some material about this subject can be found in SCF2 Chapter 11 (Introduction to Jump Processes). There, stock prices are driven by (generalized) Poisson processes in addition to Brownian motion, and Poisson processes have jumps. □

We add the assumption $\mathfrak{F}_t = \mathfrak{F}_t^W$ to Girsanov's Theorem 12.1. This results in the following corollary (SCF2 Corollary 5.3.2).

Corollary 12.1. *Let $T > 0$ and let W_t, be a Brownian motion on a probability space $(\Omega, \mathfrak{F}, P)$ Let Θ_t, be an adapted process w.r.t. the filtration $\mathfrak{F}_t^W, 0 \le t \le T$, i.e., the filtration **generated by** W_t (!) which satisfies the integrability condition*

$$\boxed{\star} \qquad E\left[\int_0^T \Theta_u^2 \, Z_u^2 \, du\right] < \infty. \tag{12.18}$$

- *Let* $Z_t := \exp\left\{-\int_0^t \Theta_u \, dW_u - \frac{1}{2}\int_0^t \Theta_u^2 \, du\right\}$,
- $\widetilde{P}(A) := \int_A Z_T \, dP$ *for all* $A \in \mathfrak{F}_T$, *i.e.,* $Z_T = \frac{d\widetilde{P}}{dP}$,
- $\widetilde{W}_t = W_t + \int_0^t \Theta_u \, du$, *i.e.,* $d\widetilde{W}_t = dW_t + \Theta_t \, dt$ *and* $\widetilde{W}_0 = 0$.
- *Let* $\widetilde{M}_t$ $(0 \le t \le T)$ *be an* $\mathfrak{F}_t^W$*–martingale under* $\widetilde{P}$ (***not P!***)

Then there exists an $\mathfrak{F}_t^W$*–adapted process* $\widetilde{\Gamma}_u$ $(0 \le u \le T)$*, such that*

$$\widetilde{M}_t = \widetilde{M}_0 + \int_0^t \widetilde{\Gamma}_u \, d\widetilde{W}_u, \; 0 \le t \le T. \tag{12.19}$$

PROOF: Will not be given here. Just one comment. More needs to be done than just combining Girsanov's Theorem with the Martingale Representation Theorem, since the process M_t is a $\widetilde{P}$–martingale with respect to a filtration $\mathfrak{F}_t^W$, and this filtration is not generated by a $\widetilde{P}$–Brownian motion, but by the P–Brownian motion W_t! ■

Remark 12.2 on p.231 discussed the significance of Girsanov's Theorem and alluded to that of the martingale representation theorem (Theorem 12.1) when modeling contingent claims with one underlying stock. We need multidimensional versions of those theorems to model claims with several underlying stocks.

12.3 Multidimensional Girsanov and Martingale Representation Theorems

We will use in this chapter the bullet notation for stochastic integrals $\int_0^t \vec{\Gamma}_u \bullet d\vec{A}_u$ and differentials $\vec{\Gamma}_t \bullet d\vec{A}_t$ which was introduced in Notations 11.1 on p.221.

The following is SCF2 Theorem 5.4.1.

Theorem 12.3 (Girsanov's Theorem in multiple dimensions)**.**

Let $T > 0$ and let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ be a filtered probability space where the filtration members $\mathfrak{F}_t$ and all stochastic processes that are used in this theorem only need to be defined for $0 \le t \le T$.
Let $\vec{W}_t$ be a multidimensional Brownian motion

$$\vec{W}_t = (W_t^{(1)}, \dots, W_t^{(d)})$$

(thus the coordinate processes $W_i(t)$ are independent). w.r.t. the filtration $\mathfrak{F}_t, 0 \le t \le T$. Let

$$\vec{\Theta}_t = (\Theta_t^{(1)}, \dots, \Theta_t^{(d)})$$

be a d–dimensional adapted process which satisfies the integrability condition

(12.20) ⋆
$$E\left[\int_0^T \|\vec{\Theta}_u\|_2^2\, Z_u^2\, du\right] < \infty.$$

Here, $\|\vec{x}\|_2 = \sqrt{\sum_{j=1}^{d} x_j^2}$ is the standard Euclidean norm in $\mathbb{R}^d$. See Example 6.2 on p.130.

Let

(12.21)
$$Z_t := \exp\left\{-\int_0^t \vec{\Theta}_u \bullet d\vec{W}_u - \frac{1}{2}\int_0^t \|\vec{\Theta}_u\|^2\, du\right\},$$

(12.22)
$$\widetilde{P} : A \mapsto \int_A Z_T\, dP, \quad \text{i.e.,} \quad Z_T = \frac{d\widetilde{P}}{dP},$$

(12.23)
$$\widetilde{\vec{W}}_t = \vec{W}_t + \int_0^t \vec{\Theta}_u\, du, \quad \text{i.e.,} \quad d\widetilde{\vec{W}}_t = d\vec{W}_t + \vec{\Theta}_t\, dt.$$

*Then **(a)** $\widetilde{P}$ is a probability equivalent to P, **(b)** $\widetilde{\vec{W}}_t$, $0 \le t \le T$, is a Brownian motion w.r.t. $\widetilde{P}$.*

Note that the vector equations in 12.23 are to be understood componentwise:

$$\widetilde{W}_t^{(j)} = W_t^{(j)} + \int_0^t \Theta_u^{(j)}\, du, \quad \text{i.e.,} \quad d\widetilde{W}_t^{(j)} = dW_t^{(j)} + \Theta_t^{(j)}\, dt \quad \text{for } j = 1, \dots, d.$$

PROOF ⋆ : Will not be given here. ■

Remark 12.4. The following aspect of the multidimensional Girsanov Theorem deserves special mention. $\widetilde{\vec{W}}_t$ being a d–dimensional Brownian motion implies that its component processes $\widetilde{W}_t^{(j)}$ are **<u>independent</u>** w.r.t. the new probability $\widetilde{P}$. This is not at all obvious from the fact that the components of the original Brownian motion $\vec{W}$ are independent under the probability P. □

Next comes the multidimensional version of Theorem 12.2 (Martingale representation, one dimension) on p.233. This is SCF2 Theorem 5.4.2.

Theorem 12.4 (Martingale representation theorem, multiple dimensions)**.** *Let T be a fixed positive time, and assume that*

- *$\vec{W}_t, 0 \le t \le T$ is a d–dimensional Brownian motion on a probability space $(\Omega, \mathfrak{F}, P)$,*
- *$\mathfrak{F}_t^{\vec{W}}, 0 \le t \le T$ is the filtration generated by this Brownian motion,*
- *$M_t, 0 \le t \le T$, is a (one dimensional) P–martingale with respect to this filtration.*

Then there is an adapted d–dimensional process $\vec{\Gamma}_u = \big(\Gamma_1(u), \dots, \Gamma_d(u)\big), 0 \le u \le T$, such that

(12.24)
$$M_t = M_0 + \int_0^t \vec{\Gamma}_u \bullet d\vec{W}_u, 0 \le t \le T.$$

We now assume in addition to the assumptions stated so far the notation and assumptions of Girsanov's Theorem in multiple dimensions (Theorem 12.3). Then the following also is true.

Let $\widetilde{M}_t, 0 \le t \le T$, be a (one dimensional) $\widetilde{P}$–martingale with respect to $\mathfrak{F}_t^{\vec{W}}, 0 \le t \le T$, the filtration generated by the original Brownian motion $\vec{W}_t$. Here $\widetilde{P}$ is the probability from Girsanov's Theorem, equivalent to P, which makes the process $\vec{\widetilde{W}}_t$ defined by

$$d\widetilde{W}_t^{(j)} = dW_t^{(j)} + \Theta_t^{(j)}\,dt \quad \text{and} \quad \widetilde{W}_t^{(j)} = 0 \text{ for } j = 1, \dots, d,$$

an $\mathfrak{F}_t^{\vec{W}}$–Brownian motion.
Then there is an adapted d–dimensional process $\vec{\widetilde{\Gamma}}_u = \big(\widetilde{\Gamma}_u^{(1)}, \dots, \widetilde{\Gamma}_u^{(d)}\big), 0 \le u \le T$, such that

(12.25)
$$\widetilde{M}_t = \widetilde{M}_0 + \int_0^t \vec{\widetilde{\Gamma}}_u \bullet d\vec{\widetilde{W}}_u, 0 \le t \le T.$$

PROOF: Will not be given here. ■

12.4 Exercises for Ch.12

None yet

13 Black–Scholes Model Part II: Risk–neutral Valuation

In this chapter we elaborate on Remark 12.2 which gave an outline of how Girsanov's Theorem (Theorem 12.1 would be crucial in pricing a contingent claim.

13.1 The One dimensional Generalized Black–Scholes Model

In Chapter 10 (Black–Scholes Model Part I: The PDE), Definition 10.2 on p.205 stated the classical assumptions of a Black–Scholes market economy. They are rather restrictive. For example, the instantanous mean rate of return and volatility that are part of the dynamics of the stock price S_t are assumed to be constant. We weaken those assumptions for most of this entire chapter 13.

Definition 13.1 (Generalized Black–Scholes market model)**.** Let $T > 0$ and let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ be a filtered probability space We only assume that the filtration $\mathfrak{F}_t$ and all stochastic processes that will be defined later exist for times $0 \leq t \leq T$ Let $W_t, 0 \leq t \leq T$, be a Brownian motion w.r.t $\mathfrak{F}_t$.

We no more require that the instantaneous mean rate of return α, the volatility σ of the stock S_t, and the interest rate r that governs investments in the bond are constant. Instead, we assume the following.

- Let $D_t, S_t, R_t, \alpha_t, \sigma_t$ be $\mathfrak{F}_t$ adapted processes.
- Assume that $\sigma_t \neq 0$ a.s. for any given t.
- Let $\Theta_t := \dfrac{\alpha_t - R_t}{\sigma_t}$, and $Z_t := e^{-\int_0^t \Theta_u \, dW_u - \frac{1}{2}\int_0^t \Theta_u^2 \, du}$. Assume that

$$E\left[\int_0^T \Theta_u^2 \, Z_u^2 \, du\right] < \infty, \tag{13.1}$$

We speak of a **generalized Black–Scholes market model** if

$$dD_t = -R_t D_t \, dt; \quad D_0 = 1; \tag{13.2}$$

$$dS_t = \alpha_t S_t \, dt + \sigma_t S_t \, dW_t; \quad S_0 \in]0, \infty[\,; \alpha_t, \sigma_t \in]0, \infty[\,; \tag{13.3}$$

(13.4) The market is efficient: No arbitrage portfolios exist.

- We interpret D_t as the discount process associated with a riskless asset (bank account): Assume that an investment will pay the amount 1 (dollar) at the future time t. Then it's worth today, at $t = 0$, only is the amount D_t, since this amount could be invested in the bank instead, where it would increase to 1 due to interest compounded at the rate R_t.
- We interpret S_t as the price process associated with a risky asset (e.g., stock). $\square$

Remark 13.1. First some remarks about the process D_t.

(1) From (13.2) we obtain

$$D_t = \exp\left[-\int_0^t R_u du\right]. \tag{13.5}$$

This follows easily from differentiating the right hand side with respect to t.

(2) We could have worked instead with the interest rate process

$$dB_t = R_t B_t dt\,; \quad B_0 = 1\,, \qquad \text{i.e.,} \qquad B_t = \exp\left[\int_0^t R_u du\right] = \frac{1}{D_t}\,;$$

but using D_t instead makes it easier to relate the contents of this chapter to the SCF2 text.

Also, be aware of the following.

(3) Formula (13.3) states that S_t is a generalized GBM with instantaneous mean rate of return α_t and volatility σ_t, for which we have the explicit representation

$$S_t = S_0 \exp\left[\int_0^t \sigma_u dW_u + \int_0^t \left(\alpha_u - \frac{1}{2}\sigma_u^2\right) du\right]. \tag{13.6}$$

See Example 9.1 on p.194, the subsequent Remark 9.8, and (9.17) on p.194. .

(4) It was not necessary to explicitly require the adaptedness of the processes S_t and D_t. Formula (13.2) (equivalently, formula (13.5)) implies that, as far as measurability is concerned, D_t only depends on the adapted process R_s for $s \leq t$, and thus only on information in $\mathfrak{F}_t$, i.e., D_t is adapted. We conclude similarly that formula (13.3) (equivalently, formula (13.6)) implies that measurability of S_t only depends on the adapted process W_s. Thus S_t is adapted.

(5) Recall from Assumption 7.1 on p.150 that we always assume that, besides being free of arbitrage, the market has complete liquidity, no transaction costs and no bid–ask spread. □

Remark 13.2. The degree of uncertainty, i.e., the risk of investing in the bank account, is significantly smaller than that of investing in the stock. These are the reasons.

Only the randomness of the process R_t within a small interval $[t, t+h]$ affects that of the change $D_{t+h} - D_t$. Since $dtdt = 0$, this results in quadratic variation $[D, D]_t = 0$. Thus

$$dD_t\, dD_t = (-R_t D_t\, dt)\,(-R_t D_t\, dt) = R_t^2 D_t^2\, dt\, dt = 0$$

In contrast the randomness of σ_t within $[t, t+h]$ is multiplied by that of the increments of the Brownian motion W_t. Those increments are so unpredictable that they result in a quadratic variation $[W, W]_t \neq 0$. As a consequence the nonzero volatility σ_t results in fluctuations of S_t which too are so unpredictable that $[S, S]_t \neq 0$. We see this from the dynamics of S_t:

$$dS_t\, dS_t = \alpha_t^2 S_t^2\, dt\, dt + 2\alpha_t \sigma_t S_t^2\, dt\, dW_t + \sigma_t^2 S_t^2\, dW_t\, dW_t = \sigma_t^2 S_t^2\, dt\,.$$

From Itô isometry we obtain the strictly positive expression

$$[S, S]_{t+h} - [S, S]_t = \int_t^{t+h} \sigma_u^2 S_u^2\, du\,.$$

In the words of SCF2,

> Unlike the price of the money market account, the stock price is susceptible to instantaneous unpredictable changes and is, in this sense, "more random" than D_t. Our mathematical model captures this effect because S_t has nonzero quadratic variation, while D_t has zero quadratic variation. □

Formula (12.9) of Remark 12.2 on p.231 already introduced the market price of risk. Here is the formal definition.

Definition 13.2. For the generalized Black–Scholes market economy of Definition 13.1 on p.237,

the **market price of risk** is the process

$$\Theta_t = \frac{\alpha_t - R_t}{\sigma_t}\,. \tag{13.7}$$

Note that Θ_t is adapted as the difference and quotient of adapted processes. □

Remark 13.3. The assumption (13.1) on p.237,

$$E\left[\int_0^T \Theta_u^2\, Z_u^2\, du\right] < \infty, \tag{13.8}$$

will allow us to apply Girsanov's Theorem to the market price of risk process. □

13.2 Risk–Neutral Measure in a Generalized Black–Scholes Market

Assumption 13.1.

We assume for the entire remainder of this Chapter 13 (Black–Scholes Model Part II: Risk–neutral Valuation) that we have a generalized Black–Scholes market as defined in Definition 13.1 on p.237. □

Introduction 13.1. We recall definitions (8.1) on p.167 and (8.2) on p.172 of the binomial asset model in which we defined a risk–neutral measure, also called there a martingale measure, as a probability measure $\widetilde{P}$ equivalent to the "true" probability which made discouned stock price D_tS_t a $\widetilde{P}$–martingale. To see that, observe that the (not continuously) compounded interest earned between times 0 and t ($t \in \mathbb{N}$) in the bank is $(1+r)^t$, thus the discount factor is

$$D_t = \frac{1}{(1+r)^t}\,.$$

We are now in a position to prove with the help of Girsanov's Theorem the existence of a risk–neutral measure for a generalized Black–Scholes market. □

Definition 13.3 (Risk–neutral measure)**.**

A **risk–neutral measure** $\widetilde{P}$ for our generalized Black–Scholes economy, also called a **martingale measure**, is the following.

(1) $\widetilde{P}$ is a probability measure on $\mathfrak{F}_T$, i.e., $\widetilde{P}(A)$ need only be defined for events $A \subseteq \Omega$ which belong to $\mathfrak{F}_T$

(2) $\widetilde{P} \sim P$, i.e., $\widetilde{P}$ and P are equivalent on $\mathfrak{F}_T$:
If $A \in \mathfrak{F}_T$ then $\widetilde{P}(A) = 0 \Leftrightarrow P(A) = 0$.

(3) Discounted stock price D_tS_t is a $\widetilde{P}$–martingale w.r.t. the filtration $\mathfrak{F}_t$. □

Proposition 13.1. *The discounted stock price has the following dynamics and explicit representation.*

(13.9) $$d\,(D_tS_t) \;= \big(\alpha_t \;-\; R_t\big)(D_tS_t)\,dt \;+\; \sigma_t(D_tS_t)\,dW_t\,,$$

(13.10) $$D_tS_t \;= S_0\,\exp\left\{\int_0^t \sigma_u\,dW_u \;+\; \int_0^t \big(\alpha_u \;-\; R_u \;-\; \frac{1}{2}\sigma_u^2\big)\,du\right\}\,.$$

Let $d\widetilde{W}_t = dW_t + \Theta_t dt$*, where* Θ_t *is the market price process given by* (13.7)*. Then*

(13.11) $$d\,S_t \;= R_tS_t\,dt \;+\; \sigma_tS_t\,d\widetilde{W}_t\,,$$

(13.12) $$d\,(D_tS_t) \;= \sigma_t(D_tS_t)\,d\widetilde{W}_t\,.$$

PROOF:
PROOF of (13.9): By (13.2), $dD_t = -R_tD_tdt$. By (13.3), $dS_t = \alpha_tS_tdt + \sigma_tS_tdW_t$.
Since dD_t has no Brownian motion differentials, It follows from Corollary 11.2 on p.224 that

$$\begin{aligned} d\,(D_tS_t) \;&= D_t\,dS_t \;+\; S_tdD_t \;=\; D_t\big(\alpha_tS_t\,dt + \sigma_tS_t\,dW_t\big) \;-\; S_tR_tD_t\,dt \\ &= D_tS_t(\alpha_t - R_t)dt + \sigma_tD_tS_tdW_t\big) \end{aligned}$$

This proves (13.9).
PROOF of (13.10): It follows from (13.9) that D_tS_t is a generalized GBM with instantaneous mean rate of return $\alpha_t' := \alpha_t - R_t$ and volatility σ_t Since $D_0S_0 = S_0$, formula (9.20) on p.194 yields

$$D_tS_t \;=\; S_0\,\exp\left\{\int_0^t \sigma_u\,dW_u \;+\; \int_0^t \big(\alpha_u' \;-\; \frac{1}{2}\sigma_u^2\big)\,du\right\}\,.$$

Since $\alpha_t' := \alpha_t - R_t$, this proves (13.10).
PROOF of (13.11): We substitute $d\widetilde{W}_t = dW_t + \Theta_tdt$ in formula (13.3) for dS_t and obtain

$$d\,S_t \;= \alpha_tS_t\,dt + \sigma_tS_t\,dW_t \;=\; \alpha_tS_t\,dt \;+\; \sigma_tS_t\,d\widetilde{W} \;-\; \sigma_tS_t\theta_t\,dt$$

Since $\sigma_t\theta_t = \alpha_t - R_t$,

$$d\,S_t \;= \alpha_tS_t\,dt \;+\; \sigma_tS_t\,d\widetilde{W} \;-\; S_t\big(\alpha_t - R_t\big)\,dt \;=\; \sigma_tS_t\,d\widetilde{W}_t \;+\; S_tR_t\,dt\,.$$

This proves (13.11).
PROOF of (13.12): We substitute $d\widetilde{W}_t = dW_t + \Theta_tdt$ in the already proven formula (13.9)

$$\begin{aligned} d\,(D_tS_t) \;&= (\alpha_t - R_t)(D_tS_t)dt \;+\; \sigma_t(D_tS_t)\big(d\widetilde{W}_t \;-\; \Theta_tdt\big) \\ &= (\alpha_t - R_t)(D_tS_t)dt \;-\; (\sigma_t\Theta_t)D_tS_t\,dt \;+\; \sigma_tD_tS_t\,d\widetilde{W}_t\,. \end{aligned}$$

Since $\sigma_t\theta_t = \alpha_t - R_t$,

$$d\,(D_tS_t) \;= (\alpha_t - R_t)D_tS_tdt \;-\; (\alpha_t - R_t)(D_tS_t)\,dt \;+\; \sigma_t(D_tS_t)\,d\widetilde{W}_t \;=\; \sigma_t(D_tS_t)\,d\widetilde{W}_t\,. \quad\blacksquare$$

This proves (13.12).

As a consequence of Girsanov's Theorem we can prove the existence of a risk–neutral measure.

Theorem 13.1.

Let the process $Z_t (0 \le t \le T)$ be defined as follows.

$$Z_t := \exp\left\{ - \int_0^t \Theta_u \, dW_u - \frac{1}{2} \int_0^t \Theta_u^2 \, du \right\},$$

Here Θ_t is the market price of risk process, $\Theta_t = \dfrac{\alpha_t - R_t}{\sigma_t}$, of Definition 13.2 on p.239.

Then,

- *the measure $\widetilde{P} : A \mapsto \int_A Z_T(\omega) \, dP(\omega) \; (A \in \mathfrak{F}_T)$ is a probability on $\mathfrak{F}_T$, and $\widetilde{P} \sim P$.*
- *The process $\widetilde{W}_t = W_t + \int_0^t \Theta_u \, du$, (equivalently, $d\widetilde{W}_t = dW_t + \Theta_t \, dt$; $\widetilde{W}_0 = 0$), is an $\mathfrak{F}_t$–Brownian motion w.r.t the new probability measure $\widetilde{P}$.*
- *Discounted stock price $D_t S_t$ is a $\widetilde{P}$–martingale.*

PROOF: We can apply Theorem 12.1 (one dimensional Girsanov) on p.231 to Θ_t, since the assumption (13.1) (p.237) implies that the integrability condition (12.3) of that theorem is satisfied. To show that $D_t S_t$ is a $\widetilde{P}$–martingale, we apply (13.12) and obtain

(13.13)
$$\begin{aligned} d\,(D_t S_t) &= \sigma_t D_t S_t \big(d\widetilde{W}_t\big), \\ \text{i.e., } D_t S_t &= S_0 + \int_0^t \sigma_u D_u S_u \, d\widetilde{W}_u. \end{aligned}$$

We are allowed above to write S_0 for $D_0 S_0$ because $D_0 = e^{-\int_0^0 R_u du} = e^0 = 1$. Since $\widetilde{W}_t$ is an $\mathfrak{F}_t$–Brownian motion under $\widetilde{P}$, $D_t S_t$ is the sum of the $\mathfrak{F}_0$–measurable constant S_0 and a $\widetilde{P}$–Itô integral of an $\mathfrak{F}_t$–Brownian motion, hence it is a $\widetilde{P}$–martingale w.r.t to $\mathfrak{F}_t$. ■

Corollary 13.1 (Existence of a risk–neutral measure)**.**

- *The probability measure $\widetilde{P}$ of Theorem 13.1 is a risk–neutral measure for the generalized Black–Scholes market in he sense of Definition 13.3 on p.239.*
- *The dynamics of discounted stock price when using $\widetilde{W}_t$ instead of W_t are*

(13.14)
$$d\,(D_t S_t) = \sigma_t (D_t S_t) \big(d\widetilde{W}_t\big).$$

PROOF: Formula (13.14) was established in the proof of Theorem 13.1. The remainder is an obvious consequence of that theorem. ■

Remark 13.4. Note the following.

