Math 454 - Additional Material

Student edition with proofs

Michael Fochler
Department of Mathematics
Binghamton University

Last update: May 17, 2021

tex-pp paramenters:

prmA	bu-ln-454
prmB	m330
prmC	teso
prmD	med
prmE	

Contents

1 Before You Start

History of Updates:

Date	Topic
2020-12-23	Created.

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, [4] 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. □

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.13) 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)**.** 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:

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

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

For the above example we have $20 \in X$, $27 - 6 \in X$, $38 \notin X$, 'Jimmy' $\notin X$. □

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

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

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

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. $\square$

Definition 2.2 (empty set)**.** $\emptyset$ or {} denotes the **empty set**. It is the one set that does not contain any elements. $\square$

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 "∼" 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.

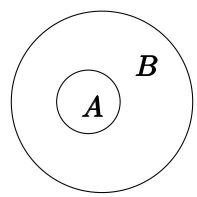


Figure 2.1: Set inclusion: $A \subseteq B,\ B \supseteq A$

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 "$B \supset A$". □

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. [2]

> 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!

[2] 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.

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. [3]

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 $A \cup B$. □

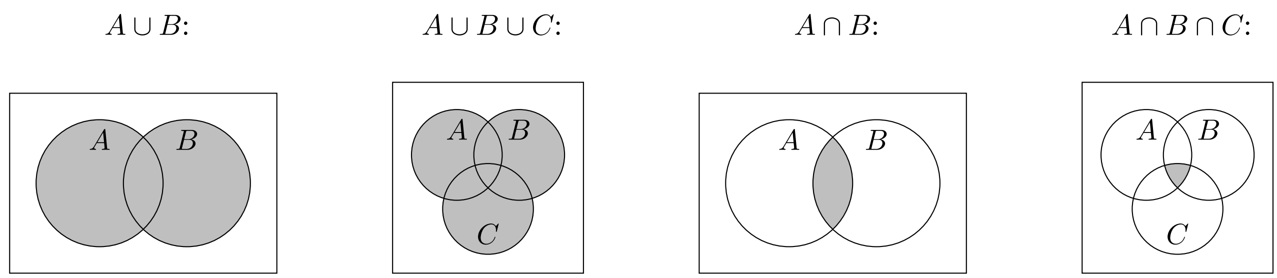


Figure 2.2: Union and intersection of sets

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 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 [2] 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!

[3] We could have shortened the phrase "all elements which belong to A or B or both" to "all elements which belong to A or B", and 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 ... 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).

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:

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

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:

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

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**. $\square$

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

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

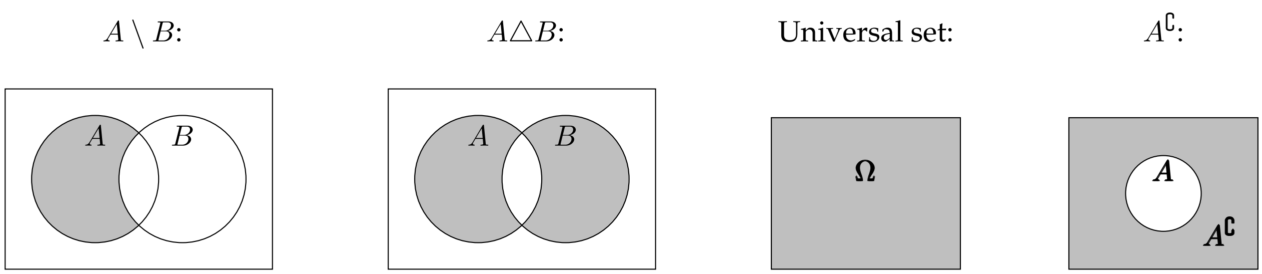


Figure 2.3: Difference, symmetric difference, universal set, complement

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., $\Omega = [0,1] = \{x \in \mathbb{R} : 0 \leqq x \leqq 1\}$. Let $a \in [0,1]$ and $\delta > 0$ and

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

[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.2.2 ("Classification of numbers") on p.14).

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

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

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*

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

PROOF: The proof is left as exercise 2.2. See p.29. ■

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*

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \tag{2.11}$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \tag{2.12}$$

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

PROOF of "$\subseteq$": Let $x \in (A \cup B) \cap C$. It follows from (2.2) on p.9 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 "$\supseteq$": 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.9 that $x \in (A \cup B) \cap C$, and we are done in this case.

[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.

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 "$\subseteq$" and of "$\supseteq$" were **proofs by cases**, i.e., we divided the proof into several cases (to be exact, two for each of "$\subseteq$" and "$\supseteq$"), 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., [2] B/G (Beck/Geoghegan) ch.6.2).