- (13.14) holds true both under the "real" probability P and the risk–neutral probability $\widetilde{P}$! It just so happens that the $\Theta_t dt$ part of $d\widetilde{W}_t = dW_t + \Theta_t dt$ prevents $D_t S_t$ from being a martingale with respect to P unless $\Theta_t = 0$, i.e., $\alpha_t = R_t$, for $0 \le t \le T$.

- Think of the above as follows. We may assume that the risk premium $\alpha_t - R_t$ in the real market, i.e., under the real world probability P, is positive on average. (See Remark 10.2 on p.206.) The redistribution of probability mass under risk–neutral probability $\widetilde{P}$ has the following effect. The upward trend of discounted stock price which happens under P as a cause of the $\Theta_t dt$ term is neutralized by $\widetilde{P}$ since this probability gives additional mass to those ω for which $\alpha_t < R_t$, at the expense of those ω for which $\alpha_t > R_t$. □

Here are some additional remarks.

Remark 13.5. This is the significance of (13.9) and (13.10) of Proposition 13.1 on p.240:

> Discounting transforms the generalized GBM S_t with an instantaneous mean rate of return α_t and volatility σ_t into another generalized GBM, $D_t S_t$, with reduced instantaneous mean rate of return $\alpha_t - R_t$.

And this is the significance of (13.11) and (13.12):

> Risk–neutral validation transforms the generalized GBM S_t with an instantaneous mean rate of return α_t and volatility σ_t into another generalized GBM, $D_t S_t$, with the same instantaneous mean rate of return R_t as the risk free asset and unchanged volatility σ_t.

Neither transformation affects the volatility. It remains σ_t in all cases.

Let us also revisit formulas (13.9)–(13.12) from the point of view that $\widetilde{P}$ is a martingale measure, and $\widetilde{W}$ is a $\widetilde{P}$–Brownian motion.

- (13.9) and its equivalent form, (13.10), both state that discounting at the riskless rate R_t decreases α_t, the instantaneous rate of return, by R_t to $\alpha_t - R_t$.
- (13.11) expresses that risk–neutral validation amounts to not considering the risk that comes with investing in the risky asset. It seems natural that the risk premium in height of $\alpha_t - R_t$ that we add to R_t, the rate of return for the riskless asset, should go away.
- Since S_t has R_t as its rate of return under $\widetilde{P}$ and discounting with D_t reduces the rate of return by R_t, discounted stock price $D_t S_t$ should have no trend to move up or down, given its current value. This is the meaning of (13.11) which shows that $D_t S_t$ is a $\widetilde{P}$–martingale. □

13.3 Dynamics of Discounted Stock Price and Portfolio Value

We saw in Chapter 10.3 (Discounted Values of Option Price and Hedging Portfolio) that in a (classical) Black–Scholes market the budget equation for a self–financing portfolio is given by formula (10.14) on p.206,

$$dV_t \;=\; Y_t dS_t \;+\; rX_t\, dt\,.$$

Here, $Y_t = H_t^S$ = stock shares, $X_t V_t^B = V_t - Y_t S_t$. [48] In the generalized Black–Scholes market we obtain dV_t by replacing the constant interest rate r with the varying interest rate $R_t(\omega)$.

[48] See Notations 10.2 on p.204.

Proposition 13.2.

The budget equation for a self–financing portfolio is

$$dV_t \;=\; Y_t\,dS_t \;+\; R_tX_t\,dt \tag{13.15}$$

Further we have the following equation for the portfolio value dynamics.

$$dV_t \;=\; R_tV_t\,dt \;+\; Y_t\sigma_tS_t\left[\,\Theta_t\,dt \;+\; dW_t\,\right]. \tag{13.16}$$

PROOF: Equation (13.15) is obvious. It just states that the number Y_t of shares held in the stock increases by the change dS_t inasset price, and the value X_t of the bond holdings changes during dt according to the interest rate, R_t.

We repeat here the proof of (13.16) as it is given in SCF2, Chapter 5.2.3 (Value of Portfolio Process Under the Risk–Neutral Measure).

$$\begin{aligned}
dV_t &= Y_t\,dS_t \;+\; R_tX_t\,dt\\
&= Y_t\left(\alpha_tS_t\,dt \;+\; \sigma_tS_t\,dW_t\right) \;+\; R_t\big(V_t - Y_t\,S_t\big)\,dt\\
&= \alpha_tY_tS_t\,dt \;+\; Y_t\sigma_tS_t\,dW_t \;+\; R_tV_t \;-\; R_tY_tS_t\,dt\,.
\end{aligned}$$

We re–order, then group the Y_tS_tdt terms, then use $\alpha_t - R_t = \Theta_t\sigma_t$ (market price of risk equation).

$$\begin{aligned}
dV_t &= R_tV_tdt \;+\; \alpha_tY_tS_t\,dt \;-\; R_tY_tS_t\,dt \;+\; Y_t\sigma_tS_t\,dW_t\\
&= R_tV_t\,dt \;+\; Y_t\big(\alpha_t - R_t\big)S_t\,dt \;+\; Y_t\sigma_tS_t\,dW_t\\
&= R_tV_t\,dt \;+\; Y_t\sigma_tS_t\left[\,\Theta_t\,dt \;+\; dW_t\,\right]. \;\blacksquare
\end{aligned}$$

Proposition 13.3.

The discounted portfolio value D_tV_t has dynamics

$$d\big(D_tV_t\big) \;=\; Y_t\sigma_tD_tS_t\,d\widetilde{W}_t. \tag{13.17}$$

PROOF: Again we follow SCF2. It follows from Corollary 11.2 on p.224 and $dD_t = -R_tD_tdt$, that

$$d(D_tV_t) \;=\; D_t\,dV_t \;+\; V_tdD_t \;=\; D_t\,dV_t \;-\; V_t(R_tD_tdt)\,.$$

Next we apply (13.16) to dV_t and obtain

$$\begin{aligned}
d(D_tV_t) &= D_t\left(R_tV_t\,dt \;+\; Y_t\sigma_tS_t\left[\,\Theta_t\,dt \;+\; dW_t\,\right]\right) \;-\; V_t(R_tD_tdt)\\
&= D_tR_tV_t\,dt \;+\; D_tY_t\sigma_tS_t\left[\,\Theta_t\,dt \;+\; dW_t\,\right] \;-\; V_tR_tD_tdt\\
&= D_tY_t\sigma_tS_t\left[\,\Theta_t\,dt \;+\; dW_t\,\right].
\end{aligned}$$

This proves (13.17). $\blacksquare$

It follows from Proposition 13.3 that D_tV_t is a martingale under $\widetilde{P}$, thus

$$D_tV_t \;=\widetilde{E}[D_TV_T\,|\,\mathfrak{F}_t] \;\text{ for all } 0 \le t \le T. \tag{13.18}$$

Now assume that V_t is the value of the hedging portfolio for a contingent claim $\mathcal{X}$. We denote the arbitrage free price process of $\mathcal{X}$ by $\Pi_t(\mathcal{X})$, and we recall that $\Pi_T(\mathcal{X}) = \mathcal{X}$, since $\mathcal{X}$ denotes the payoff at time T of the derivative on which this claim is based.

According to the pricing principle, $V_t = \Pi_t(\mathcal{X})$ holds for all $t \leq T$ to avoid arbitrage. Of course, this implies that $D_t V_t = D_t \Pi_t(\mathcal{X})$ for all $t \leq T$. We obtain from Proposition 13.3 the following

Corollary 13.2. *Assume that V_t is the value process of a hedging portfolio for a contingent claim with price process $\Pi_t(\mathcal{X})$ for $0 \leq t \leq T$. Then*

$$\begin{aligned} D_t\,\Pi_t(\mathcal{X}) &= \widetilde{E}[D_T\,\mathcal{X}\,|\,\mathfrak{F}_t],\ 0 \leq t \leq T. \\ \Pi_t(\mathcal{X}) &= \widetilde{E}\left[e^{-\int_t^T R_u du}\mathcal{X}\,\middle|\,\mathfrak{F}_t\right],\ 0 \leq t \leq T. \end{aligned}$$

PROOF: The equation for $D_t\Pi_t(\mathcal{X})$ results from this process being a $\widetilde{P}$–martingale. The formula for $\Pi_t(\mathcal{X})$ is then obtained by noting that

$$D_T \;=\; \exp\left(-\int_0^T R_u\,du\right) \;=\; \exp\left(-\int_0^t R_u\,du\right)\,\exp\left(-\int_t^T R_u\,du\right)$$

and observing that the exponential $e^{-\int_0^t R_u\,du}$ is $\mathfrak{F}_t$ measurable and can be pulled out of the conditional expectation. ■

Definition 13.4 (Risk–neutral valuation formula)**.** We call either one of the Corollary 13.2 formulas,

(13.19) $$D_t\,\Pi_t(\mathcal{X}) = \widetilde{E}[D_T\,\mathcal{X}\,|\,\mathfrak{F}_t],\quad 0 \leq t \leq T.$$

(13.20) $$\Pi_t(\mathcal{X}) = \widetilde{E}\left[e^{-\int_t^T R_u du}\mathcal{X}\,\middle|\,\mathfrak{F}_t\right],\quad 0 \leq t \leq T.$$

the **risk–neutral pricing formula**, also the **risk–neutral valuation formula** for a contingent claim with contract function $\mathcal{X}$. □

13.4 Risk–Neutral Pricing of a European Call

Assumption 13.2. For this entire subchapter we assume the following.

- The instantaneous mean rate of return is constant: $\alpha_t(\omega) = \alpha$.
- The volatility is constant: $\sigma_t(\omega) = \sigma$.
- The interest rate is constant: $R_t(\omega) = r$.
- the derivative is a European call, i.e., the payoff is $\mathcal{X} = \Phi(S_T) = (S_T - K)^+$. □

We now derive the Black–Scholes formula for the price of this European call. [49] Since the contract

[49] SCF2 does not ask that α_t be constant, presumably because this variable does not directly show in the formula

$$c(t, S_t) \;=\; \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+\,|\,\mathfrak{F}_t\right].$$

But without that assumption S_t would not be a GBM, only a generalized GBM which is not necessarily Markov, since part or all of the past could enter the dynamics $dS_t = \alpha_t S_t dt + \sigma S_t dt$.

function for a European call is

$$\mathcal{X} \;=\; \Phi(S_T) \;=\; (S_T \,-\, K)^+,$$

the risk–neutral valuation formula (13.20) on p.244 for V_t reads

$$\Pi_t(\mathcal{X}) \;=\; \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+ \,|\mathfrak{F}_t\right]. \tag{13.21}$$

We are looking for a way to evaluate this expression only using data known at time t. This could be accomplished if there was a function $(t,x) \mapsto c(t,x)$ of time t and stock price x such that

$$c\big(t, S_t\big) \;= \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+ \,|\, \mathfrak{F}_t\right]. \tag{13.22}$$

There is hope to find such a function because the geometric Brownian motion S_t is a Markov process, thus the right–hand side of (13.22) only depends on stock price S_t and time t, but not on the stock price prior to time t.

To achieve that goal, we fix a time $0 \leq t \leq T$ and define

$$\tau \;:= T \,-\, t; \qquad\qquad Y \;:=\; -\,\frac{\widetilde{W}_T \,-\, \widetilde{W}_t}{\sqrt{\tau}}. \tag{13.23}$$

$$h(t;x,y) \;:= e^{-r\tau}\left(x \cdot \exp\left\{-\sigma\sqrt{\tau}y \,+\, \left(r \,-\, \frac{\sigma^2}{2}\right)\tau\right\} \,-\, K\right)^+. \tag{13.24}$$

Note that Y is standard normal w.r.t. $\widetilde{P}$ since $\widetilde{W}_t, t \geq 0$, is a $\widetilde{P}$–Brownian motion.

We next provide three lemmas which have the following purpose.

- Lemma 13.1 shows that we can work with $h(t;S_t,Y)$ instead of $e^{-r\tau}(S_T - K)^+$.
- Lemma 13.2 gives the definition of $c(t,x)$ in terms of $h(t;x,y)$.
- Lemma 13.3 allows us to actually compute $c(t,x)$. The result will be formula (10.30) of Theorem 10.1 on p.210 which was stated there without proof.

Lemma 13.1. *With the above definitions we can rewrite the risk–neutral valuation formula* (13.21) *for a European call as follows.*

$$\widetilde{E}\left[e^{-r\tau}(S_T - K)^+ \,|\mathfrak{F}_t\,\right] \;= \widetilde{E}\left[h(t;S_t,Y) \,|\, \mathfrak{F}_t\,\right] \tag{13.25}$$

PROOF: According to (13.10) on p.240,

$$S_t \;= S_0\, \exp\left\{\int_0^t \sigma_s\, d\widetilde{W}_s \,+\, \int_0^t \left(R_s ds \,-\, \frac{1}{2}\,\sigma_s^2\right) ds\right\} \;=\; S_0\, \exp\left\{\sigma\, \widetilde{W}_t \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\, t\right\}.$$

For $t = T$ we obtain similarly that $S_T \;=\; S_0\, \exp\left\{\sigma\, \widetilde{W}_T \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\, T\right\}$. Thus,

$$\begin{aligned}\frac{S_T}{S_t} \;&=\; \exp\left\{\left[\sigma\, \widetilde{W}_T \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\, T\right] \,-\, \left[\sigma\, \widetilde{W}_t \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\, t\right]\right\}\\ &=\; \exp\left\{\sigma\, (\widetilde{W}_T \,-\, \widetilde{W}_t) \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\, (T-t)\right\},\end{aligned}$$

thus

$$\begin{aligned} S_T &= S_t \cdot \exp\left\{\sigma\,(\widetilde{W}_T - \widetilde{W}_t) + \left(r - \frac{1}{2}\,\sigma^2\right)(T-t)\right\} \\ \textbf{(A)} \qquad &= S_t \cdot \exp\left\{-\sigma\tau\,\frac{-(\widetilde{W}_T - \widetilde{W}_t)}{\tau} + \left(r - \frac{1}{2}\,\sigma^2\right)(T-t)\right\} \\ &\overset{(13.23)}{=} S_t \cdot \exp\left\{-\sigma\tau\,Y + \left(r - \frac{1}{2}\,\sigma^2\right)(T-t)\right\}. \end{aligned}$$

It follows that

$$\begin{aligned} h(t; S_t, Y) &= e^{-r\tau}\left(S_t \cdot \exp\left\{-\sigma\sqrt{\tau}Y + \left(r - \frac{\sigma^2}{2}\right)\tau\right\} - K\right)^+ \\ &\overset{\textbf{(A)}}{=} e^{-r\tau}(S_T - K)^+. \end{aligned}$$

We apply conditional expectations $\widetilde{E}[\cdots \mid \mathfrak{F}_t]$ to both sides and assertion (13.25) follows. ■

We remember our goal: find a function $(t,x) \mapsto c(t,x)$ such that (13.22) holds:

$$c\big(t, S_t\big) = \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+ \,\middle|\, \mathfrak{F}_t\right]. \tag{13.26}$$

Lemma 13.1 allows us to reformulate this problem as follows: Let $h(t;x,y)$ be the function given in formula (13.24). We want to find a function $(t,x) \mapsto c(t,x)$ such that

$$c(t, S_t) = \widetilde{E}\left[h(t; S_t, Y) \,\middle|\, \mathfrak{F}_t\right]. \tag{13.27}$$

The next lemma shows how to define this function $c(t,x)$.

Lemma 13.2. *Let*

$$c(t,x) := \widetilde{E}[h(t;x,Y)], \tag{13.28}$$

where $h(t;x,y)$ is the function defined in (13.24). Then $c(t,S_t)$ satisfies (13.27) and hence also the risk–neutral pricing formula (13.22), i.e.,

$$c(t, S_t) = \widetilde{E}\left[e^{-r\tau}(S_T - K)^+ \,\middle|\, \mathfrak{F}_t\right]. \tag{13.29}$$

PROOF: We fix $0 \le t \le T$. Since S_t is $\mathfrak{F}_t$–measurable and $Y = -\frac{\widetilde{W}_T - \widetilde{W}_t}{\sqrt{\tau}}$ is, as a function of the Brownian increment $\widetilde{W}_T - \widetilde{W}_t$, independent of $\mathfrak{F}_t$, it follows for each tixed $0 \le t \le T$ from the Independence Lemma (Lemma 5.7 on p.121) [50] that

$$c(t, S_t) = \widetilde{E}\left[h(t; S_t, Y) \,\middle|\, \mathfrak{F}_t\right].$$

50 There we wrote $h(x,y)$ instead of $h(t;x,y)$,
and $g(x) = E[h(x,Y)]$ instead of $c(t,x) = \widetilde{E}[h(t;x,Y)]$.

This proves the validity of (13.27). We apply Lemma 13.1 and (13.29) follows. ■

We have shown that the function $c(t,x) = \widetilde{E}[h(t;x,Y)]$ allows us to price a European call option, at time t, conditioned on the stock price S_t at that time, via the risk–neutral pricing formula

$$\Pi_t(\mathcal{X}) = c\big(t,S_t\big) = \widetilde{E}\left[e^{-r(T-t)}(S_T-K)^+ \,\big|\, \mathfrak{F}_t\right]. \tag{13.30}$$

It follows from the definition of $h(t;x,y)$ given in (13.24) that

$$c(t,x) = \widetilde{E}[h(t;x,Y)] = \widetilde{E}\left[e^{-r\tau}\left(x\cdot\exp\left\{-\sigma\sqrt{\tau}Y + \left(r-\frac{\sigma^2}{2}\right)\tau\right\} - K\right)^+\right].$$

This is an ordinary expected value of a function which depends on ω only by means of the $\widetilde{P}$–standard normal random variable Y. This we have learned to work with and we are able to obtain a concrete representation of $c(t,x)$ by computing this expected value. We use again the symbols $d_-(\tau,x)$ and $d_+(\tau,x)$ introduced in Theorem 10.1 on p.210:

$$d_\pm(\tau,x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r\pm\frac{\sigma^2}{2}\right)\tau\right], \tag{13.31}$$

Lemma 13.3. *The pricing function $c(t,x)$ for a European call option is given by the formula*

$$c(t,x) = x\,N\big(d_+(\tau,x)\big) - e^{-r\tau}K\,N\big(d_-(\tau,x)\big). \tag{13.32}$$

PROOF: It is true for any random variable U with a $\widetilde{P}$–density $f_U(u)$, and for any deterministic (measurable) function $u\mapsto\varphi(u)$, that $\widetilde{E}[\varphi(U)] = \int\limits_{-\infty}^{\infty}\varphi(u)f_U(u)\,du$.

We apply this to the random variable Y which has density $f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/y}$ since it is standard normal, and to the function $h(t;x,Y)$ of Y. We obtain

$$\begin{aligned} c(t;x) &\overset{(13.28)}{=} \widetilde{E}[h(t;x,Y)] = \int_{-\infty}^{\infty} h(t;x,y)\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{y^2}{2}}\,dy \\ &\overset{(13.24)}{=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-r\tau}\left(x\cdot\exp\left\{-\sigma\sqrt{\tau}y + \left(r-\frac{\sigma^2}{2}\right)\tau\right\} - K\right)^+ e^{-\frac{y^2}{2}}\,dy\,. \end{aligned}$$

Since the function $u\mapsto\log(u)$ is strictly increasing: $u<u' \Leftrightarrow \log u < \log u'$, and since always $e^{-r\tau}>0$, the integrand is positive (i.e., not zero) if and only if

$$\begin{aligned} &\log x + \left\{-\sigma\sqrt{\tau}y + \left(r-\frac{\sigma^2}{2}\right)\tau\right\} > \log K \\ \Leftrightarrow\ &\log x - \log K + \left(r-\frac{\sigma^2}{2}\right)\tau > \sigma\sqrt{\tau}y \\ \Leftrightarrow\ &\sigma\sqrt{\tau}y < \log\left(\frac{x}{K}\right) + \left(r-\frac{\sigma^2}{2}\right)\tau \\ \Leftrightarrow\ &y < \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{x}{K}\right) + \left(r-\frac{\sigma^2}{2}\right)\tau\right] = d_-(\tau,x)\,. \end{aligned} \tag{13.33}$$

Therefore,

$$c(t,x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(\tau,x)} e^{-r\tau} \left(x \exp\left\{ -\sigma\sqrt{\tau}y + \left(r - \frac{1}{2}\sigma^2 \right)\tau \right\} - K \right) e^{-\frac{1}{2}y^2}\, dy\,.$$

We simplify

$$e^{-r\tau}\, x\, e^{-\sigma\sqrt{\tau}y + \left(r - \frac{1}{2}\sigma^2\right)\tau} = x\, e^{-r\tau}\, e^{-\sigma\sqrt{\tau}y} e^{r\tau}\, e^{-\frac{\sigma^2}{2}\tau} = x\, e^{-\sigma\sqrt{\tau}y}\, e^{-\frac{\sigma^2}{2}\tau}\,,$$

and obtain

$$\begin{aligned} c(t,x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(\tau,x)} x \exp\left\{ -\frac{y^2}{2} - \sigma\sqrt{\tau}y - \frac{\sigma^2\tau}{2} \right\} dy - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_-(\tau,x)} e^{-r\tau} K e^{-\frac{1}{2}y^2}\, dy \\ &= \frac{x}{\sqrt{2\pi}} \int_{-\infty}^{d_-(\tau,x)} \exp\left\{ -\frac{1}{2}(y + \sigma\sqrt{\tau})^2 \right\} dy - e^{-r\tau} K\, N\big(d_-(\tau,x)\big)\,. \end{aligned}$$

The last equation was obtained by replacing the integral $\int\limits_{-\infty}^{d_-(\tau,x)} e^{-\frac{1}{2}y^2}\, dy$ over the standard normal density with the CDF, $N\big(d_-(\tau,x)\big)$. Thus

$$\begin{aligned} c(t,x) &= \frac{x}{\sqrt{2\pi}} \int_{-\infty}^{d_-(\tau,x)+\sigma\sqrt{\tau}} \exp\left\{ -\frac{z^2}{2} \right\} dz - e^{-r\tau} K N\big(d_-(\tau,x)\big) \\ &= x\, N\big(d_+(\tau,x)\big) - e^{-r\tau} K\, N\big(d_-(\tau,x)\big)\,. \end{aligned}$$

We have proven formula (13.32). The last equation holds because, according to (13.31),

$$\begin{aligned} d_+(\tau,x) &= d_-(\tau,x) + \sigma\sqrt{\tau} \\ &= \frac{1}{\sigma\sqrt{\tau}} \left[\log\frac{x}{K} + \left(r + \frac{1}{2}\sigma^2 \right)\tau \right]\,. \quad \blacksquare \end{aligned} \tag{13.34}$$

This was indeed the proof of Theorem 10.1 on p.210, since the classical Black–Scholes market condidions under which it was stated satisfy the assumptions 13.2 on p.244. The difference is that the function $c(t,x)$ was given there as the solution to the (deterministic) Black–Scholes PDE (10.25)

$$c_t(t,x) + rx\, c_x\big(t,x\big) + \frac{1}{2}\sigma^2 x^2 c_{xx}(t,x) = r\, c\big(t,x\big),\ x \geq 0\,,$$

with terminal condition

$$c(T,x) = (x - K)^+,$$

whereas we derived the same function in this chapter as an application of the risk–neutral valuation formula.

The next theorem just reformulates the results of the preceding lemmas.

Theorem 13.2. *We defined in Remark 10.7 on p. 211, for $\tau = T - t$, i.e., $t = T - \tau$,*

$$BSM(\tau, x; K, r, \sigma) := c(t,x), \quad \textit{where} \quad c(t,x) = x\, N\big(d_+(\tau,x)\big) - e^{-r\tau} K\, N\big(d_-(\tau,x)\big). \tag{13.35}$$

If we redefine BSM $(\tau, x; K, r, \sigma)$ to be

$$BSM(\tau, x; K, r, \sigma) = \widetilde{E}\left[e^{-r\tau}\left(x \exp\left\{-\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} - K\right)^+\right], \tag{13.36}$$

where Y is a standard normal random variable under $\widetilde{P}$, then the following holds true:

$$BSM(\tau, x; K, r, \sigma) = x\,N\big(d_+(\tau, x)\big) - e^{-r\tau}K\,N\big(d_-(\tau, x)\big). \tag{13.37}$$

PROOF: Follows from the preceding Lemmas and the fact that the right–hand side of (13.37) matches the definition of $c(t, x)$ given in (13.35). ■

13.5 Completeness of the One dimensional Generalized Black–Scholes Model

We have seen in Corollary 13.2 on p.244 that any contingent claim $\mathcal{X}$ that can be replicated can be priced by means of the risk–neutral valuation formula.

$$\Pi_t(\mathcal{X}) = \widetilde{E}\left[e^{-\int_t^T R_u du}\mathcal{X}\,\middle|\,\mathfrak{F}_t\right], \quad 0 \le t \le T. \tag{13.38}$$

The question that has not been aswered is the following. What claims can be hedged? We will explore that in this chapter.

We assume that we operate in a generalized Black–Scholes market as was defined in Definition 13.1 on p.237, in particular, that the market price of risk process Θ_t is such that the integrability condition (13.1) given in that definition is satisfied and thus Girsanov's Theorem can be applied.

Assumption 13.3. We need to apply the martingale representation theorem and must make the following additional assumptions.

The filtration $\mathfrak{F}_t$ is generated by the Brownian motion W_t and $\mathfrak{F}$ only contains information generated that Brownian motion up to time T. In other words,

$$\mathfrak{F}_t = \mathfrak{F}_t^W = \sigma\{W_u : u \le t\} \text{ for all } 0 \le t \le T,$$
$$\mathfrak{F} = \mathfrak{F}_T^W.$$

We have the following result. See SCF2, ch.5.3.2 (Hedging with One Stock).

Theorem 13.3 (Completeness of the one dimensional Generalized Black–Scholes market)**.** *Given the additional assumptions 13.3, we have the following.*

The one dimensional Generalized Black–Scholes market is complete, i.e., every contingent claim can be hedged. Further, if $0 \le t \le T$, the quantity Y_t of the replicating portfolio is given by either of

$$Y_t\sigma_t D_t S_t = \widetilde{\Gamma}_t, \tag{13.39}$$

$$Y_t = \frac{\widetilde{\Gamma}_t}{\sigma_t D_t S_t}. \tag{13.40}$$

Here the process $\widetilde{\Gamma}_t$ is implicitly defined by the equation

$$D_t\Pi_t(\mathcal{X}) = \Pi_0(\mathcal{X}) + \int_0^t \widetilde{\Gamma}_u \, d\widetilde{W}_u \quad (0 \le t \le T)\,, \tag{13.41}$$

$$\text{i.e.,}\quad d(D_t\Pi_t(\mathcal{X})) = \widetilde{\Gamma}_t \, d\widetilde{W}_t \quad (0 \le t \le T)\,. \tag{13.42}$$

PROOF: We create the hedge $\vec{H}_t$ by first looking at the pricing function $\Pi_t(\mathcal{X})$ of the claim $\mathcal{X}$ that the value process V_t of $\vec{H}_t$ must replicate for each t. This will allow us to determine the quantity Y_t of the underlying stock (and thus the bond holdings $X_t = V_t - S_tY_t$) for $\vec{H}_t$.

Since $\vec{H}$ replicates $\mathcal{X}$, the pricing principle mandates $V_t = \Pi_t(\mathcal{X})$ for all t. From risk–neutral validation (13.38) we obtain

$$\Pi_t(\mathcal{X}) = \widetilde{E}\left[e^{-\int_t^T R_u du}\mathcal{X} \,\middle|\, \mathfrak{F}_t\right], \quad 0 \le t \le T\,. \tag{13.43}$$

Since $\Pi_t(\mathcal{X}) = V_t$, $D_t\Pi_t(\mathcal{X}) = D_tV_t$. This plus the other risk–neutral validation formula which expresses the fact that the discounted portfolio value D_tV_t is a $\widetilde{P}$–martingale, yields

$$D_t\Pi_t(\mathcal{X}) = \widetilde{E}\left[D_T\mathcal{X} \,\middle|\, \mathfrak{F}_t\right], \quad 0 \le t \le T\,. \tag{13.44}$$

It now follows from Corollary 12.1 (p.233) to the martingale representation theorem in one dimension that there exists an $\mathfrak{F}_t^W$–adapted process $\widetilde{\Gamma}_u, 0 \le u \le T$, such that (13.41) holds. Here we made use of the fact that

$$D_0 = e^{-\int_0^0 R_u du} = e^0 = 1, \qquad \text{hence,} \quad D_0\Pi_0(\mathcal{X}) = \Pi_0(\mathcal{X})\,.$$

We compare (13.42) to formula (13.17) on p.243 for the differential of $D_t\Pi_t(\mathcal{X})$,

$$d\big(D_tV_t\big) = Y_t\sigma_tD_tS_t \, d\widetilde{W}_t\,.$$

Since $\sigma_tD_tS_t > 0$ as the product of three strictly positive quantities, we obtain the desired quantity Y_t for the number of shares of a hedge $\vec{H}$ for our claim according to either of (13.39) or (13.40). ■

Remark 13.6. Note that the formulas for Y_t given in the preceding theorem are of no practical value to compute this process, since the process $\widetilde{\Gamma}_t$ cannot be constructed: The martingale representation theorem is an existence only theorem. □

13.6 Multidimensional Financial Market Models

Necessary changes for Ch.13.6 (Multidimensional Financial Market Models):

- MPoR (Market price of risk equations are defined too late
- Review entire chapter for typos/errors
- Write $\sigma_t^{(**)}, \sigma_t^{(i*)}, \sigma_t^{(*j)}$. Introduce general matrix notation intro ch.2 or 3
- Check the proof of Prop.13.6 (SCF2 Lemma 5.4.5) on p.253.

Assumption 13.4. For this entire subchapter we assume the following.
Given are a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$, a d–dimensional Brownian motion

$$\vec{W}_t \;=\; \big(W_t^{(1)}, W_t^{(2)} \dots, W_t^{(d)}\big)$$

w.r.t. the filtration $\mathfrak{F}_t$ ($d \in \mathbb{N}$), and m risky assets (stocks)

$$\vec{\mathscr{A}} \;=\; \big(\mathscr{A}^{(0)}, \mathscr{A}^{(1)}, \dots, \mathscr{A}^{(m)}\big),$$

with stock prices $\vec{S}_t = \big(S_t^{(1)}, \dots, S_t^{(m)}\big)$.

We assume that each stock price $S_t^{(i)}$ is driven by $\vec{W}_t$, with dynamics

$$dS_t^{(i)} \;=\; \alpha_t^{(i)} S_t^{(i)}\,dt \;+\; S_t^{(i)} \sum_{j=1}^{d} \sigma_{ij}(t)\,dW_t^{(j)}, \quad i = 1, \dots, m, \tag{13.45}$$

and that we have the usual discount process which is based on an adapted interest rate process R_t.

$$dD_t \;=\; -R_t D_t\,dt, \;\; D_0 \;=\; 1, \quad \text{i.e.,} \quad D_t \;=\; \exp\left(-\int_0^t R_u\,du\right). \tag{13.46}$$

In the above we assume that the vector valued process $\vec{\alpha}_t = \big(\alpha_t^{(1)}, \dots, \alpha_t^{(m)}\big)$ which we call the **mean rate of return** vector, and the matrix valued adapted process $\big(\sigma_{ij}(t)\big)_{i=l,\dots,m;j=l,\dots,d}$ which we call the **volatility matrix** both are $\mathfrak{F}_t$–adapted processes.
We further define the processes

$$\sigma_t^{(i)} \;:=\; \sqrt{\sum_{j=1}^{d} \sigma_{ij}^2(t)}, \quad i = 1, \dots, m. \tag{13.47}$$

$$B_t^{(i)} \;:=\; \sum_{j=1}^{d} \int_0^d \frac{\sigma_{ij}(u)}{\sigma_u^{(i)}}\,dW_u^{(j)}, \quad i = 1, \dots, m. \tag{13.48}$$

$$\rho_{ij}(t) \;:=\; \frac{1}{\sigma_t^{(i)}\sigma_t^{(j)}} \sum_{k=1}^{d} \sigma_{ik}(t)\sigma_{jk}(t), \quad i, k = 1, \dots, m. \tag{13.49}$$

We also assume that $\sigma_t^{(i)} > 0$ for all t. $\square$

We have the following result.

Proposition 13.4. ⋆

Each process $B_t(i)$ is a Brownian motion. The multiplication table is

(13.50) $$dB_t^{(i)}\,dB_t^{(i)} = dt, \quad i = 1,\dots,m,$$

(13.51) $$dB_t^{(i)}\,dB_t^{(j)} = \rho_{ij}(t)\,dt, \qquad i,j = 1,\dots,m,\ i \neq j\,.$$

The covariances are

(13.52) $$Cov\big[B_t^{(i)}B_t^{(j)}\big] \;=\; E\int_0^t \rho_{ij}(u)\,du\,.$$

Further, each $S^{(i)}$ is a $B_t^{(i)}$–driven generalized GBM with volatility $\sigma_t^{(i)}$ and unchanged drift $\alpha_t^{(i)}$:

(13.53) $$dS_t^{(i)} \;=\; \alpha_t^{(i)} S_t^{(i)}\,dt \;+\; \sigma_t^{(i)} S_t^{(i)} dB_t^{(i)}.$$

PROOF: See Chapter 5.4.2 (Multidimensional Market Model) in SCF2. ■

Corollary 13.3. ⋆ *Assume that $\big((\sigma_{ij}(t,\omega)\big($ is constant in t and ω. We define*

(13.54) $$\sigma_{ij} \;:=\; \sigma_{ij}(t,\omega), \quad \sigma^{(i)} \;:=\; \sigma_t^{(i)}(\omega), \quad \rho_{ik} \;:=\; \rho_{ik}(t)(\omega).$$

The latter is possible since the right hand side of (13.49) also is constant in t and ω. Then

(13.55) $$\rho_{ik} \;=\; \frac{1}{\sigma^{(i)}\sigma^{(k)}}\sum_{j=1}^{d}\sigma_{ij}\sigma_{kj} \;\; \text{for } i,k = 1,\dots,m,$$

(13.56) $$Cov\big[B_t^{(i)},B_t^{(k)}\big] \;=\; \rho_{ik}\,t\,,$$

and the correlation between $B_t^{(i)}$ and $B_t^{(j)}$ is ρ_{ik}.

PROOF: ⋆ The proof of (13.55) and (13.56) is trivial. The last assertion follows from

$$\mathrm{Var}\big[B_t^{(i)}\big] \;=\; t \quad \text{for all } i = 1,\dots,m. \;\blacksquare$$

Now some terminology.

Definition 13.5. ⋆

If the volatility matrix has entries which are **not** constant in t and ω,
we call $\rho_{ij}(t) = \rho_{ij}(t,\omega)$ the **instantaneous correlation** between $B_t^{(i)}$ and $B_t^{(j)}$, and we call $\sigma_t(i)$ the **instantaneous standard deviation** of the relative change in S_i. □

Remark 13.7. The reason for the term "relative change" is that $\sigma_t(i)$ is tied to the "relative differential" $dS_t^{(i)}/S_t^{(i)}$ as follows. From

$$\begin{aligned} dS_t^{(i)} &= \alpha_t^{(i)} S_t^{(i)}\,dt \;+\; \sigma_t^{(i)} S_t^{(i)} dB_t^{(i)}\,,\\ dt\,dB_t^{(i)} &= dB_t^{(i)}\,dt \;=\; dtdt \;=\; 0, \quad dB_t^{(i)}\,dB_t^{(j)} \;=\; \rho_{ij}dt\,, \end{aligned}$$

we obtain

$$\begin{aligned} dS_t^{(i)}\,dS_t^{(j)} &= \left(\sigma_t^{(i)} S_t^{(i)} dB_t^{(i)}\right)\left(\sigma_t^{(j)} S_t^{(j)} dB_t^{(j)}\right) \\ &= \sigma_t^{(i)}\sigma_t^{(j)} S_t^{(i)} S_t^{(j)} \left(dB_t^{(i)} dB_t^{(j)}\right) \;=\; \sigma_t^{(i)}\sigma_t^{(j)} S_t^{(i)} S_t^{(j)} \rho_{ij}\, dt\,. \end{aligned}$$

$$\text{Thus,} \qquad \left(\frac{dS_t^{(i)}}{S_t^{(i)}}\right)\left(\frac{dS_t^{(j)}}{S_t^{(j)}}\right) \;=\; \sigma_t^{(i)}\sigma_t^{(j)}\rho_{ij}\, dt\,.$$

We can express this last formula as follows. The product of the relative instantaneous changes of $S^{(i)}$ and $S^{(j)}$ is the product of the instantaneous standard deviations and the instantaneous correlation. □

Proposition 13.5. ⋆

Given the dynamics (13.45) for $\vec{S}_t$ and (13.46) for D_t, the discounted stock price vector $D_t\vec{S}_t$ has dynamics

$$d\left(D_t S_t^{(i)}\right) \;=\; D_t S_t^{(i)} \left[\left(\alpha_t^{(i)} - R_t\right) dt \;+\; \sum_{j=1}^{d} \sigma_{ij}(t)\, dW_t^{(j)} \right]. \tag{13.57}$$

PROOF: See Chapter 5.4.2 (Multidimensional Market Model) in SCF2. ■

We must generalize the definition of risk–neutral measure given in Definition 13.3 on p.239 for a financial market with a single risky asset price driven by a single Brownian motion to the multidimensional model.

Definition 13.6 (Risk–neutral measure for multiple risky assets)**.**

A **risk–neutral measure** or **martingale measure** $\widetilde{P}$ in the multitimensional market model given in the assumptions 13.4 on p.251 is the following.

- **(1)** $\widetilde{P}$ is a probability measure on $\mathfrak{F}_T$, i.e., $\widetilde{P}(A)$ need only be defined for events $A \subseteq \Omega$ which belong to $\mathfrak{F}_T$
- **(2)** $\widetilde{P} \sim P$, i.e., $\widetilde{P}$ and P are equivalent on $\mathfrak{F}_T$:
  If $A \in \mathfrak{F}_T$ then $\widetilde{P}(A) = 0 \Leftrightarrow P(A) = 0$.
- **(3)** Discounted stock price $D_t S_t^{(i)}$ is a $\widetilde{P}$–martingale w.r.t. the filtration $\mathfrak{F}_t$ for **ALL** $i = 1, \dots, m$. □

Proposition 13.6 (SCF2 Lemma 5.4.5)**.**

Let $\widetilde{P}$ be a risk–neutral measure, and let V_t be the value of a self–financing portfolio. Then discounted portfolio value $D_t V_t$ is a $\widetilde{P}$–martingale, and its differential is

$$d\big(D_t V_t\big) \;=\; D_t\left(dV_t \;-\; R_t V_t\, dt\right) \;=\; \sum_{i=1}^{m} Y_t^{(i)}\, d\big(D_t S_t^{(i)}\big). \tag{13.58}$$

PROOF: See the proof of SCF2, Lemma 5.4.5. ■

Remark 13.8. We restate here for the reader's convenience the definition 7.7 of an arbitrage portfolio on p.150.
A portfolio $\vec{H}_t$ is an arbitrage portfolio if its value process V_t satisfies

$$V_0 = 0, \tag{13.59}$$
$$P\{V_T \geq 0\} = 1, \tag{13.60}$$
$$P\{T > 0\} > 0. \tag{13.61}$$
□

Here is how we define the vector valued version of a market price of risk process.

Definition 13.7.

If it exists, then the **market price of risk** process is an adapted process

$$\vec{\Theta}_t = (\Theta_t^{(1)}, \dots, \Theta_t^{(d)})$$

which **(a)** solves the system of equations, called the **market price of risk equations**,

$$\alpha_i(t) - R_t = \sum_{j=1}^{d} \sigma_{ij}(t)\Theta_t^{(j)}, \quad i = 1, \dots, m, \tag{13.62}$$

and **(b)** satisfies the Girsanov integrability condition (formula (12.20) on p.235). □

Remark 13.9. The existence of a market price of risk process is of central importance for an efficient market.

(1) If there is no solution to the market price of risk equations, then we have a financial market model which is not free of arbitrage. It is not suitable for pricing contingent claims. For a simple example of a model which does not have a solution to the market price of risk equations and an arbitrage portfolio that this allows to be created, see SCF2 Example 5.4.4.

(2) SCF2 does not state Girsanov integrability as a condition for $\vec{\Theta}$ but we do it here because, if Girsanov's Theorem cannot be applied, then there is no guarantee that a risk–neutral measure $\widetilde{P}$ exists. We then would not able to rule out the existence of arbitrage portfolios. See the first fundamental theorem of asset pricing below (Theorem 13.5 on p.256). □

Theorem 13.4.

If a solution to the market price of risk equations

$$\alpha_i(t) - R_t = \sum_{j=1}^{d} \sigma_{ij}(t)\Theta_t^{(j)}, \quad i = 1, \dots, m,$$

exists then the market model possesses a risk–neutral probability measure.

PROOF: [★] Let $\widetilde{P}$ be the probability equivalent to P which is created in Theorem 12.3 (Girsanov's Theorem in multiple dimensions) on p.234. We recall that the process

$\vec{\widetilde{W}}_t = \left(\widetilde{W}_t^1, \dots, \widetilde{W}_t^d\right)$ with dynamics

$$d\widetilde{W}_t^{(j)} = dW_t^{(j)} + \Theta_t^{(j)}\,dt, \qquad \widetilde{W}_0^{(j)} = 0, \tag{13.63}$$

is a d–dimensional $\mathfrak{F}_t$–Brownian motion under the probability $\widetilde{P}$. We plug the market price of risk equations into formula (13.57) on p.253 and obtain

$$\begin{aligned} d\left(D_tS_t^{(i)}\right) &= D_tS_t^{(i)}\left[\sum_{j=1}^{d}\sigma_{ij}(t)\Theta_t^{(j)}\,dt + \sum_{j=1}^{d}\sigma_{ij}(t)\,dW_t^{(j)}\right] \\ &= D_tS_t^{(i)}\sum_{j=1}^{d}\sigma_{ij}(t)\left[\Theta_t^{(j)}\,dt + dW_t^{(j)}\right]. \end{aligned}$$

We apply formula (13.63) and obtain

$$d\left(D_tS_t^{(i)}\right) = D_tS_t^{(i)}\sum_{j=1}^{d}\sigma_{ij}(t)d\widetilde{W}_t^{(j)}. \tag{13.64}$$

Since each $\widetilde{W}_t^{(j)}$ is a $\widetilde{P}$–martingale, this also is true for each discounted stock price $D_tS_t^{(i)}$. It follows that $\widetilde{P}$ is a risk–neutral probability measure. ■

Remark 13.10. Let $\mathcal{X}$ be a contingent claim with price process $\Pi_t(\mathcal{X})$, [51] We would like to be able to create a hedge for that claim.
We can define $D_t\Pi_t(\mathcal{X})$ and $\Pi_t(\mathcal{X})$ by the risk–neutral pricing formulas (13.19) and (13.20) on p.244,

$$\begin{aligned} D_t\Pi_t(\mathcal{X}) &= \widetilde{E}[D_T\Pi_T(\mathcal{X})\,|\,\mathfrak{F}_t],\ 0 \le t \le T. \\ \Pi_t(\mathcal{X}) &= \widetilde{E}\left[e^{-\int_t^T R_u du}\Pi_T(\mathcal{X})\,\middle|\,\mathfrak{F}_t\right],\ 0 \le t \le T. \end{aligned}$$

Since $D_T\Pi_T(\mathcal{X})$ is constant in t, and $D_t\Pi_t(\mathcal{X})$ is the $\widetilde{P}$–conditional expectation of $D_T\Pi_T(\mathcal{X})$, this process is a martingale under $\widetilde{P}$. According to the Martingale Representation Theorem for multiple dimensions (Theorem 12.4 on p.235), there are processes $\widetilde{\Gamma}_1(u), \dots, \widetilde{\Gamma}_d(u)$ such that

$$D_t\Pi_t(\mathcal{X}) = \Pi_0(\mathcal{X}) + \sum_{j=1}^{d}\int_0^t \widetilde{\Gamma}_j(u)\,d\widetilde{W}_u^{(j)},\ \ 0 \le t \le T. \tag{13.65}$$

Consider a self–financing portfolio $\vec{H}_t$ with value process V_t. By (13.58) on p.253 and (13.64) on p.255,

$$\begin{aligned} d\big(D_tV_t\big) &= \sum_{i=1}^{m} Y_t^{(i)}\,d\left(D_tS_t(i)\right) \\ &= \sum_{j=1}^{d}\sum_{i=1}^{m} Y_t^{(i)}D_tS_t(i)\sigma_{ij}(t)\,d\widetilde{W}_t^{(j)}. \end{aligned} \tag{13.66}$$

[51] Mathematically speaking, any nonnegative, $\mathfrak{F}_T$–measurable and integrable random variable will do.

(The first equation holds because $\vec{H}_t$ is self–financing.) Equivalently,

$$D_tV_t = V_0 + \sum_{j=1}^{d}\int_0^t \sum_{i=1}^{m} Y_t^{(i)} D_u S_u^{(i)} \sigma_{ij}(u)\, d\widetilde{W}_u^{(j)}. \tag{13.67}$$

We compare the integrands of (13.65) and (13.67) and obtain

$$\widetilde{\Gamma}_j(u) \;=\; D_u \sum_{i=1}^{m} Y_t^{(i)} S_u^{(i)} \sigma_{ij}(t), \quad j = 1\dots, d,$$

To hedge the short position, we should take $V_0 = \Pi_0(\mathcal{X})$ and choose the portfolio process $\vec{Y}_t = Y_t^{(1)}, \dots, Y_t^{(m)}$ so that the **hedging equations**

$$\frac{\widetilde{\Gamma}_j(t)}{D_t} \;=\; \sum_{i=1}^{m} Y_t^{(i)} S_t(i) \sigma_{ij}(t), \quad j = 1\dots, d, \tag{13.68}$$

are satisfied. Note that these are d equations in m unknown processes $Y_t^{(1)}, \dots, Y_t^{(m)}$.

□

Next comes SCF2 Theorem 5.4.7.

Theorem 13.5.

First fundamental theorem of asset pricing:
If the market model given in Assumption 13.4 on p.251 has a risk–neutral probability measure, then it does not admit arbitrage.

PROOF: ⋆ Let $\widetilde{P}$ be a risk–neutral measure and assume that $\vec{H}$ is a self–financing portfolio with initial value $V_0 = 0$. Since D_tV_T is a $\widetilde{P}$–martingale and thus has constant expectation across all times $0 \le t \le T$ and $D_0 = e^{-\int_0^0 R_u du} = e^0 = 1$ we have

$$\widetilde{E}\big[\, D_TV_T \,\big] \;= \widetilde{E}\big[\, D_0V_0 \,\big] \;=\; V_0 \;=\; 0. \tag{13.69}$$

Assume further that $\vec{H}$ satisfies condition (13.60), $P\{V_T \ge 0\} \;=\; 1$.

$$\text{Then } \; P\{V_T < 0\} \;=\; 0, \quad \text{thus } \; \widetilde{P}\{V_T < 0\} \;=\; 0. \tag{13.70}$$

If we can show that it is impossible for $\vec{H}$ to satisfy (13.61): $P\{V_T > 0\} \;>\; 0$, then we are done since this means that no self–financing portfolio can satisfy all three conditions (13.59) (13.60), (13.61) of an arbitrage portfolio. So,

(**A**) let us assume to the contrary that $P\{V_T \;>\; 0\} \;>\; 0$.

Since $P \sim \widetilde{P}$ and thus both probabilities assign zero to the same events, we obtain $\widetilde{P}\{V_T > 0\} \;>\; 0$. Moreover, $\{V_T > 0\} \;= \{D_TV_T > 0\}$, because $D_T(\omega)$ is strictly positive for all ω as an exponential.

Let $A_j := \{D_T V_T \geq \frac{1}{j}\}$ and $A := \{D_T V_T > 0\}$. If we write $2a$ for $\widetilde{P}(A)$ then $a > 0$. Since

$$A \,=\, \bigcup_{j\in\mathbb{N}} A_j \qquad \text{and thus, by (4.39a) on p.58,} \qquad \widetilde{P}(A_j) \uparrow 2a,$$

there is some index j_0 such that $\widetilde{P}(A_{j_0}) \geq a$. We have

$$0 \stackrel{(13.69)}{=} \widetilde{E}[D_T V_T] \,=\, \int_\Omega D_T V_T \, d\widetilde{P} \,=\, \int_A D_T V_T \, d\widetilde{P} + \int_{\{D_T V_T = 0\}} D_T V_T \, d\widetilde{P} + \int_{\{D_T V_T < 0\}} D_T V_T \, d\widetilde{P}.$$

The second integral of the right hand expression is zero because the integrand vanishes on $\{D_T V_T = 0\}$. The third integral of the right hand expression is zero by (13.70), since any integral over a set of measure zero is zero. This follows from Proposition 4.20 on p.99. Hence,

$$\int_A D_T V_T \, d\widetilde{P} \,=0.$$

Since $A_{j_0} \subset A$ and $D_T V_T > 0$ on A,

$$0 \,=\, \int_A D_T V_T \, d\widetilde{P} \,\geq\, \int_{A_{j_0}} D_T V_T \, d\widetilde{P} \,\geq\, \int_{A_{j_0}} \frac{1}{j_0} \, d\widetilde{P} \,=\, \frac{1}{j_0}\,\widetilde{P}(A_{j_0}) \,\geq\, \frac{a}{j_0} \,>\, 0.$$

Thus assumption (**A**) has lead us to the contradiction $0 > 0$. This proves that $P\{V_T \,>\, 0\} \,>\, 0$; thus $\vec{H}$ is not an arbitrage portfolio. Since $\vec{H}$ was an arbitrary, self–financing portfolio, we have shown that the model is free of arbitrage. ■

Remark 13.11. Take a moment to reflect on how the proof of that last theorem was able to switch between the equivalent probabilities P and $\widetilde{P}$ by making use of

$$\begin{aligned} \widetilde{P}(...) \,&=\, 0 \,\Leftrightarrow P(...) \,=\, 0,\\ \widetilde{P}(...) \,&>\, 0 \,\Leftrightarrow P(...) \,>\, 0,\\ \widetilde{P}(...) \,&=\, 1 \,\Leftrightarrow P(...) \,=\, 1. \end{aligned}$$

Theorem 13.3 (Completeness of the one dimensional Generalized Black–Scholes market) in Subchapter 13.5 (Completeness of the One dimensional Generalized Black–Scholes Model) gave conditions under which the one dimensional market is complete, i.e., every contingent claim that is reasonably integrable can be hedged. See Definition 7.9 (Hedging/Replicating Portfolio) on p.151. We now want to examine under which conditions the multidimensional market is complete.

Assumption 13.5. We add to Assumption 13.4 the following conditions.

(1) The market price of risk equations of Definition 13.7 on p.254,

$$\alpha_i(t) \,-\, R_t \,=\, \sum_{j=1}^d \sigma_{ij}(t)\Theta_t^{(j)}, \quad i = 1, \dots, m,$$

have a solution process $\vec{\Theta}_t = \big(\Theta_t^{(1)}, \dots, \Theta_t^{(d)}\big)$.

(2) $\mathfrak{F}_t \,=\, \mathfrak{F}_t^{\vec{W}}$, i.e., $\mathfrak{F}_t$ is generated by the d–dimensional Brownian motion $\vec{W}_t$. □

Remark 13.12. The first of the above conditions implies that the conditions of Theorem 13.4 on p.254 are satisfied, hence there exists a risk–neutral probability $\widetilde{P}$.
Both conditions together ensure that the multidimensional martingale representation theorem is satisfied: Every $\mathfrak{F}_t$–martingale M_t under risk–neutral probability $\widetilde{P}$ is of the form

$$M_t \;=\; M_0 \,+\, \sum_{j=1}^{d} \int_0^t \widetilde{\Gamma}_j(u)\, d\widetilde{W}_u^{(j)}.$$

Here the process $\widetilde{\vec{W}}_t$ is the $\widetilde{P}$–d–dimensional Brownian motion

$$\widetilde{\vec{W}}_t \;=\; \vec{W}_t \,+\, \int_0^t \vec{\Theta}_u\, du\,. \;\square$$

The next theorem is SCF2 Theorem 5.4.9.

Theorem 13.6.

Second fundamental theorem of asset pricing:
Assume that a risk–neutral probability measure exists. Then

The market is complete $\Leftrightarrow$ *The risk–neutral probability measure is unique.*

The proof is not given here. See SCF2! ■

13.7 Exercises for Ch.13

Exercise 13.1. Prove the formula (13.9) of Proposition 13.1 on p.240:

$$d\,(D_tS_t) \;= \big(\alpha_t - R_t\big) D_tS_t\, dt \,+\, \sigma_t D_tS_t\, dW_t$$

directly from the dynamics given in Definition 13.1 on p.237,

$$\begin{aligned} dD_t \;&=\; -\, R_tD_t\, dt, \\ dS_t \;&= \alpha_tS_t\, dt \,+\, \sigma_tS_t\, dW_t, \end{aligned}$$

by applying the Itô product rule or one of its corollaries to $d(D_tS_t)$. $\square$

Exercise 13.2. Prove the "⇒" direction of Theorem 13.6 (Second fundamental theorem of asset pricing) on p.258 of this document: If the multidimensional market is complete then the risk–neutral probability measure is unique. $\square$

14 Dividends

Many if not most stocks pay a dividend per share at discrete times, say, anually or semi–annually or quarterly. We also consider stocks that pay dividends continually. Such stocks do not exist in reality but they can be used to model the kind of mutual fund which holds many different kinds of stocks which pay their dividends at different times.

Note that whatever money is paid out as a dividend to shareholders diminishes the company assets and thus reduces the share value accordingly.

- If a quarterly dividend of 2 dollars per share is paid at time t then stock price per share S_t will go down by 2 dollars.
- If dividends are paid continuously at a rate $A_t(\omega)$ per unit time then a dividend of (approximately) A_tS_tdt is paid per share during $[t, t+dt]$. We must subtract A_tS_tdt from dS_t.

Both cases will yield more powerful results if we specialize to constant dividend rates which vary neither with time t nor with randomness ω. Accordingly, we subdivide this chapter into

- continuously paying dividends
- dividends paid at discrete times,
- constant dividend rates.

We will limit ourselves to the one dimensional case: A single (one dimensional) Brownian motion which drives a single underlying risky asset (stock).

We try to use SCF2 notation whenever feasible.

Proposition 14.2 on p.261 will show that the probability measure $\widetilde{P}$ which is constructed in Girsanov's Theorem by means of the market price of risk process Θ_t no longer transforms the discounted stock price D_tS_t into a martingale. Accordingly, $\widetilde{P}$ no longer is a risk–neutral measure. [52] However, discounted portfolio value D_tX_t for a self–financing portfolio remains a $\widetilde{P}$–martingale.

We thus decide to use in this chapter on dividends the term **Girsanov measure** or **Girsanov probability** rather than risk–neutral measure for that probability $\widetilde{P}$.

14.1 Continuously Paying Dividends

Assumption 14.1. Unless stated otherwise we assume that we have a generalized Black–Scholes market as defined in Definition 13.1 (Generalized Black–Scholes market model) on p.237, with the following **modification**.

We assume that the stock pays a continuous dividend at a rate of $A_t(\omega)$ per unit time and that this continuous time **dividend rate process** A_t is $\mathfrak{F}_t$–adapted and nonnegative. We noted in the introduction to this chapter that this will result in the subtraction of A_tS_tdt from dS_t. Thus we replace formula (13.3) for the stock price dynamics with the following.

(14.1) $$dS_t = \alpha_t S_t\,dt + \sigma_t S_t\,dW_t - A_t S_t\,dt; \quad S_0 \in]0,\infty[\,; \ \alpha_t, \sigma_t \in]0,\infty[\,;$$

All other processes remain unchanged. In particular we have the same discount process D_t, market price of risk process Θ_t, Girsanov measure $\widetilde{P}$, and the process $\widetilde{W}_t = W_t + \int_0^t \Theta_u du$ which becomes a Brownian motion under $\widetilde{P}$. □

[52] See Definition 13.3 on p.239.

We thus have

$$dD_t = -R_t D_t\,dt\,; \quad D_0 = 1\,, \tag{14.2}$$

$$\Theta_t = \frac{\alpha_t - R_t}{\sigma_t}, \tag{14.3}$$

$$d\widetilde{W}_t = dW_t + \Theta_t\,dt\,; \quad \widetilde{W}_0 = 0\,. \;\square \tag{14.4}$$

Proposition 14.1.

The value and discounted value of a self–financing portfolio have the following dynamics.

$$dV_t = R_t V_t\,dt + Y_t S_t \sigma_t\big(\Theta_t\,dt + dW_t\big) = R_t V_t\,dt + Y_t S_t \sigma_t\,d\widetilde{W}_t, \tag{14.5}$$

$$d\big(D_t V_t\big) = Y_t D_t S_t \sigma_t\,d\widetilde{W}_t. \tag{14.6}$$

In particular, the discounted portfolio process $D_t V_t$ is a $\widetilde{P}$–martingale.

For the proof see SCF2 ch.5.5.1. ■

Remark 14.1. A. Discounted portfolio value being a $\widetilde{P}$–martingale is all it takes to use risk–neutral valuation for contingent claims. Let $\vec{H}_t$ with portfolio value V_t be a hedge for a contingent claim $\mathcal{X}$ with pricing process $\Pi_t(\mathcal{X})$. Then $V_T = \mathcal{X}$, thus $D_T\mathcal{X} = D_T V_T$ and, according to the pricing principle, $\Pi_t(\mathcal{X}) = V_t$ for all $0 \le t \le T$. Moreover, since $D_t V_t$ is an $\mathfrak{F}_t$–martingale under $\widetilde{P}$,

$$D_t\Pi_t(\mathcal{X}) = D_t V_t = \widetilde{E}\big[D_T V_T \mid \mathfrak{F}_t\big] = \widetilde{E}\big[D_T\mathcal{X} \mid \mathfrak{F}_t\big] \text{ for } 0 \le t \le T,$$

$$\text{thus} \quad \Pi_t(\mathcal{X}) = \widetilde{E}\big[D_t^{-1} D_T\mathcal{X} \mid \mathfrak{F}_t\big] = \widetilde{E}\big[e^{-\int_t^T R_u du}\mathcal{X} \mid \mathfrak{F}_t\big] \text{ for } 0 \le t \le T.$$

B. Note that formula (14.5) for dV_t matches formula 13.16 on p,243, and note that formula (14.6) for $d(D_t V_t)$ matches formula 13.17 on p,243. **Neither formula references the dividend rate process A_t!**

C. A closer inspection of the proof of Theorem 13.3 (Completeness of the one dimensional Generalized Black–Scholes market) on p.249 shows that it only depends on risk–neutral valuation and what was shown in parts **A** and **B** of this remark. We will use this observation in the proof of the next theorem. □

Theorem 14.1. *Given the assumptions 13.3 on p.249 in addition to the assumptions 14.1 made at the beginning of this chapter we have the following.*

The one dimensional Generalized Black–Scholes market with continuous dividend payments is complete, i.e., every contingent claim can be hedged. Further, the quantity Y_t of the replicating portfolio satisfies, for any $0 \le t \le T$,

$$Y_t \sigma_t D_t S_t = \widetilde{\Gamma}_t, \tag{14.7}$$

$$Y_t = \frac{\widetilde{\Gamma}_t}{\sigma_t D_t S_t}. \tag{14.8}$$

Here the process $\widetilde{\Gamma}_t$ is implicitly defined by the equation

$$D_t \Pi_t(\mathcal{X}) = \Pi_0(\mathcal{X}) + \int_0^t \widetilde{\Gamma}_u \, d\widetilde{W}_u \quad \text{for } 0 \le t \le T, \tag{14.9}$$

$$\text{i.e.,} \quad d(D_t \Pi_t(\mathcal{X})) = \widetilde{\Gamma}_t \, d\widetilde{W}_t \quad \text{for } 0 \le t \le T. \tag{14.10}$$

PROOF: ⋆ We can copy the proof of Theorem 13.3 word for word This follows from the previous remark and the fact that the definitions of Θ_t and thus $\widetilde{P}$ and $\widetilde{W}_t$ have not changed. ■

We have seen in Proposition 14.1 on p.260 that discounted portfolio value of a self–financing portfolio behaves the same under continuous dividends and no dividend payments. In particular, discounted portfolio value is a martingale under risk–neutral measure. The next proposition shows that this is no more true for discounted stock price.

Proposition 14.2. ⋆

If $A_t \neq 0$, then

***(a)** The process $D_t S_t$ is not a $\widetilde{P}$–martingale.*

***(b)** However, the process $e^{\int_0^t A_u du} D_t S_t$ is a $\widetilde{P}$–martingale, and this process satisfies*

$$e^{\int_0^t A_u du} D_t S_t = S_0 \exp\left\{ \int_0^t \sigma_u \, d\widetilde{W}_u - \frac{1}{2} \int_0^t \sigma_u^2 \, du \right\}. \tag{14.11}$$

PROOF (Outline): We rewrite (14.1) on p.259 as follows

$$dS_t = (\alpha_t - A_t) S_t \, dt + \sigma_t S_t \, dW_t .$$

Clearly, S_t behaves like stock price in the ordinary generalized Black–Scholes market model, except that the mean rate of return drops from α_t to $\alpha'_t = \alpha_t - A_t$. In particular, S_t is a generalized GBM with unchanged volatility σ_t and can be explicitly written as

$$S_t := S_0 e^{X_t} = S_0 \exp\left[\int_0^t \sigma_u dW_u + \int_0^t \left(\alpha'_u - \frac{1}{2}\sigma_u^2 \right) du \right].$$

See (9.20) on p.194. From there one obtains that the process $M_t := \exp \int\limits_{u=0}^{t} D_t S_t$ equals

$$M_t = S_0 \exp\left[\int_0^t \sigma_u d\widetilde{W}_u + \int_0^t \left(\sigma_u - \frac{1}{2}\sigma_u^2 \right) du \right]. \quad ■$$

14.2 Dividends Paid at Discrete Times

We now examine the case when the stock pays its dividend not at all times t, but only at times $0 < t_1 < t_2 < \cdots < t_n < T$.

At each time t_j the stock loses value in height of the dividend that is paid. If we assume that the dividend paid at time t_j is $a_j S_{t_j}$, i.e., the dividend rate is a_j, then stock price will go down by that amount.

To work with these assumptions, we need to know how to work with continuous time processes that possess a jump at some time t^*.

Definition 14.1.

Let $t \mapsto f(t)$ be a function of time t, let t^* be a fixed time, and asume that $\lim_{t \uparrow t^*} f(t)$ exists. We write

$$f(t^*-) \; := \; \lim_{t \uparrow t^*} f(t)$$

and call this expression the **left sided limit** of f at t^*. We often use subscripts X_t rather than parenthesized time arguments for stochastic processes $X_t(\omega)$ and write X_{t^*-} for $X(t^*-)$. $\square$

We must modify the assumptions 14.1 of Chapter 14.1 (Continuously Paying Dividends) accordingly.

Assumption 14.2. Unless stated otherwise, we assume that we have a generalized Black–Scholes market as defined in Definition 13.1 (Generalized Black–Scholes market model) on p.237, with the following **modifications**.

(1) The stock pays its dividend only at the discrete times $0 < t_1 < t_2 < \cdots < t_n < T$. The **dividend rate** at time t_j is denoted by $a_j = a_j(\omega)$ We assume that those rates are $\mathfrak{F}_t$–adapted in the sense that each a_j is $\mathfrak{F}_{t_j}$–adapted. We further assume that $0 \leq a_j \leq 1$ since the dividend cannot exceed the value of the stock. We write $t_0 := 0$ and $t_{n+1} := T$, and $a_0 := a_{n+1} := 0$ in case that no dividend is paid at those dates.

(2) We assume that S_t is a generalized geometric Brownian motion for each interval $[t_j, t_{j+1}[$. The initial condition absorbs the drop in stock price:

$$dS_t = \alpha_t S_t \, dt + \sigma_t S_t \, dW_t, \quad \text{where } \alpha_t, \sigma_t \in]0, \infty[\,; \tag{14.12}$$

$$S_{t_j} = S_{t_j -} - a_j S_{t_j -} \,. \tag{14.13}$$

(4) All other items remain unchanged. In particular, this applies to the following:
• discount process D_t • market price of risk process Θ_t • Girsanov measure $\widetilde{P}$
• $\widetilde{W}_t = W_t + \int_0^t \Theta_u du$, which becomes a Brownian motion under $\widetilde{P}$. Thus,

$$dD_t = -R_t D_t \, dt\,; \quad D_0 = 1\,, \tag{14.14}$$

$$\Theta_t = \frac{\alpha_t - R_t}{\sigma_t}\,, \tag{14.15}$$

$$d\widetilde{W}_t = dW_t + \Theta_t \, dt\,; \quad \widetilde{W}_0 = 0\,. \; \square \tag{14.16}$$

Remark 14.2.

(1) Since the dividend rate at t_j is a_j, the dividend paid on a share of stock is $a_j S_{t_j-}$. Thus stock price S_{t_j} after the dividend payment is the difference

$$S(t_j) \;=\; S(t_j-) \,-\, a_j S(t_j-) \;=\; (1-a_j) S(t_j-)\,. \tag{14.17}$$

(2) If $a_j = 0$, then no dividend is paid, and $S_{t_j} = S_{t_j-}$.
(3) If $a_j = 1$, then the full value of the asset is paid, and $S_t = 0$ for all $t \geq t_j$. □

Proposition 14.3.

The value of a self–financing portfolio has the same dynamics as in the case of no dividends or a continuously paid dividend. See Proposition 14.1 on p.260

$$dV_t \;= R_t V_t\,dt \,+\, Y_t S_t \sigma_t \big(\Theta_t\,dt \,+\, dW_t\big) \;=\; R_t V_t\,dt \,+\, Y_t S_t \sigma_t\,d\widetilde{W}_t\,, \tag{14.18}$$

$$d\big(D_t V_t\big) \;= Y_t D_t S_t \sigma_t\,d\widetilde{W}_t\,. \tag{14.19}$$

In particular, discounted portfolio value $D_t V_t$ is a $\widetilde{P}$–martingale, and risk–neutral validation still applies:

$$D_t \Pi_t(\mathcal{X}) \;= D_t V_t \;=\; \widetilde{E}\big[D_T \mathcal{X} \mid \mathfrak{F}_t\big] \quad \textit{for}\;\; 0 \leq t \leq T\,,$$
$$\textit{thus}\quad \Pi_t(\mathcal{X}) \;= \widetilde{E}\big[D_t^{-1} D_T \mathcal{X} \mid \mathfrak{F}_t\big] \;=\; \widetilde{E}\big[e^{-\int_t^T R_u du} \mathcal{X} \mid \mathfrak{F}_t\big] \quad \textit{for}\;\; 0 \leq t \leq T\,.$$

PROOF: ⋆ For the proof see SCF2 ch.5.5.2. ■

14.3 Constant Dividend Rates

First the continuous time case.

Assumption 14.3.

We not only assume that $a := A_t(\omega)$ is constant in t and ω, but that the same is true for $r := R_t$, $\alpha := \alpha_t$, $\sigma := \sigma_t$. In other words, we have a classical Black–Scholes market as in Chapter 10 (Black–Scholes Model Part I: The PDE). □

In the case of no divdidends we had seen in Subchapter 10.5 (The Black–Scholes PDE for a European Call) that the pricing function of a European call is

$$c(t,x) \;= x N\big(d_+(T-t,x)\big) \,-\, K e^{-r(T-t)} N\big(d_-(T-t,x)\big), \;\; 0 \leq t < T,\, x > 0, \tag{14.20}$$

where

$$d_\pm(\tau,x) \;= \frac{1}{\sigma\sqrt{\tau}} \left[\log\frac{x}{K} \,+\, \left(r \pm \frac{\sigma^2}{2}\right)\tau\right], \tag{14.21}$$

Here is the main result in the case of continuous and constant dividend payments with rate a.

We derived in Chapter 13.4 (Risk–Neutral Pricing of a European Call) the formula

$$\pi(t,x) \;=\; x\,N\big(d_+(\tau,x)\big) \;-\; e^{-r\tau}K\,N\big(d_-(\tau,x)\big).$$

for the pricing function of a European call. See Theorem 13.2 on p.248. This was for a stock that does not pay a dividend. We now derive the corresponding formula for the the case of a constant dividend rate a. The proof is very similar to that of the no dividend case. Accordingly, there will be quite a few references to Chapter 13.4.

To achieve our goal, let $0 \le t \le T$ be a fixed time, and

$$\tau \;:= T \,-\, t, \qquad r' \;:=\; r - a, \qquad Y \;:=\; -\,\frac{\widetilde{W}_T \,-\, \widetilde{W}_t}{\sqrt{\tau}}, \tag{14.22}$$

$$h(t;x,y) \;:= e^{-r\tau}\left(x\cdot\exp\left\{-\sigma\sqrt{\tau}y \,+\, \left(r' \,-\, \frac{\sigma^2}{2}\right)\tau\right\} \,-\, K\right)^+. \tag{14.23}$$

Note that Y is standard normal w.r.t. $\widetilde{P}$ since $\widetilde{W}_t, t \ge 0$, is a $\widetilde{P}$–Brownian motion.

We next adapt Lemma 13.1, Lemma 13.2, Lemma 13.3 to the presence of a nonzero dividend rate.

Lemma 14.1. *With the above definitions we can express the risk–neutral valuation formula for a European call as follows.*

$$\widetilde{E}\left[e^{-r\tau}(S_T - K)^+\,\middle|\,\mathfrak{F}_t\,\right] \;= \widetilde{E}\left[h(t;S_t,Y)\,\middle|\,\mathfrak{F}_t\,\right] \tag{14.24}$$

PROOF: According to (13.10) on p.240,

$$S_t \;= S_0\,\exp\left\{\int_0^t \sigma_s\,d\widetilde{W}_s \,+\, \int_0^t \left((R_s - A_s) \,-\, \frac{1}{2}\,\sigma_s^2\right)ds\right\} \;=\; S_0\,\exp\left\{\sigma\,\widetilde{W}_t \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,t\right\}.$$

For $t = T$, we obtain similarly that $S_T \;=\; S_0\,\exp\left\{\sigma\,\widetilde{W}_T \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,T\right\}$. Thus

$$\begin{aligned}\frac{S_T}{S_t} \;&= \exp\left\{\left[\sigma\,\widetilde{W}_T \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,T\right] \,-\, \left[\sigma\,\widetilde{W}_t \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,t\right]\right\}\\ &= \exp\left\{\sigma\,(\widetilde{W}_T \,-\, \widetilde{W}_t) \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,(T-t)\right\},\end{aligned}$$

thus

$$\begin{aligned}S_T \;&= S_t\cdot\exp\left\{\sigma\,(\widetilde{W}_T \,-\, \widetilde{W}_t) \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,(T-t)\right\}\\ &= S_t\cdot\exp\left\{-\sigma\tau\,\frac{-(\widetilde{W}_T \,-\, \widetilde{W}_t)}{\tau} \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,(T-t)\right\}\\ &\overset{(14.22)}{=}\; S_t\cdot\exp\left\{-\sigma\tau\,Y \,+\, \left(r' \,-\, \frac{1}{2}\,\sigma^2\right)\,(T-t)\right\}.\end{aligned}$$

It follows from that equation for S_T that

$$\begin{aligned}h(t;S_t,Y) \;&\overset{(14.23)}{=}\; e^{-r\tau}\left(S_t\cdot\exp\left\{-\sigma\sqrt{\tau}Y \,+\, \left(r' \,-\, \frac{\sigma^2}{2}\right)\tau\right\} \,-\, K\right)^+\\ &= e^{-r\tau}(S_T \,-\, K)^+.\end{aligned}$$

We apply conditional expectations $\widetilde{E}[\cdots \mid \mathfrak{F}_t]$ to both sides and assertion (14.24) follows. ■

Our goal is to find a function $(t,x) \mapsto \pi(t,x)$ such that $\Pi_t(\mathcal{X}) = \pi(t, S_t)$, i.e.,

$$\pi\big(t, S_t\big) \;= \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+ \,\middle|\, \mathfrak{F}_t\right]. \tag{14.25}$$

Lemma 14.1 allows us to reformulate this problem as follows: Let $h(t;x,y)$ be the function given in formula (14.23). We want to find a function $(t,x) \mapsto \pi(t,x)$ such that

$$\pi(t, S_t) \;= \widetilde{E}\left[h(t; S_t, Y) \,\middle|\, \mathfrak{F}_t\right]. \tag{14.26}$$

The next lemma shows how to define this function $c(t,x)$.

Lemma 14.2. *Let*

$$\pi(t,x) \;:= \widetilde{E}[h(t;x,Y)]\,, \tag{14.27}$$

where $h(t;x,y)$ is the function defined in (14.23). Then $\pi(t,S_t)$ satisfies (14.26) and hence, also the risk–neutral pricing formula

$$\pi(t, S_t) \;=\; \widetilde{E}\left[e^{-r\tau}(S_T - K)^+ \,\middle|\mathfrak{F}_t\right]. \tag{14.28}$$

PROOF: We fix $0 \le t \le T$. Since S_t is $\mathfrak{F}_t$–measurable and $Y = -\frac{\widetilde{W}_T - \widetilde{W}_t}{\sqrt{\tau}}$ is, as a function of the Brownian increment $\widetilde{W}_T - \widetilde{W}_t$, independent of $\mathfrak{F}_t$, it follows for each tixed $0 \le t \le T$ from the Independence Lemma (Lemma 5.7 on p.121) [53] that

$$\pi(t, S_t) \;= \widetilde{E}\left[h(t; S_t, Y) \,\middle|\, \mathfrak{F}_t\right].$$

This proves the validity of (14.26). We apply Lemma 14.1, and (14.28) follows. ■

We have shown the following. If $\mathcal{X}$ is a European call which is based on a stock which pays a continuous dividend at the rate a, then the function $\pi(t,x) \;=\; \widetilde{E}[h(t;x,Y)]$ allows us to price that option, at time t, by means of the risk–neutral pricing formula

$$\Pi_t(\mathcal{X}) \;=\; \pi\big(t, S_t\big) \;= \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+ \,\middle|\, \mathfrak{F}_t\right]. \tag{14.29}$$

It follows from the definition of $h(t;x,y)$ given in (14.23) that

$$\pi(t,x) \;= \widetilde{E}[h(t;x,Y)] \;=\; \widetilde{E}\left[e^{-r\tau}\left(x \cdot \exp\left\{-\sigma\sqrt{\tau}Y \,+\, \left(r' \,-\, \frac{\sigma^2}{2}\right)\tau\right\} \,-\, K\right)^+\right].$$

This is an ordinary expected value of a function which depends on ω only by means of the $\widetilde{P}$–standard normal random variable Z. This we have learned to work with and we are able to obtain

[53] There we wrote $h(x,y)$ instead of $h(t;x,y)$, and $g(x) = E[h(x,Y)]$ instead of $\pi(t,x) \;= \widetilde{E}[h(t;x,Y)]$.

a concrete representation of $\pi(t,x)$ by computing this expected value. We redefine the symbols $d_-(\tau,x)$ and $d_+(\tau,x)$, defined in Theorem 10.1 on p.210:

$$d_\pm(\tau,x) \ = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right],$$

to take into account the dividend rate a, as follows:

$$d_\pm(\tau,x) \ := \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left((r-a) \pm \frac{\sigma^2}{2}\right)\tau\right], \tag{14.30}$$

Lemma 14.3. *The pricing function $\pi(t,x)$ for a European call option on a stock which pays a constant, continuous dividend rate a, is*

$$\pi(t,x) \ = xe^{-a\tau}N\big(d_+(\tau,x)\big) \ - \ e^{-r\tau}K\,N\big(d_-(\tau,x)\big). \tag{14.31}$$

PROOF: It is true for any random variable U with a $\widetilde{P}$–density $f_U(u)$, and for any deterministic (measurable) function $u \mapsto \varphi(u)$, that $\widetilde{E}[\varphi(U)] \ = \ \int\limits_{-\infty}^{\infty}\varphi(u)f_U(u)\,du$.

We apply this to the random variable Y which has density $f_Y(y) \ = \ \frac{1}{\sqrt{2\pi}}e^{-y^2/y}$ since it is standard normal, and to the function $h(t;x,Y)$ of Y. We obtain

$$\begin{aligned}\pi(t;x) \ &\overset{(14.27)}{=} \ \widetilde{E}[h(t;x,Y)] \ = \int_{-\infty}^{\infty} h(t;x,y)\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{y^2}{2}}\,dy \\ &\overset{(14.23)}{=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-r\tau}\left(x\cdot\exp\left\{-\sigma\sqrt{\tau}y \ + \ \left(r' \ - \ \frac{\sigma^2}{2}\right)\tau\right\} \ - \ K\right)^+ \ e^{-\frac{y^2}{2}}\,dy\,.\end{aligned}$$

Since the function $u \ \mapsto \ \log(u)$ is strictly increasing: $u \ < \ u' \ \Leftrightarrow \ \log u \ < \ \log u'$, and since always $e^{-r\tau} > 0$, the integrand is positive (i.e., not zero) if and only if

$$\begin{aligned} &\log x \ + \ \left\{-\sigma\sqrt{\tau}y \ + \ \left(r' \ - \ \frac{\sigma^2}{2}\right)\tau\right\} \ > \ \log K \\ \Leftrightarrow \ &\log x \ - \ \log K \ + \ \left(r' \ - \ \frac{\sigma^2}{2}\right)\tau \ > \ \sigma\sqrt{\tau}y \\ \Leftrightarrow \ &\sigma\sqrt{\tau}y \ < \ \log\left(\frac{x}{K}\right) \ + \ \left(r' \ - \ \frac{\sigma^2}{2}\right)\tau \\ \Leftrightarrow \ &y \ < \ \frac{1}{\sigma\sqrt{\tau}}\left[\log\left(\frac{x}{K}\right) \ + \ \left(r' \ - \ \frac{\sigma^2}{2}\right)\tau\right] \ \overset{(14.30)}{=} \ d_-(\tau,x)\,.\end{aligned} \tag{14.32}$$

Therefore,

$$\pi(t,x) \ = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)} e^{-r\tau}\left(x\exp\left\{-\sigma\sqrt{\tau}y \ + \ \left(r' \ - \ \frac{1}{2}\sigma^2\right)\tau\right\} \ - \ K\right)e^{-\frac{1}{2}y^2}\,dy\,.$$

Since $r' = r - a$, and thus,

$$\begin{aligned} &e^{-r\tau}\,x\,e^{-\sigma\sqrt{\tau}y + \left(r' - \frac{1}{2}\sigma^2\right)\tau}e^{-\frac{y^2}{2}} \ = \ e^{-r\tau}\,x\,e^{-\sigma\sqrt{\tau}y + \left(r - a - \frac{\sigma^2}{2}\right)\tau}e^{-\frac{y^2}{2}} \\ &\quad = xe^{-a\tau}\,e^{-r\tau}\,e^{-\sigma\sqrt{\tau}y}e^{r\tau}\,e^{-\frac{\sigma^2}{2}\tau}e^{-\frac{y^2}{2}} \ = \ xe^{-a\tau}\,e^{-\frac{\sigma^2}{2}\tau}e^{-\sigma\sqrt{\tau}y}\,e^{-\frac{y^2}{2}}\,, \ = \ xe^{-a\tau}\,e^{-\frac{1}{2}\left(y+\sigma\sqrt{\tau}\right)^2}\,,\end{aligned}$$

it follows that

$$\begin{aligned}\pi(t,x) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)} xe^{-a\tau}\, e^{-\frac{1}{2}\left(y+\sigma\sqrt{\tau}\right)^2} dy \;-\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)} e^{-r\tau}Ke^{-\frac{1}{2}y^2}\,dy\\ &= \frac{xe^{-a\tau}}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)} e^{-\frac{1}{2}\left(y+\sigma\sqrt{\tau}\right)^2} dy \;-\; e^{-r\tau}K\,N\big(d_-(\tau,x)\big)\,.\end{aligned}$$

The last equation holds, because $N(\alpha) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\alpha} e^{-\frac{1}{2}y^2}dy$ is true for all $\alpha\in\mathbb{R}$.

In the last integral, we substitute $u := y+\sigma\sqrt{\tau}$. Then $du = dy$, and the integration bounds change from $-\infty$ and $d_-(\tau,x)$ to $-\infty$ and $d_-(\tau,x)+\sigma\sqrt{\tau}$. One easily sees from (14.30) that the formula $d_+(\tau,x) = d_-(\tau,x) + \sigma\sqrt{\tau}$ which had been established previously for the case $a=0$ remains valid, and it follows that

$$\begin{aligned}\pi(t,x) &= \frac{xe^{-a\tau}}{\sqrt{2\pi}}\int_{-\infty}^{d_-(\tau,x)+\sigma\sqrt{\tau}} \exp\left\{-\frac{z^2}{2}\right\} dz \;-\; e^{-r\tau}KN\big(d_-(\tau,x)\big)\\ &= xe^{-a\tau}\,N\big(d_+(\tau,x)\big) \;-\; e^{-r\tau}K\,N\big(d_-(\tau,x)\big)\,.\end{aligned}$$

We have proven formula (14.31). ■

We have collected the necessary tools to prove the next proposition.

Proposition 14.4. *Under the assumptions 14.3, the pricing process V_t for European call can be written as a function $c(t,S_t)$ of time t and stock price S_t where $c(t,x)$ is the following function:*

(14.33) $$c(t,x) = xe^{-a\tau}N\big(d_+(\tau,x)\big) \;-\; Ke^{-r\tau}N\big(d_-(\tau,x)\big)\,.$$

Here $0\le t<T$, $x>0$, $\tau = T-t$ and, ***differently from 14.21****,*

(14.34) $$d_\pm(\tau,x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r-a\pm\frac{\sigma^2}{2}\right)\tau\right].$$

As usual N is the cumulative standard normal distribution

(14.35) $$N(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{z^2}{2}}\,dz = \frac{1}{\sqrt{2\pi}}\int_{-y}^{\infty} e^{-\frac{z^2}{2}}\,dz.$$

PROOF: Follows from Lemma 14.1, Lemma 14.2, and Lemma 14.3. ■

Now we switch to discrete time dividend payments.

Assumption 14.4. We replace the assumptions 14.3 with the following.

(a) We assume that the processes $r := R_t,\ \alpha := \alpha_t,\ \sigma := \sigma_t,$ are constant in t and ω.

In addition, we now also have finite list of discrete time dividend rates a_j, as we had defined in the assumptions 14.2 of Subchapter 14.2 (Dividends Paid at Discrete Times).

(b) We assume that those rates a_j are deterministic.

Note that **(a)** implies that we have a classical Black–Scholes market as in Chapter 10 (Black–Scholes Model Part I: The PDE). □

Under these assumption we will derive, for a European call, the price $\Pi_0(\mathcal{X})$ at time zero.

Proposition 14.5.

Under the assumptions 14.4, the price at time zero for a European call is

$$\Pi_0(\mathcal{X}) = S_0 \prod_{j=0}^{n} (1 - a_{j+1}) N(d_+^*) - K e^{-r(T)} N(d_-^*)\,, \tag{14.36}$$

$$\text{where} \quad d_\pm^* = \frac{1}{\sigma\sqrt{T}} \left[\log \frac{S_0}{K} + \sum_{j=0}^{n-1} \log(1 - a_{j+1}) + \left(r \pm \frac{\sigma^2}{2} \right) T \right]. \tag{14.37}$$

As usual, N is the cumulative standard normal distribution

$$N(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{z^2}{2}}\,dz = \frac{1}{\sqrt{2\pi}} \int_{-y}^{\infty} e^{-\frac{z^2}{2}}\,dz. \tag{14.38}$$

For the proof see SCF2 ch.5.5.1. ■

Remark 14.3. A similar formula holds for the call price at times t between 0 and T. In those cases, one includes only the terms $(1 - a_{j+1})$ corresponding to the dividend dates between times t and T. □

Remark 14.4. The software suggested earlier to calculate the parameters for Black–Scholes contract functions also handles the case of a constant, continuous dividend:

- **a.** Magnimetrics Excel implementation:
  https://magnimetrics.com/black-scholes-model-first-steps/
- **b.** Drexel U Finance calculator:
  https://www.math.drexel.edu/~pg/fin/VanillaCalculator.html
- **b.** EasyCalculation.com:
  https://www.easycalculation.com/statistics/black-scholes-mode.php □

14.4 Forward Contracts and Zero Coupon Bonds

We now assume that a dividend is **NOT paid** for the stock, thus discounted stock price D_tS_t is a martingale under the Girsanov measure $\widetilde{P}$ and $\widetilde{P}$ is a genuine risk–neutral measure. We also assume that $\bar{T}$ is a time so large, that all securities we consider in this chapter will have an expiration date before $\bar{T}$.

When we speak of having bought a \$100 zero–coupon bond with a maturity date T, then we mean that we bought a bond which will pay us \$100 at time T without paying any interest beforehand. We will follow SCF2 and think of this as owning 100 zero coupon bonds which pay one dollar each at time T.

Definition 14.2.

- A **zero–coupon bond** is a contingent claim with contract value $\mathcal{X} = 1$ at time T. We call T the **maturity date** of the zero–coupon bond.
- We write
$$B(t,T) \quad (0 \le t \le T \le \bar{T}),$$
for the price of such a zero–coupon bond at time t. □

Proposition 14.6.

If $\widetilde{P}$ is a risk–neutral probability, then $D_tB(t,T)$ is a $\widetilde{P}$–martingale, and

$$B(t,T) \;:=\; \frac{1}{D_t}\widetilde{E}[D_T \mid \mathfrak{F}_t], \;\; \text{for } 0 \le t \le T \le \bar{T}. \tag{14.39}$$

PROOF: Formula (14.39) is risk–neutral validation for a contingent claim with constant value 1 at T. Thus,

$$D_tB(t,T) \;=\widetilde{E}[D_T \mid \mathfrak{F}_t]$$

is a martingale, since conditioning with respect to $\mathfrak{F}_t$ is done on an ordinary random variable which is constant in t. ■

We modify Definition 14.3 (Forward price For$_t$) on p.269 by including the expiration date and price process of the underlying risky asset into the symbol of the forward price.

Definition 14.3 (Forward price)**.**

Given is a forward contract with a strike price K (set at time 0) at expiration date T.

$$\text{For}_S(t,T), \text{ the } T\text{–}\textbf{forward price} \text{ of the underlying asset with price } S = S_t \text{ at time } t,$$

is that strike price, re-evaluated at t, for which the forward contract would have value zero at time t. □

The following is SCF2, Theorem 5.6.2.

Theorem 14.2. ⋆

Assume that there is unlimited liquidity in the market for zero–coupon bonds with maturity dates before $\bar{T}$. Let $\mathcal{X}$ be a forward contract with expiration date $T \le \bar{T}$ for an underlying asset with price S_t. Then the following holds, regardless of the strike price of that contract.
The T–forward price For$_t$ at time t is

$$\text{For}_S(t,T) \;=\; \frac{S_t}{B(t,T)} \quad \text{for } 0 \le t \le T \le \bar{T}. \tag{14.40}$$

PROOF: The proof given here is the one to be found in SCF2 Remark 5.6.3.

We apply risk–neutral validation to the forward contract. Since it has strike price K, its value at time T is $\mathcal{X} = S_T - K$. Thus,

$$\begin{aligned}\Pi_t(\mathcal{X}) &= \frac{1}{D_t}\,\widetilde{E}[D_T\left(S_T - K\right) \mid \mathfrak{F}_t] \\ &= \frac{1}{D_t}\widetilde{E}[D_T S_T \mid \mathfrak{F}_t] - \frac{K}{D_t}\widetilde{E}[D_T \mid \mathfrak{F}_t]\,. \end{aligned} \qquad \textbf{(A)}$$

Note that $D_t S_t$ is a martingale under risk–neutral probability $\widetilde{P}$, and so is $D_t\Pi_t(\mathcal{X}')$, if $\Pi_t(\mathcal{X}')$ is the pricing function of a claim with contract value $\mathcal{X}' = 1$, i.e., of a zero–coupon bond with maturity T. Note that $D_T = D_T \cdot 1 = D_T\mathcal{X}'$, and that $\Pi_t(\mathcal{X}') = B(t,T)$ by the very definition of $B(t,T)$ (and Proposition 14.6). It follows from **(A)** that

$$\Pi_t(\mathcal{X}) = \frac{1}{D_t}\,D_t S_t - \frac{K}{D_t} D_t\,B(t,T) = S_t - K\,B(t,T).$$

The forward price $\mathrm{For}_S(t,T)$ was defined as that strike price K that would make the foward contract a fair deal for both parties at time t, i.e., that would result in a zero value for the price $\Pi_t(\mathcal{X})$ of that contract at time t. Thus,

$$0 = S_t - \mathrm{For}_S(t,T)\,B(t,T)\,,$$

and we have obtained (14.40). ■

14.5 Exercises for Ch.14

Exercise 14.1. Theorem 14.2 on p.269 was done by means of a risk–neutral measure argument. In SCF2 a proof of this theorem (Theorem 5.6.2 on p.241 in the book) is given by means of a no arbitrage allowed argument, but only case 1 where the "seller" of the forward contract is not allowed to make a profit is covered in detail.

The last four lines of the proof indicate what must be done for the proof of case 2: The seller cannot have a loss: »..... If it is negative, the agent could instead have taken the opposite position .....«

Give a detailed proof of that case 2 by modifying the proof of case1. □

15 Stochastic Methods for Partial Differential Equations

15.1 Stochastic Differential Equations

Definition 15.1 (Stochastic differential equation)**.** Let $W_t, t \geq 0$, be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ and let

$$\beta : [0,T] \times \mathbb{R} \to \mathbb{R}, \qquad (t,x) \mapsto \beta(t,x),$$
$$\gamma : [0,T] \to \mathbb{R}, \qquad (t,x) \mapsto \gamma(t,x),$$

be two (measurable) deterministic functions. Given are a stochastic differential and initial condition

(15.1) $$dX_t = \beta\big(t, X_t\big)\, dt + \gamma\big(t, X_t\big)\, dW_t,$$
(15.2) $$X_{t_0} = x_0, \quad \text{where } 0 \leq t_0 \leq T \text{ and } x_0 \in \mathbb{R}\,.$$

We call (15.1) a **stochastic differential equation** (short: **SDE**) with **drift coefficient** β and **diffusion coefficient** γ. We call a process $X = (X_t)_{t_0 \leq t \leq T}$ that satisfies both (15.1) and (15.2) a **solution** of the SDE (15.1) for the **initial condition** (15.2). □

A word on notation. We will often write $X_u = a$ for the initial condition. This does not look as intuitive as $X_{t_0} = x_0$, but we often will write $X_t^{u,a}$ for the SDE solution with initial condition $X_u = a$, and that is more readable than $X_t^{t_0,x_0}$.

Remark 15.1. Note that the differential $dY_t = \Theta_t dt + \Delta_t dW_t$ of an Itô process Y_t is more general than that given by (15.1), since $(t,\omega) \mapsto \Theta_t(\omega)$ and $(t,\omega) \mapsto \Delta_t(\omega)$ are merely adapted $\mathfrak{F}_t$–processes, whereas $\beta(t, X_t(\omega))$ and $\gamma(t, X_t(\omega))$ are functions of t and $X_t(\omega)$, not just of ω. □

Fact 15.1. *The SDE* (15.1)*, with an intial condition $X_u = a$, possesses a unique solution*

(15.3) $$X^{u,a} = (X_t^{u,a})_{u \leq t \leq T}$$

under very general conditions on drift $\beta(t,x)$ and diffusion $\gamma(t,x)$. □

It is absolutely OK if you skip the following technical note.

Note 15.1 (Technical note on the Markov property of SDE solutions)**.** ⋆
For $0 \leq u \leq T$ and $a \in \mathbb{R}$, let $X^{u,a}$ be the unique SDE solution of Fact 15.1. Let

(15.4) $$P(u,a,t,B) := P\{X_t^{u,a} \in B\} \quad (u \leq t \leq T, B \in \mathfrak{B}^1)\,.$$

Then $(u,a) \mapsto P(u,a,t,B)$ is measurable in u and a, and $B \mapsto P(u,a,t,B)$ is a probability measure on the Borel σ–algebra. In addition, it satisfies the so–called Chapman–Kolmogorov equations. [54] Such a function is customarily called a **Markov transition function**, a **transition probability function**, or a **transition probability** (on $\mathbb{R}$).

[54] See Definition 15.5 on p.283 of the optional subchapter 15.4 (Markov Processes With Transition Probability Functions).

Let us ignore the role of the SDE solutions $X_t^{u,a}$ in the definition of $P(\cdot,\cdot,\cdot,\cdot)$ and just think of it as a function of three real numbers and a Borel set as arguments. If $Z = Z(\omega)$ is any (real–valued) random variable, then it is perfectly fine to plug in $Z(\omega)$ for the second argument and examine the properties of the random variable $\omega \mapsto P(t_0, Z(\omega), t_1, B')$, just as long as $t_0 \leq t_1 \leq T$ and B' is a Borel set. Let $X = X^{0,x}$ be the SDE solution for $X_0 = x$.

Assume in all that follows that $0 \leq u \leq t \leq T$ and $B \in \mathfrak{B}^1$. Then it can be proven that

(15.5)
$$P\{X_t^{0,x} \in B \mid \mathfrak{F}_u\} \;=\; P\{X_t^{0,x} \in B \mid X_u^{0,x}\} \;=\; P(u, X_u^{0,x}, t, B)\,.$$

Since one and the same process $X^{0,x}$ occurs in all four places of (15.5), it is customary to drop the superscripts and write X_t for $X_t^{0,x}$. We obtain

(15.6)
$$P\{X_t \in B \mid \mathfrak{F}_u\} \;=\; P\{X_t \in B \mid X_u\} \;=\; P(u, X_u, t, B)\,.$$

We often follow SCF2 notation and write

(15.7)
$$P^{u,a}\{X_t \in B\} \;:=\; P(u,a,t,B) \;\overset{(15.4)}{=}\; P\{X_t^{u,a} \in B\}\,.$$

Recall from Definition 4.25 (Expected value of a random variable) on p.82 the connection between a probability P and the expectation E. If Z is a non–negative or P–integrable random variable, then

$$E[Z] \;=\; \int Z\,dP \;=\; \int Z(\omega)P(d\omega)\,.$$

Also recall from Definition 4.13 (Image measure) on p.66 the connection between P and the image probability (distribution) P_Z of the random variable Z, $P_Z(B) = P\{Z \in B\}$. Also recall Theorem 4.16 on p.94 which states for Borel measurable functions $g(z) (z \in \mathbb{R})$ of a random variable Z,

$$\int_\Omega g\big(Z(\omega)\big)P(d\omega) \;=\; \int_{\mathbb{R}} g(z)P_Z(dz)\,.$$

In our setting, $P(u,a,t,B) = P^{u,a}\{X_t \in B\}$ states that $P(u,a,t,\cdot) = P^{u,a}_{X_t}$ (the distribution of $X_t^{u,a}$). Since $P^{u,a}(t,\cdot)$ is a probability measure, it comes with a corresponding expectation $E^{u,a}$ which also is parametrized by t. We limit ourselves to random variables $h(X_t)$ for Borel measurable functions $h(x)$. That allows us to further abuse notation and write $E^{u,a}\big[h(X_t)\big]$ to indicate that the probability measure associated with that expectation is $P^{u,a}(t,\cdot)$. Thus,

(15.8)
$$\begin{aligned} E^{u,a}h(X_t) \;&=\; \int_\Omega h \circ X_t(\omega)P^{u,a}(d\omega) \;=\; \int_{\mathbb{R}} h(x)P^{u,a}_{X_t}(dx) \\ &=\; \int_{\mathbb{R}} h(x)P_{X_t^{u,a}}(dx) \;\overset{(15.7)}{=}\; \int_{\mathbb{R}} h(x)P(u,a,t,dx)\,. \end{aligned}$$

The second equation is the definition of the image of $P^{u,a}$ under the random variable X_t, the third equation is the relation $P^{u,a}\{X_t \in B\} = P\{X_t^{u,a} \in B\}$, which follows from (15.7). In terms of expectations, (15.6) becomes

(15.9)
$$E\{h(X_t) \mid \mathfrak{F}_u\} \;=\; E\{h(X_t) \mid X_u\} \;=\; \int_{\mathbb{R}} h(x)P(u, X_u, t, dx)\,.$$

We obtain a formula without reference to the transition probability by combining (15.8) and (15.9) and replacing the real number a with the real number $X_u(\omega)$ and then dropping as usual, the reference to ω:

(15.10)
$$E^{u,X_u}h(X_t) \;=\; E\{h(X_t) \mid \mathfrak{F}_u\} \;=\; E\{h(X_t) \mid X_u\}\,. \;\square$$

Remark 15.2. This remark is meant to provide more intuition of a Markov process as one, for which its future development does not depend on the past, only on the present. See Proposition 6.2 on p.124.

(a) In the special case where $h(x) = \mathbf{1}_B(x)$ for some Borel set B, (15.10) reads

$$P^{u,X_u}\{X_t \in B\} \;=\; P\{X_t \in B \mid \mathfrak{F}_u\} \;=\; P(u, X_u, t, B)\,. \;=\; P\{X_t \in B \mid X_u\}\,.$$

(b) We recall that X_t was just a convenience symbol which actually denotes $X_t^{0,x}$, the PDE solution which starts at time 0 in an arbitrary state x. If we happen to know that $X_u(\omega) = a$, i.e., we condition on $X_u = a$, then we obtain

$$P^{u,a}\{X_t^{0,x} \in B\} \;=\; P\{X_t^{0,x} \in B \mid \mathfrak{F}_u\} \;=\; P\{X_t^{0,x} \in B \mid X_u^{0,x} = a\} \;=\; P(u, a, t, B)\,.$$

(c) Since the expression $P(u, a, t, B)$ does not depend on x, the following must be true. No matter where the process was at $t = 0$, the probability of ending up in the set B (and thus, the entire distribution of X_t, since $B \in \mathfrak{B}^1$ was arbitrary), only depends on knowing that $X_u = a$, i.e., knowing the state of the process at time u. □

The following is SCF2 Theorem 6.3.1.

Theorem 15.1. *The original expectation $E[\dots]$ of $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ is intimately related to the expectations $E^{u,a}[\dots]$ belonging to the initial conditions (u, a) by means of conditioning:*

$$E^{u,X_u}\big[h(X_t)\big] \;=\; E\{h(X_t) \mid X_u\} \;=\; E\{h(X_t) \mid \mathfrak{F}_u\}. \tag{15.11}$$

PROOF: This is formula (15.10) of the preceding technical notes. ■

The following is SCF2 Theorem 6.4.1.

Theorem 15.2 (Feynman–Kac Theorem)**.**

Let $T > 0$. We examine again the SDE with differential (15.1) and initial conditions (15.2),

$$d\,X_t \;:= \beta\big(t, X_t\big)\,dt \;+\; \gamma\big(t, X_t\big)\,dW_t; \quad X_{t_0} = x_0 \quad \text{for } 0 \le t_0 < T,\ x_0 \in \mathbb{R}\,. \tag{15.12}$$

Let $x \mapsto \Phi(x)$ be Borel–measurable such that $E^{t,x}[\Phi\big(X_T\big)] < \infty$, for all $0 \le t \le T$ and $x \in \mathbb{R}$. Let $(t, x) \mapsto f(t, x)$ be the function

$$f(t, x) \;:=\; E^{t,x}[\Phi\big(X_T\big)] \tag{15.13}$$

Then $f(t, x)$ is a solution to the PDE plus terminal condition

$$f_t(t, x) + \beta(t, x) f_x(t, x) \;+\; \frac{1}{2}\gamma^2(t, x) f_{xx}(t, x) \;=\; 0\,, \tag{15.14}$$

$$f(T, x) \;=\; \Phi(x) \quad \text{for all } x. \tag{15.15}$$

You can find an outline of the proof in the SCF2 text. ■

The following is SCF2 Theorem 6.4.3.

Theorem 15.3 (Discounted Feynman–Kac)**.**

Let $T > 0$. We examine again the SDE with differential (15.1) and initial conditions (15.2),

(15.16) $$d\,X_t \;:=\; \beta\big(t, X_t\big)\,dt \;+\; \gamma\big(t, X_t\big)\,dW_t; \quad X_{t_0} = x_0 \;\; \textit{for} \;\; 0 \leq t_0 < T,\; x_0 \in \mathbb{R}\big)\,.$$

Let $x \mapsto \Phi(x)$ be Borel–measurable such that $E^{t,x}[\Phi\big(X_T\big)] < \infty$, for all $0 \leq t \leq T$ and $x \in \mathbb{R}$. Let $(t,x) \mapsto f(t,x)$ be the function

(15.17) $$f(t,x) \;:=\; E^{t,x}[e^{-r(T-t)}\Phi\big(X_T\big)]$$

Then $f(t,x)$ is a solution to the following PDE plus terminal condition

(15.18) $$f_t(t,x) + \beta(t,x) f_x(t,x) \;+\; \frac{1}{2}\gamma^2(t,x) f_{xx}(t,x) - r f(t,x) \;=\; 0,$$

(15.19) $$f(T,x) \;=\; \Phi(x) \;\; \textit{for all } x.$$

You can find an outline of the proof in the SCF2 text. ■

Remark 15.3. The two Feynman–Kac theorems are general theorems which relate the solution of an SDE to that of an associated PDE + terminal condition. In stochastic finance we do option pricing by means of risk–neutral validation, and we need a suitable setup in the model. Here is a very important case.

- The SDE describes the dynamics $dS_t = \dots$ of stock price.
- The PDE solution $f(t,x)$ will be the arbitrage free price $\Pi_t(\mathcal{X})$, at time t, of a simple claim $\mathcal{X} = \Phi(S_T)$), given that stock price at t is $S_t = x$,
- The terminal condition $f(T,x) = \Phi(x)$ will be the contract function of $\mathcal{X}$.
- $f(t,x) \;=\; E^{t,x}[e^{-r(T-t)}\Phi\big(X_T\big)]$ is guaranteed to be the solution of the PDE $f_t + \beta f_x + \frac{1}{2}\gamma^2 f_{xx} - r f = 0,$ but what is that good for if $E[...]$ is not risk neutral measure, and $E^{t,S_t}[e^{-r(T-t)}\Phi\big(X_T\big)]$ is NOT the arbitrage free price $\Pi_t(\mathcal{X})$ of the option?

So the following must be done: Find the market price of risk process Θ_t to find $\widetilde{P}$ and $\widetilde{W}_t$ and rewrite the dynamics

$$dS_t \;=\; \beta(t, S_t)\,dt \;+\; \gamma(t, S_t)\,dW_t,$$

with new coefficients β' and γ', and with the $\widetilde{P}$– Brownian motion $\widetilde{W}_t$:

$$dS_t \;=\; \beta'(t, S_t)\,dt \;+\; \gamma'(t, S_t)\,d\widetilde{W}_t.$$

Now (discounted) Feyman Kac gives you the correct PDE

$$f_t(t,x) + \beta'(t,x) f_x(t,x) \;+\; \frac{1}{2}\gamma'^2(t,x) f_{xx}(t,x) - r f(t,x) \;=\; 0 \quad \text{for} \;\; 0 \leq t_0 < T,\; x_0 \in \mathbb{R}\,.$$
$$f(T,x) \;=\; \Phi(x) \quad \text{for all } x\,,$$

for which the solution, $f(t,x) = \widetilde{E}^{t,x}[e^{-r(T-t)}\Phi\big(X_T\big)]$, does the desired: $\Pi_t(\mathcal{X}) = f(t, S_t)$.
Examples for this are SCF2 Example 6.4.4 - Options on a geometric Brownian motion, and the interest rate models of SCF2 Chapter 6.5. $\square$

15.2 Interest Rates Driven by Stochastic Differential Equations

Given is a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ with a risk–neutral probability $\widetilde{P}$ and an $\mathfrak{F}_t$–adapted Brownian motion $\widetilde{W}$ under $\widetilde{P}$.
We assume we have a market model in which the interest rate $R_t(\omega)$ is a stochastic process, but not of the most general kind, i.e., just $\mathfrak{F}_t$–adapted and nothing more. We rather assume that R_t is modeled by a stochastic Differential Equation

(15.20) $$d\,R_t \;= \beta\big(t, R_t\big)\,dt \;+\; \gamma\big(t, R_t\big)\,d\widetilde{W}_t.$$

Since interest rates for short–term borrowing are modeled by such an SDE we speak of a **short–rate model** for R_t. Very simple models for fixed income markets fall into this category.
We recall from Definition 7.4 (Discount process) on p.156, that

$$B_t \;=\; \exp\left\{\int_0^t R_s\,ds\right\}$$

is the money market account price process and

$$D_t \;=\; \frac{1}{B_t} \;= \exp\left\{-\int_0^t R_s\,ds\right\}$$

is the discount process of the bank account.
Clearly, the dynamics of those processes are

$$dD_t \;=\; -\,R_t D_t\,dt, \qquad dB_t \;= B_t\,R_t\,dt.$$

We saw in Chapter 14.4 (Forward Contracts and Zero Coupon Bonds) that a zero–coupon bond with maturity date T is a contingent claim with constant contract value $V_T = 1$ and that the (arbitrage free) price $B(t,T)$ at time $0 \le t \le T$ is, under risk–neutral probability $\widetilde{P}$,

(15.21) $$B(t,T) \;= \frac{1}{D_t}\,\widetilde{E}[D_T \mid \mathfrak{F}_t] \;=\; \widetilde{E}[e^{-\int_t^T R_s\,ds} \mid \mathfrak{F}_t]\,.$$

Definition 15.2 (Yield)**.**

We define the zero–coupon bond **yield** between times t and T as

$$Y(t,T) \;:=\; -\,\frac{1}{T-t}\,\log B(t,T) \;\;\square$$

Remark 15.4. Formula (15.2) is equivalent to

(15.22) $$B(t,T) \;=\; e^{-Y(t,T)\,(T-t)}.$$

In other words, $Y(t,T)$ is that constant rate of continuously compounding interest between times t and T which corresponds to the price $B(t,T)$ of a zero–coupon bond maturing at T. $\square$

Proposition 15.1.

Given the dynamics of (15.20) for the interest rate R_t, one can write $B(t,T) = f(t, R_t)$. Here $f(t,x)$ is a function of time t and $x \geq 0$ which satisfies the PDE plus terminal condition

$$f_t(t,x) + \beta(t,x)\, f_x(t,x) + \frac{1}{2}\gamma^2(t,x)\, f_{xx}(t,x) = x f(t,x)\,, \tag{15.23}$$

$$f(T,x) = 1 \text{ for all } x. \tag{15.24}$$

For the proof see SCF2 Chapter 6.5. ■

15.3 Stochastic Differential Equations and their PDEs in Multiple Dimensions

As in Chapter 13.6 (Multidimensional Financial Market Models), the material discussed here can be generalized to SDEs, in which an m–dimensional processes $\vec{X}_t = \big(X_t^{(1)}, \dots, X_t^{(m)}\big)$ is driven by a d–dimensional Brownian motion $\vec{W}_t = \big(W_t^{(1)}, \dots, W_t^{(d)}\big)$. However, the notation is complicated enough when we restrict ourselves to a two dimensional process $\vec{X}_t = \big(X_t, \dots, Y_t\big)$ which is driven by a 2–dimensional Brownian motion $\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)}\big)$. Doing so will drastically reduce the amount of superscripts you will encounter.

Definition 15.3. Let $\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)}\big)$, $t \geq 0$, be a two dimensional Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$, and let

$$\beta_1,\, \beta_2,\, \gamma_{11},\, \gamma_{12},\, \gamma_{21},\, \gamma_{22},\, :\; [0,T] \times \mathbb{R}^2 \to \mathbb{R}\,,$$

be six (measurable) deterministic functions $\beta_i(t,x,y)$, $\gamma_{ij}(t,x,y)$, where $i,j = 1,2$.
Given are the stochastic differentials and initial conditions

$$\begin{aligned} dX_t &= \beta_1\big(t, X_t, Y_t\big)\, dt + \gamma_{11}\big(t, X_t, Y_t\big)\, dW_t^{(1)} + \gamma_{12}\big(t, X_t, Y_t\big)\, dW_t^{(2)}\,, \\ dY_t &= \beta_2\big(t, X_t, Y_t\big)\, dt + \gamma_{21}\big(t, X_t, Y_t\big)\, dW_t^{(1)} + \gamma_{22}\big(t, X_t, Y_t\big)\, dW_t^{(2)}\,, \end{aligned} \tag{15.25}$$

$$X_{t_0} = x_0\,, \quad Y_{t_0} = y_0\,, \quad \text{where } 0 \leq t_0 \leq t \leq T \text{ and } x_0, y_0 \in \mathbb{R}\,. \tag{15.26}$$

We call (15.25) a **stochastic differential equation** (short: **SDE**) with **drift vector** $\vec{\beta} = \big(\beta_1, \beta_2\big)$, and **diffusion matrix** $\gamma^{**} = \big(\gamma_{ij}\big)_{ij}$, where $i = 1,2$, $j = 1,2$.
We call a process $\vec{X} = (X_t, Y_t)_{t_0 \leq t \leq T}$ that satisfies both (15.25) and (15.26) a **solution** of the SDE (15.25) for the **initial condition** (15.26). □

Similar to the onedimensional case we often define $\vec{X}_t = (X_t, Y_t)$, $\vec{a} = (a,b)$, and write

$$\vec{X}_u = \vec{a}\,, \quad \text{i.e.,} \quad X_u = a\,, \;\; Y_u = b$$

for the initial condition. Again, this is done to improve readability of superscripts.

Fact 15.2. *The SDE (15.25), with an intial condition $\vec{X}_u = \vec{a}$, possesses a unique solution*

$$\vec{X}^{u,\vec{a}} \;=\; \left(\vec{X}_t^{u,\vec{a}}\right)_{u\leq t\leq T} \tag{15.27}$$

under very general conditions on drift vector $\vec{\beta} = (\beta_1, \beta_2)$ and diffusion matrix $\gamma^{**} = \left(\gamma_{ij}\right)_{ij}$. □

Note 15.1 on p.271 generalizes to the multidimensional case. It follows next. Feel free to skip this note. If you study it, be sure to remember the concepts discussed in Note15.1.

Note 15.2 (Technical note on the Markov property of SDE solutions). ⋆

For $0 \leq u \leq T$ and $a \in \mathbb{R}$, let $\vec{X}^{u,\vec{a}}$ be the unique SDE solution of Fact 15.2. Let

$$P(u,\vec{a},t,B) \;:=\; P\{\vec{X}_t^{u,\vec{a}} \in B\} \quad (u \leq t \leq T, B \in \mathfrak{B}^2)\,. \tag{15.28}$$

Then $(u,\vec{a}) \mapsto P(u,\vec{a},t,B)$ is measurable in u and $\vec{a}$, $B \mapsto P(u,\vec{a},t,B)$ is a probability measure on $\mathfrak{B}^2$, and $P(\cdot,\cdot,\cdot,\cdot)$ satisfies the Chapman–Kolmogorov equations. [55] We call such a function a **Markov transition function**, a **transition probability function**, or a **transition probability** (on $\mathbb{R}^2$).

As in the onedimensional case, we ignore the role of the SDE solutions $\vec{X}_t^{u,\vec{a}}$, and we simply consider $P(\cdot,\cdot,\cdot,\cdot)$ as a function of two time parameters, a two dimensional vector, and a Borel set as arguments. If $\vec{Z} = \vec{Z}(\omega)$ is a twodimensional random vector, then it is perfectly fine to plug in $\vec{Z}(\omega)$ for the second argument and examine the properties of the random variable $\omega \mapsto P(t_0, \vec{Z}(\omega), t_1, B')$, just as long as $t_0 \leq t_1 \leq T$ and B' is a Borel set. Let $\vec{X} = \vec{X}_t^{0,\vec{x}}$ be the SDE solution for $\vec{X}_0 = \vec{x}$.
Assume in all that follows that $0 \leq u \leq t \leq T$ and $B \in \mathfrak{B}^2$. Then it can be proven that

$$P\{\vec{X}_t^{0,\vec{x}} \in B \mid \mathfrak{F}_u\} \;=\; P\{\vec{X}_t^{0,\vec{x}} \in B \mid \vec{X}_u^{0,\vec{x}}\} \;=\; P(u,\vec{X}_u^{0,\vec{x}},t,B)\,. \tag{15.29}$$

Since one and the same process $\vec{X}_t^{0,\vec{x}}$ occurs in all four places of (15.29), it is customary to drop the superscripts and write $\vec{X}_t$ for $\vec{X}_t^{0,\vec{x}}$. We obtain

$$P\{\vec{X}_t \in B \mid \mathfrak{F}_u\} \;=\; P\{\vec{X}_t \in B \mid \vec{X}_u\} \;=\; P(u,\vec{X}_u,t,B)\,. \tag{15.30}$$

We often follow SCF2 notation and write

$$P^{u,\vec{a}}\{\vec{X}_t \in B\} \;:=\; P(u,\vec{a},t,B) \;\overset{(15.28)}{=}\; P\{\vec{X}_t^{u,\vec{a}} \in B\}\,. \tag{15.31}$$

Recall from Definition 4.13 (Image measure) on p.66 the connection between P and the image probability (distribution) $P_{\vec{Z}}$ of a twodimensional random vector $\vec{Z} = (Z_1, Z_2)$. $P_{\vec{Z}}(B) = P\{\vec{Z} \in B\}$. Also recall Theorem 4.16 on p.94 which states for Borel measurable functions $f(\vec{z})(\vec{z} = (z_1,z_2) \in \mathbb{R}^2)$ of a twodimensional random vector $\vec{Z} = (Z_1, Z_2)$,

$$\int_\Omega g\big(\vec{Z}(\omega)\big)P(d\omega) \;=\; \int_{\mathbb{R}^2} g(\vec{z})P_{\vec{Z}}(d\vec{z})\,. \;=\; \int_{\mathbb{R}^2} g(z_1,z_2)P_{(Z_1,Z_2)}(d(z_1,z_2))\,.$$

In our setting, $P(u,\vec{a},t,B) = P^{u,\vec{a}}\{\vec{X}_t \in B\}$ states that $P(u,\vec{a},t,\cdot) = P^{u,\vec{a}}_{\vec{X}_t}$ (the distribution of $\vec{X}_t^{u,\vec{a}}$).

[55] As in the onedimensional case, we refer you to Definition 15.5 on p.283 of the optional subchapter 15.4.

Since $P^{u,\vec{a}}(t,\cdot)$ is a probability measure, it comes with a corresponding expectation $E^{u,\vec{a}}$ which also is parametrized by t. We limit ourselves to random variables $h(\vec{X}_t)$ for Borel measurable functions $h(\vec{x})$. That allows us to further abuse notation and write $E^{u,\vec{a}}\big[h(\vec{X}_t)\big]$ to indicate that the probability associated with that expectation is $P^{u,\vec{a}}(t,\cdot)$. Thus,

$$\begin{aligned} E^{u,\vec{a}}h(\vec{X}_t) &= \int_\Omega h\circ\vec{X}_t(\omega)P^{u,\vec{a}}(d\omega) = \int_{\mathbb{R}} h(x)P^{u,\vec{a}}_{\vec{X}_t}(d\vec{x}) \\ &= \int_{\mathbb{R}} h(x)P_{\vec{X}_t^{u,\vec{a}}}(d\vec{x}) \stackrel{(15.31)}{=} \int_{\mathbb{R}^2} h(\vec{x})P(u,\vec{a},t,d\vec{x})\,. \end{aligned} \tag{15.32}$$

The second equation is the definition of the image of $P^{u,\vec{a}}$ under the random variable $\vec{X}_t$, the third equation is the relation $P^{u,\vec{a}}\{\vec{X}_t\in B\} = P\{\vec{X}_t^{u,\vec{a}}\in B\}$, which follows from (15.31). In terms of expectations, (15.30) becomes

$$E\{h(\vec{X}_t)\mid\mathfrak{F}_u\} = E\{h(\vec{X}_t)\mid\vec{X}_u\} = \int_{\mathbb{R}^2} h(\vec{x})P(u,\vec{X}_u,t,d\vec{x})\,. \tag{15.33}$$

We obtain a formula without reference to the transition probability by combining (15.32) and (15.33) and replacing the vector $\vec{a}$ wih the real number $X_u(\omega)$ and then dropping as usual, the reference to ω:

$$E^{u,\vec{X}_u}h(\vec{X}_t) = E\{h(\vec{X}_t)\mid\mathfrak{F}_u\} = E\{h(\vec{X}_t)\mid\vec{X}_u\}\,.\ \square \tag{15.34}$$

We generalize now the Feynman–Kac to two dimensions.

Theorem 15.4 (Two dimensional Feynman–Kac)**.**

Let $\vec{X}_t := (X_t, Y_t)$ be the solution of the SDE of Definition 15.3 on p.276. and let $(x,y)\mapsto h(x,y)$ be Borel-measurable. Corresponding to the initial condition $\vec{X}_{t'} = \vec{x}' = (x',y')$, where $0\le t'\le T$ and $x',y'\in\mathbb{R}$, we define

$$g(t',x',y') := E^{t',\vec{x}'}\,h\big(X_T,Y_T\big), \tag{15.35}$$

$$f(t',x',y') := E^{t',\vec{x}'}\left[e^{-r(T-t)}\,h\big(X_T,Y_T\big)\right] \tag{15.36}$$

Then g and f are solutions to the PDEs

$$\begin{aligned} &g_t + \beta_1 g_x + \beta_2 g_y \\ &\quad + \frac{1}{2}(\gamma_{11}^2+\gamma_{12}^2)g_{xx} + (\gamma_{11}\gamma_{21}+\gamma_{12}\gamma_{22})g_{xy} + \frac{1}{2}(\gamma_{21}^2+\gamma_{22}^2)g_{xy} = 0\,, \end{aligned} \tag{15.37}$$

$$\begin{aligned} &f_t + \beta_1 f_x + \beta_2 f_y \\ &\quad + \frac{1}{2}(\gamma_{11}^2+\gamma_{12}^2)f_{xx} + (\gamma_{11}\gamma_{21}+\gamma_{12}\gamma_{22})f_{xy} + \frac{1}{2}(\gamma_{21}^2+\gamma_{22}^2)f_{xy} = rf\,. \end{aligned} \tag{15.38}$$

Further, these PDE solutions $f(t,x,y)$ and $g(t,x,y)$ also satisfy the terminal conditions

$$g(T,x,y) = f(T,x,y) = h(x,y) \quad \text{for all } x \text{ and } y\,.$$

PROOF: See SCF2 Chapter 6.6 ■

We demonstrate the use of the multidimensional Feynman–Kac Theorem in the context of determining the price of an Asian option. This is SCF2 Example 6.6.1.

Definition 15.4.

An **Asian option** with a strike price of K is a contract written at time 0, which specifies that, at the time of expiration $T > 0$, the holder of this option will receive the amount

$$\mathcal{X} = \left(\frac{1}{T} \int_0^T S_u\, du - K \right)^+ . \tag{15.39}$$

Here, S_t is a geometric Brownian motion and $K > 0$.

Remark 15.5 (The Asian option is not Markov)**.** Because the contract value depends on the entire history from 0 to t of the stock price trajectory, $\Pi_t(\mathcal{X})$ is not a Markov process, and thus cannot be written as a function $F(t, S_t)$ of time and stock price. It should be clear that the entire history $S_u(\omega)$ for $0 \le u \le t \le T$ has a bearing on $\Pi_t(\mathcal{X})$, since a history of high stock prices drives up $\int\limits_{u=0}^{t} S_u du$ and thus makes it more likely to obtain a big payoff $\mathcal{X} = \left(\frac{1}{T} \int\limits_0^T S_u\, du - K \right)^+$. Of course, this will result in a higher option price $\Pi_t(\mathcal{X})$.

Surprisingly, if we define $A_t := \int_0^T S_u du$, the twodimensional process $\big(S_t, A_t\big)$ is Markov. This is so because we can model this process by the SDE

$$\begin{aligned} dS_t &= rS_t\, dt + \sigma S_t\, d\widetilde{W}_t\,, \\ dA_t &= S_t\, dt\,, \end{aligned} \tag{15.40}$$

with deterministic initial conditions $A_0 = 0$ and S_0. Be sure to understand the following:

Even though A_t by itself is not a Markov process, the vector process $\big(S_t, A_t\big)$ is Markov because the drift and diffusion coefficients of the SDE system (15.40) only possess S_t and A_t (and, of course, time t) as arguments. $\square$

Proposition 15.2. *Assume that we operate in a classical Black–Scholes market, i.e., we have constant interest rate $r \ge 0$ and constant volatility $\sigma > 0$.*

We specify the dynamics of S_t terms of the Brownian motion $\widetilde{W}_t$ under risk–neutral measure $\widetilde{P}$. Thus,

- *$\widetilde{W}_t$ is the process $d\widetilde{W}_t = dW_t + \Theta_t dt$, where $\Theta_t = \Theta = (\alpha - r)/\sigma$ is the market price of risk.*

Then the stochastic differential equation for S_t specifies the interest rate r rather than the stock's instantaneous rate of return, α_t, as its drift coefficient. Since the interest rate is constant,

$$dS_t = rS_t\, dt + \sigma S_t\, d\widetilde{W}_t. \tag{15.41}$$

$$dD_t = -rD_t\, dt, \quad D_0 = 1, \quad \textit{i.e.,} \quad D_t = e^{-rt}. \tag{15.42}$$

(15.43) *Let* $A_t := \int_{u=0}^{t} S_u\,du$, *i.e.,* $dA_t = S_t\,dt$, $A_0 = 0$.

Then the option price is $\Pi_t(\mathcal{X}) = \pi(t, S_t, A_t)\,,\ (0 \le t \le T)$,

where the function $(t,x,y) \mapsto \pi(t,x,y)$ *solves the partial differential equation*

(15.44) $\pi_t(t,x,y) + rx\pi_x(t,x,y) + x\pi_y(t,x,y) + \frac{1}{2}\sigma^2 x^2 \pi_{xx}(t,x,y) - r\pi(t,x,y) = 0\,,$

and satisfies at time of expiry T *the boundary condition*

(15.45) $\pi(T, S_T, A_T) = \mathcal{X} = \left(\frac{1}{T}\int_0^T S_u\,du - K\right)^+.$

First PROOF: (Outline. For details, see SCF2 Example 6.6.1.)

One can prove this proposition without using Theorem 15.4 (Two dimensional Feynman–Kac]) on p.278 by applying the Itô formula to compute the differential $d\Big(e^{-rt}\,\pi\big(t,S_t,A_t\big)\Big)$, where the Itô processes S_t, A_t are defined by the SDE system, (15.40), and the function $\pi(t,x,y)$ is implicitly defined as follows:

$$\pi\big(t, S_t, A_t\big) = \Pi_t(\mathcal{X}) = \widetilde{E}\left[e^{-r(T-t)}\left(\frac{1}{T}A_T - K\right)^+ \,\middle|\, \mathfrak{F}_t\right].$$

Such a function must exist due to the Markovian nature of the process $\big(S_t, A_t\big)$. One obtains from Corollary 11.2 on p.224, followed by the use of Itô's formula to evaluate $d\pi\big(t,S_t,A_t\big)$,

(E)
$$\begin{aligned} d\Big(e^{-rt}\,\pi\big(t,S_t,A_t\big)\Big) &= e^{-rt}\Big[-r\pi(\cdot,\cdot,\cdot) + \pi_t + \pi_x r S_t + \pi_y S_t + \frac{1}{2}\sigma^2 S_t^2 \pi_{xx}\Big]\,dt \\ &\quad + e^{-rt}\,\sigma S_t \pi_x\, d\widetilde{W}_t\,. \end{aligned}$$

We wrote $\pi(\cdot,\cdot,\cdot)$ to avoid confusion with the number π, and we omitted the arguments everywhere else. One shows that $e^{-rt}\pi\big(t,S_t,A_t\big)$ is a martingale. As a consequence, the dt term of **(E)** vanishes. Replacing S_t with x one obtains (15.44). Since the expressioni under the conditional expectation is $\mathfrak{F}_T$–measurable, and $r(T-T)=0$,

$$\pi\big(T, S_T, A_T\big) = \widetilde{E}\left[e^0\left(\frac{1}{T}A_T - K\right)^+ \,\middle|\, \mathfrak{F}_T\right]. = \left(\frac{1}{T}A_T - K\right)^+ = \mathcal{X}\,.$$

This proves (15.45). ■

Alternate proof:

This second proof is based on the multidimensional Feynman–Kac Theorem 15.4 on p.278. Let

(F)
$$h(y) := \left(\frac{1}{T}y - K\right)^+; \qquad \pi(t,x,y) := \widetilde{E}^{t,x,y}\left[e^{-r(T-t)}h(A_t)\,\middle|\,\mathfrak{F}_t\right].$$

We translate the SDE system (15.40)

$$\begin{aligned} dS_t &= rS_t\,dt + \sigma S_t\,d\widetilde{W}_t\,, \\ dA_t &= S_t\,dt\,, \end{aligned}$$

to match Definition 15.3 on p.276, since we want to apply Feynman–Kac:

$$\begin{aligned} \beta_1(t,x,y) &= rx, \quad \beta_2(t,x,y) \,=\, x, \\ \gamma_{11}(t,x,y) &= \sigma x, \quad \gamma_{12}(t,x,y) \,=\, \gamma_{21}(t,x,y) \,=\, \gamma_{22}(t,x,y) \,=\, 0. \end{aligned}$$

Then (15.38) becomes

$$\pi_t + rx\pi_x \,+\, x\pi_y \,+\, \frac{1}{2}\sigma^2 x^2 \pi_{xx} \,=\, r\pi(\cdot,\cdot,\cdot)\,.$$

This is formula (15.44) of this proposition. According to the multidimensional Feynman–Kac Theorem, the function $\pi(\cdot,\cdot,\cdot)$ is a solution to this PDE, and it satisfies

$$\pi(T,x,y) \,= h(y) \stackrel{\textbf{(F)}}{=} \left(\frac{1}{T}y \,-\, K\right)^+.$$

$$\text{Thus,} \qquad \pi\big(T, S_T, A_T\big) \,= h(A_T) \,=\, \left(\frac{1}{T}A_T \,-\, K\right)^+ \,=\, \mathcal{X}\,.$$

This proves formula (15.45) of this proposition. ■

The following remark refers back to the proof of Proposition 15.2. It is intended to deepen your understanding about hedging portfolios.

Remark 15.6. Since the dt term of **(E)** is zero, we obtain

$$d\big(e^{-rt}\,\Pi_t(\mathcal{X})\big) \,= d\Big(e^{-rt}\,\pi\big(t,S_t,A_t\big)\Big) \,=\, e^{-rt}\,\sigma S_t\,\pi_x\big(t,S_t,A_t\big)\,d\widetilde{W}_t.$$

By the pricing principle, by $e^{-rt} = D_t$, and by (13.17) on p.243,

$$d\big(e^{-rt}\,\Pi_t(\mathcal{X})\big) \,= d\big(e^{-rt}V_t\big) \,=\, e^{-rt}\,\sigma S_t Y_t\,d\widetilde{W}_t.$$

We equate the right hand sides and obtain

$$e^{-rt}\,\sigma S_t\,\pi_x\big(t,S_t,A_t\big)\,d\widetilde{W}_t \,= e^{-rt}\,\sigma S_t Y_t\,d\widetilde{W}_t\,.$$

Not surprisingly, we have again obtained the Delta hedging formula,

$$Y_t \,= \pi_x\big(t,S_t,A_t\big)\,.$$

If we sell the Asian option at time zero for $\pi(0,S_0,0)$ and use this as the initial capital for a hedging portfolio (i.e., take $V_0 := \pi(0,S_0,0)$), and at each time t adhere to the portfolio strategy in which we set

$$\text{\# of stock shares} \,=\, Y_t \,:=\, \pi_x\big(t,S_t,A_t\big),$$

then we will have

$$d\big(e^{-rt}\,V_t\big) \,= d\Big(e^{-rt}\,\pi\big(t,S_t,A_t\big)\Big)$$

for all times t, and hence

$$V_T \,= \pi\big(T,S_T,A_T\big)\big) \,=\, \left(\frac{1}{T}A_T \,-\, K\right)^+.$$

We will be able to purchase an Asian option at time T to cover our short position in the.option with the proceeds from the sale of the portfolio. In other words, this portfolio is a hedge for an Asian option.

$$\text{The delta–hedging rule,} \qquad Y_t \;=\; \partial/(\partial x)(\text{ option price})\,,$$

is the same for Asian options as for the European calls and puts (see (10.22) on p.208). But be aware that the PDE we obtained for $\pi(\cdot,\cdot,\cdot)$ is structurally different from the one for $c(t,x)$. For example, it contains a term $x\pi_y(t,x,y)$ which has no counterpart in the PDE for $c(t,x)$. $\square$

15.4 Markov Processes With Transition Probability Functions ⋆

The presentation of this material follows [11] Friedman, Avner: Stochastic Differential Equations and Applications.

Introduction 15.1. We have seen in Chapter 6.5 (Brownian Motion as a Markov Process) that one can associate with a Brownian motion W_t a transition density, i.e., a function $p(\tau,x,y)$, such that the formula (6.32),

$$E[f(W_{s+\tau}) \mid \mathfrak{F}_s] \;=\; E[f(W_{s+\tau}) \mid W_s] \;=\; \int_{-\infty}^{\infty} f(y)\,p(\tau,W_s,y)\,dy\,, \tag{15.46}$$

holds true for $s \geq 0, \tau > 0$, and nonnegative, Borel measurable $f:\mathbb{R}\to\mathbb{R}$. Now let X_t be some Markov process, not necessarily Brownian motion, which possesses a transition density $p(\tau,x,y)$. For the function $f(y)=1$ we obtain, when conditioning on $X_s = x$,

$$1 \;=\; E[1 \mid X_s = x] \;=\; \int_{-\infty}^{\infty} 1\cdot p(\tau,x,y)\,dy\,.$$

Thus, for each fixed τ and x, the assignment

$$B \mapsto P(\tau,x,B) \;:=\; \int_B p(\tau,x,y)\,dy\,,$$

defines a probability measure $P(\tau,x,\cdot)$ on the Borelsets of $\mathbb{R}$. According to (6.36),

$$P(\tau,x,B) \;=\; \int_B p(\tau,x,y)\,dy \;=\; P\{X_{s+\tau}\in B \mid X_s = x\}\,.$$

This gives $P(\tau,x,B)$ an interpretation as the probability that $X_{s+\tau}$ will land in B, given that its trajectory has value x at time s.

Brownian motion is a special kind of Markov process, since it possesses **stationary increments**, i.e., the distribution of $W_{t+\tau} - W_t$ does not change with t. We also call such a Markov process **time–homogeneous**. Time–homogeneity usually is not satisfied for the Markov processes we obtain as solutions of stochastic differential equations. If X_t is such a solution, and if the drift and/or diffusion coefficient of the SDE has time as an argument, then the distribution of $X_{t+\tau} - X_t$ will change with t. Rather than just considering $\tau = t - s$, we must keep track separately of the time s at which we condition $X_s = x$, and the later time $t = s+\tau$ at which we examine the event $X_t = B$.

A transition density for X_t should then be a function $p(s,x,t,y)$ such that the analoque of (15.46) holds:

$$E[f(X_t) \mid \mathfrak{F}_s] \;=\; E[f(X_t) \mid X_s] \;=\; \int_{-\infty}^{\infty} f(y)\, p(s, X_s, t, y)\, dy\,,$$

for $0 \le s \le t \le T$, and nonnegative, Borel measurable $f : \mathbb{R} \to \mathbb{R}$. Now,

$$B \mapsto P(s,x,t,B) \;:=\; \int_B p(s,x,t,y)\, dy$$

is a probability measure, and $P(s,x,t,B)$ can be interpreted as the probability that X_t will land in B, given that its trajectory has value x at time s. One could also say that it gives the probability that $X_s = x$ transitions into the set B at time t. This function $P(s,x,t,B)$ is the transition probability function we discussed in the technical notes 15.1 on p.271 and, for the mulitidimensional case, [56] in 15.2 on p.277. □

The observations of this introduction lead us to the definition of a Markov transition function even if no stochastic differential equations and their solution processes are involved.

Definition 15.5. Let $P(s, \vec{x}, t, B) \ge 0$ be a function of $0 \le s < t < \infty, \vec{x} \in \mathbb{R}^d, B \in \mathfrak{B}^d$, such that

- **(1)** $\vec{x} \mapsto P(s, \vec{x}, t, B)$ is $\mathfrak{B}^d$–measurable for fixed s, t, B,
- **(2)** $B \mapsto P(s, \vec{x}, t, B)$ is a probability measure for fixed $\vec{x}, s, t$,
- **(3)** For any $0 \le s < t < u < \infty, \vec{x} \in \mathbb{R}^d$, and $B \in \mathfrak{B}^d$, $P(s,x,t,B)$ satisfies the **Chapman–Kolmogorov equation**

$$\int_{\mathbb{R}^d} P(s, \vec{x}, t, d\vec{y})\, P(t, \vec{y}, u, B) \;=\; P(s, \vec{x}, u, B)\,. \tag{15.47}$$

Then we call p a **Markov transition function**, a **transition probability function**, or a **transition probability** (on $\mathbb{R}^d$). □

Example 15.1. The purpose of this example is to understand the connection between Markov transition functions and Definition 6.2 on p.124 of a Markov process.

Let $X = \left(X_t\right)_{t \ge 0}$ be a stochastic process on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ as follows.

The state space of the process is the set of n numbers $S = \{b_1, \dots, b_n\}$. Thus,

$$\sum_{j=1}^{n} P\{X_t = b_j\} \;=\; 1 \quad \text{for all } t \ge 0\,.$$

We assume that X_t is Markov. We will work with the alternate definition of such a process given in Proposition 6.2 on p.124. If $0 \le s \le t \le T$, and φ is an arbitrary, nonnegative or bounded, Borel–measurable function $x \mapsto \varphi(x)$, then

$$E[\varphi(X_t) \mid \mathfrak{F}_s] \;=\; E[\varphi(X_t) \mid X_s]. \tag{15.48}$$

[56] Yes, there are multidimensional analogues for transition densities and corresponding transition probability functions.

For $0 \le s < t$ and $i, j = 1, 2, \dots, n$, let

$$p(s,x,t,y) \ := P\{X_t = y \mid X_s = x\}\,.$$

We combine this with (15.48) and obtain that, for $X_s(\omega) = a$,

$$E[\varphi(X_t) \mid \mathfrak{F}_s] \ = E[\varphi(X_t) \mid X_s] \ = \ \sum_{y\in S} \varphi(y) p(s,a,t,y)\,. \tag{15.49}$$

We will show that

$$P(s,x,t,B) \ := \ \sum_{y\in B} p(s,x,t,y) \tag{15.50}$$

is a Markov transition probability, i.e., it satisfies the Chapman–Kolmogorov equation.

Since B is finite, integration simplifies to summation with respect to the finitely many elements $b_1, \dots, b_n$ of S. The right hand side of (15.49) exemplifies this. Thus the Chapman–Kolmogorov equation we want to prove is

$$P(u,x,t,B) \ = \ \sum_{y\in S} P(u,x,s,\{y\})\, P(s,y,t,B) \quad \text{for } 0 \le u \le s \le t,\ u \in S,\ B \subseteq S\,.$$

Since measures are additive, is suffices to show the above for singletons $B = \{z\}$, where $z \in S$. Since $P(u,x,t,\{z\}) = p(u,x,t,z)$, the last formula is equivalent to

$$p(u,x,t,z) \ = \ \sum_{y\in S} p(u,x,s,y)\, p(s,y,t,z) \quad \text{for } 0 \le u \le s \le t,\ u,z \in S\,. \tag{15.51}$$

We will show more generally that, for a nonnegative function $\varphi : S \to \mathbb{R}$,

$$\sum_{z\in S} \varphi(z) p(u,x,t,z) \ = \ \sum_{z\in S} \varphi(z) \sum_{y\in S} p(u,x,s,y)\, p(s,y,t,z)\,, \quad \text{for } \ 0 \le u \le s \le t,\ x \in S\,. \tag{15.52}$$

We obtain (15.51) from this formula by setting $\varphi := \mathbf{1}_{\{z\}}$ for arbitrary $z \in S$.

Let $0 \le u \le s \le t$ and $\varphi : S \to \mathbb{R}$. Iterated conditioning yields

$$E[\varphi(X_t) | \mathfrak{F}_u] \ = E\big[E[\varphi(X_t) \mid \mathfrak{F}_s] \,\big|\, \mathfrak{F}_u\big]\,. \tag{15.53}$$

Use of the Markov property shows that, if $a \in S$ and $X_u(\omega) = a$, the left hand side of (15.53) equals

$$E[\varphi(X_t) | X_u](\omega) \ = \ \sum_{z\in S} \varphi(z)\, P\{X_t = z \mid X_u = a\} \ = \ \sum_{z\in S} \varphi(z)\, p(u,a,t,z)\,. \tag{\textbf{LS}}$$

Even though the conditional expectation $E[\varphi(X_t) \mid X_s]$ is a function of ω, it is constant on the atoms $\{X_s = b\} = \{\omega : X_s(\omega) = b\}$, i.e., it can be written as a function

$$\psi(b) \ = \ E[\varphi(X_t) \mid X_s = b]\,.$$

Note that

$$\psi(b) \ = \sum_{z\in S} \varphi(z)\, P\{X_t = z \mid X_s = b\} \ = \ \sum_{z\in S} \varphi(z)\, p(s,b,t,z)\,. \tag{15.54}$$

If $X_u(\omega) = a$, the right hand side of (15.53) thus equals

$$\textbf{(RS)} \qquad E\big[E[\varphi(X_t)\ |X_s] \,\big|\, X_u\big] \;=\; E\big[\psi(X_s) \,\big|\, X_u\big] \;=\; \sum_{b\in S} \psi(b) P\{X_s = b \mid X_u\}$$
$$= \sum_{b\in S} \psi(b) p(u,a,s,b)\,. \;=\; \sum_{b\in S}\sum_{z\in S} \varphi(z)\, p(s,b,t,z)\, p(u,a,s,b)\,.$$

Since **(LS)** = **(RS)**, we obtain for $X_u(\omega) = a$,

$$\sum_{z\in S} \varphi(z)\, p(u,a,t,x)\,. \;=\; \sum_{b\in S}\sum_{z\in S} \varphi(z)\, p(u,a,s,b)\, p(s,b,t,z)\,.$$

This proves that (15.51) holds true, thus $P(s,x,t,B)$ satisfies the Chapman–Kolmogorov equation and is indeed a Markov transition function. □

We thus have shown the following in the previous example.

Proposition 15.3.

Any Markov process with a finite state space possesses a Markov transition function.

PROOF: See Example 15.1. ■

One could say that any reasonable process that is a Markov process is associated with a Markov transition function. We confine the next definition to real–valued processes, even though it has counterparts for multidimensional state spaces.

Definition 15.6.

Let $(\Omega, \mathfrak{F}, \mathfrak{F}_t)$ be a filtered measurable space. For each $0 \le t \le T$, let $X_t : \Omega \to \mathbb{R}$ be adapted to the filtration, i.e., X_t is $\mathfrak{F}_t$–$\mathfrak{B}$–measurable. We are reluctant to call $X = (X_t)_t$ a stochastic process, since there is no probability measure (yet). That comes now. Let $\big(P^{s,x}\big)_{s\ge 0, x\in\mathbb{R}}$ be a family of probability measures on $(\Omega, \mathfrak{F}, \mathfrak{F}_t)$.

Thus, for each time $s \ge 0$ and for each $x \in \mathbb{R}$, $X = (X_t)_t$ is an adapted process on the filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P^{s,x})$.

Let $P(s,x,t,B)$ be a Markov transition function on $\mathbb{R}$. Assume that the following is true.

(1) $P^{s,x}\{X_s = x\} \;=\; 1$, for all $s \ge 0$ and $x \in \mathbb{R}$.

(2) $P^{0,x}\{\vec{X}_t \in B \mid \mathfrak{F}_s\} \;=\; P(s, X_s, t, B)\ \ P^{0,x}$ – a.s. , for $0 \le s < t$ and $x \in \mathbb{R}$.

Then we call X_t a **Markov process with transition function** $P(s,x,t,B)$. □

In the following, $E^{s,x}[\dots]$ denotes the expectation with respect to $P^{s,x}$. In other words,

$$E^{s,x}[Z] \;=\; \int Z dP^{s,x} \;=\; \int_\Omega Z(\omega)\, P^{s,x}(d\omega)\,,$$

for any $P^{s,x}$–integrable random variable Z.

Fact 15.3. *If X_t is a Markov process with transition function $P(s,x,t,B)$, then*

(1) $P^{0,x}\{\vec{X}_t \in B \mid \mathfrak{F}_s\} \;=\; P^{0,x}\{\vec{X}_t \in B \mid X_s\} \;=\; P(s,X_s,t,B)\;\; P^{0,x} - a.s.\,,$
for $0 \le s < t$ and $x \in \mathbb{R}$. That is the Markov property

(2) $E^{0,x}\{f\big(\vec{X}_t\big) \mid \mathfrak{F}_s\} \;=\; E^{0,x}\{f\big(\vec{X}_t\big) \mid X_s\} \;=\; \int_{\mathbb{R}} f(y)P(s,X_s,t,dy)\;\; P^{0,x} a.s.\,,$
for $0 \le s < t$, $x \in \mathbb{R}$, and nonnegative or bounded, Borel measurable f. *See (6.5) on p.124.*

(3) *If $x \in \mathbb{R}, s < t_1 < t_2 < \cdots < t_n$ and $B_1, \dots, B_n \in \mathfrak{B}$, then*

$$P^{s,x}\{X_{t_1} \in B_1, \dots, X_{t_n} \in B_n\} \;=\; \int_{B_1} P(s,x,t_1,dx_1) \cdots \int_{B_n} P(t_{n-1},x_{n-1},t_n,dx_n)\,.$$

□

Remark 15.7. Note the following significant structural differences between the solutions of an SDE as Markov processes and Markov processes with transition function.

In Note 15.1 (Technical note on the Markov property of SDE solutions) on p.271 we have:

(1) a fixed probability P on $(\Omega, \mathfrak{F}, \mathfrak{F}_t)$
(2) a separate stochastic process $X_t^{s,x}$ for each initial condition $X_s = x$
(3) a resulting Markov transition function $P(s,x,t,B) \;=\; P\{X_t^{s,x} \in B\}$.

When defining a Markov process with transition function, we have

(1) a family of probabilities $P^{s,x}$ on $(\Omega, \mathfrak{F}, \mathfrak{F}_t)$
(2) one and the same stochastic process X_t for each $(\Omega, \mathfrak{F}, \mathfrak{F}_t), P^{s,x})$
(3) a Markov transition function $P(s,x,t,B) \;=\; P^{s,x}\{X_t \in B\}$.

It feels much more natural to work with the second scenario, since dealing with one and the same process $X_t(\omega)$ makes it seem natural to think of $P^{s,x}\{\dots\}$ as a conditional probability $\widetilde{P}\{\cdots \mid X_s = x\}$, i.e.,

$$P^{s,x}\{X_{t_1} \in B_1, \dots, X_{t_n} \in B_n\} \;=\; \widetilde{P}\{X_{t_1} \in B_1, \dots, X_{t_n} \in B_n \mid X_s = x\}\,.$$

(Careful here! No claim is made that such a probability $\widetilde{P}$ actually exists as a mathematical object!)
□

Wouldn't it be nice if we could have the SDE solutions $X_t^{s,x}$ given by a single Markov process with transition function? This can in fact be done, but it comes at a significant cost. We must abandon the original filtered measurable space $(\Omega, \mathfrak{F}, \mathfrak{F}_t)$ (and also, of course the probability P and Brownian motion W_t) and create that single process which incorporates all solutions $X_t^{s,x}$ on a new filtered measurable space $(\widetilde{\Omega}, \widetilde{\mathfrak{F}}, \widetilde{\mathfrak{F}}_t)$.

An important reason why that is possible is the following. A Markov transition function $P(s,x,t,B)$ is defined without reference to Ω. Rather, the probabilities $P(s,x,t,\cdot)$ are defined on the Borel sets of $\mathbb{R}$.

The following can be shown.

Theorem 15.5.

Let $(s,x,t,B) \mapsto P(s,x,t,B)$ be a Markov transition function for $(\mathbb{R},\mathfrak{B}^1)$. Then there exist a measurable space $(\widetilde{\Omega},\widetilde{\mathfrak{F}})$, a filtration $\big(\widetilde{\mathfrak{F}}_t\big)_{t\geq 0}$, a real–valued function

$$\widetilde{X} : [0,\infty[\,\times\widetilde{\Omega}\,; \qquad (t,\tilde{\omega}) \mapsto \widetilde{X}_t(\tilde{\omega})\,,$$

and a family $\big(\widetilde{P}^{s,x}\big)_{s\geq 0, x\in\mathbb{R}}$ of probability measures on $\widetilde{\mathfrak{F}}$ as follows.
$\widetilde{X}$ is a Markov process with transition function $P(\cdot,\cdot,\cdot,\cdot)$. In other words,

(1) *$\widetilde{X}$ is an adapted process on the filtered probability space $\big(\widetilde{\Omega},\widetilde{\mathfrak{F}},\widetilde{\mathfrak{F}}_t,\widetilde{P}^{s,x}\big)$, for each $s\geq 0$ and $x\in\mathbb{R}$.*
(2) *$\widetilde{P}^{s,x}\{\widetilde{X}_s = x\} \,=\, 1$, for all $s\geq 0$ and $x\in\mathbb{R}$.*
(3) *$\widetilde{P}^{0,x}\{\widetilde{X}_t\in B \mid \widetilde{\mathfrak{F}}_s\} \,=\, P(s,\widetilde{X}_s,t,B)\ \ \widetilde{P}^{0,x}$ – a.s., for $0\leq s<t$ and $x\in\mathbb{R}$.*

PROOF: See the proof of Theorem 2.1.1 of [11] Friedman, Avner: Stochastic Differential Equations and Applications. ■

Remark 15.8.

(1) There is a multidimensional version of Theorem 15.5.
(2) one can choose for $\widetilde{\Omega}$ the set $C([0,\infty[,\mathbb{R})$ of all real–valued, continuous functions
$\tilde{\omega} : [0,\infty[,\mathbb{R}\,; \quad t\mapsto\tilde{\omega}(t)\,.$
(3) If the Markov transition function is associated with an SDE

$$dX_t \;=\; \beta\big(t,X_t\big)\,dt \,+\, \gamma\big(t,X_t\big)\,dW_t,$$

then we not only have to consider the measurable space $(\Omega,\mathfrak{F})$ and the filtration $\big(\mathfrak{F}_t\big)_t$, but also the Brownian motion W_t and the specific probability P that makes W_t a Brownian motion, i.e., $W_{t+tau} - W_t$ has normal distribution with mean zero and variance τ under P, and the trajectories of W are continuous P–a.s. This can be dealt with:
(4) One can construct a generic filtered probability space $\big(\widehat{\Omega},\widehat{\mathfrak{F}},\widehat{\mathfrak{F}}_t,\widehat{P}\big)$ with a Brownian motion $\widehat{W}_t$, a real–valued function $(t,\hat{\omega})\mapsto\widehat{X}_t(\hat{\omega})$, and a family $\big(\widehat{P}^{s,x}\big)_{s\geq 0,x\in\mathbb{R}}$ of (additional) probability measures on $\widehat{\mathfrak{F}}$ as follows. $\widehat{X}$ is a Markov process with transition function $P(\cdot,\cdot,\cdot,\cdot)$, and $\widehat{X}_t$ is a solution of the SDE with initial condition = with respect to the specific probability $\widehat{P}^{s,x}$. For a proof, see Theorem IV.1.1 of [12] Ikeda & Watanabe: Stochastic Differential Equations and Diffusion Processes.
(5) The construction done in **(4)** lets us keep the essence of what it means that a stochastic process $\widehat{X}$ is a solution of the SDE given in **(4)** with initial condition $\widehat{X}_u = x$:

$$\widehat{X}_t \;=\; x \,+\, \int_u^t \beta\big(s,\widehat{X}_s\big)\,ds \,+\, \gamma\big(t,\widehat{X}_s\big)\,d\widehat{W}_s\,.$$

At the same time, we managed to gain the advantage we had hoped for before stating Theorem 15.5: There now is a single process $\widehat{X}_t$ with enough trajectories to represent the multitude of solutions $X_t^{u,x}$ for the various initial conditions $X_u^{u,x} = x$.
(6) There is no magic. Different probability measures give nonzero probability to very different parts of $\widehat{\Omega}$, and thus to very different trajectories of $\widehat{X}$. Consider the sets

$$A(u,x_j) \;:=\; \{\,\hat{\omega} : \widehat{X}_u = x_j\,\}\,, \quad \text{for } j=1,2,\ u\geq 0, \ \text{ and different } x_1,x_2\in\mathbb{R}\,.$$

Then $\widehat{P}^{u,x_1}\big(A(u,x_1)\big) \;=\; \widehat{P}^{u,x_2}\big(A(u,x_2)\big) \;=\; 1,$
but $\widehat{P}^{u,x_1}\big(A(u,x_2)\big) \;=\; \widehat{P}^{u,x_2}\big(A(u,x_1)\big) \;=\; 0.$

(7) There is special terminology for specifying solutions of an SDE without referring to a specific carrier space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ and Brownian motion W_t. They are referred to as **weak solutions**. [57] □

15.5 Exercises for Ch.15

Exercise 15.1. Let T, X_t, $\Phi(x)$, $f(t,x)$ be as defined in Theorem 15.2 (Feynman–Kac Theorem) on p.273. Prove that the process

$$M_t \;:=\; f(t, X_t) \;=\; E^{t,x}[\Phi(X_T)]$$

is a martingale. **Hint**: Use formula (15.11) on p.273. □

[57] There is an entire litany of classifications of the solutions of an SDE Even worse, different authors sometimes choose the same definition to describe solutions with different properties.

16 Other Appendices

16.1 Greek Letters

The following section lists all greek letters that are commonly used in mathematical texts. You do not see the entire alphabet here because there are some letters (especially upper case) which look just like our latin alphabet letters. For example: $A =$ Alpha $B =$ Beta. On the other hand there are some lower case letters, namely epsilon, theta, sigma and phi which come in two separate forms. This is not a mistake in the following tables!

α	alpha	θ	theta	ξ	xi	ϕ	phi
β	beta	ϑ	theta	π	pi	φ	phi
γ	gamma	ι	iota	ρ	rho	χ	chi
δ	delta	κ	kappa	ϱ	rho	ψ	psi
ϵ	epsilon	$\varkappa$	kappa	σ	sigma	ω	omega
ε	epsilon	λ	lambda	ς	sigma		
ζ	zeta	μ	mu	τ	tau		
η	eta	ν	nu	υ	upsilon		

Γ	Gamma	Λ	Lambda	Σ	Sigma	Ψ	Psi
Δ	Delta	Ξ	Xi	Υ	Upsilon	Ω	Omega
Θ	Theta	Π	Pi	Φ	Phi		

16.2 Notation

This appendix on notation has been provided because future additions to this document may use notation which has not been covered in class. It only covers a small portion but provides brief explanations for what is covered.

For a complete list check the list of symbols and the index at the end of this document.

Notation 16.1. **a)** If two subsets A and B of a space Ω are disjoint, i.e., $A \cap B = \emptyset$, then we often write $A \biguplus B$ rather than $A \cup B$ or $A + B$. Both $A^{\complement}$ and, occasionally, $\complement A$ denote the complement $\Omega \setminus A$ of A.

b) $\mathbb{R}_{>0}$ or $\mathbb{R}^+$ denotes the interval $]0, +\infty[$, $\mathbb{R}_{\geq 0}$ or $\mathbb{R}_+$ denotes the interval $[0, +\infty[$,

c) The set $\mathbb{N} = \{1, 2, 3, \cdots\}$ of all natural numbers excludes the number zero. We write $\mathbb{N}_0$ or $\mathbb{Z}_+$ or $\mathbb{Z}_{\geq 0}$ for $\mathbb{N} \biguplus \{0\}$. $\mathbb{Z}_{\geq 0}$ is the B/G notation. It is very unusual but also very intuitive. $\square$

References

[1] Martin Anthony and Norman Biggs. Mathematics for Economics and Finance - Methods and Modelling. Cambridge University Press, 2nd edition, 2024.

[2] Louis Bachelier. Théorie de la spéculation. Annales Scientifiques de l'École Normale Supérieure, vol. 3, no. 17, pp. 21-86, 1900.

[3] Heinz Bauer. Probability Theory (Translated by R.Burckel). De Gruyter, Berlin, 1st edition, 1996.

[4] Heinz Bauer. Measure and Integration Theory. De Gruyter, Berlin, 1st edition, 2001.

[5] Matthias Beck and Ross Geoghegan. The Art of Proof. Springer, 1st edition, 2010.

[6] Peter J. Bickel and Kjell A. Doksum. Mathematical Statistics. Holden-Day, San Francisco, 1st edition, 1977.

[7] Thomas Björk. Arbitrage Theory in Continuous Time. Oxford University Press, 2nd edition, 2004.

[8] Ovidiu Calin. An Introduction to Stochastic Calculus with Applications to Finance. Lecture Notes, Eastern Michigan University, 2012.

[9] Richard M. Dudley. Real Analysis and Probability. Cambridge University Press, Cambridge, New York, 2nd edition, 2002.

[10] Michael Fochler. Lecture Notes for Math 447 - Probability. Published on the author's Binghamton University Math home page, 2024.

[11] Avner Friedman. Stochastic Differential Equations and Applications. Academic Press, New York, 1st edition, 1975.

[12] Nobuyuki Ikeda and Shinzo Watanabe. Stochastic Differential Equations and Diffusion Processes. North-Holland, 2nd edition, 1989.

[13] Bernt Øksendal. Stochastic Differential Equations: An Introduction With Applications. Springer, 6th edition, 2003.

[14] Steve E. Shreve. Stochastic Calculus for Finance II: Continuous-Time Models. Springer, 1st edition, 2004.

[15] James Stewart. Single Variable Calculus. Thomson Brooks Cole, 7th edition, 2012.

[16] D. Wackerly, W. Mendenhall, and R.L. Scheaffer. Mathematical Statistics with Applications. Thomson Brooks/Cole, 7th edition, 2008.

List of Symbols

Index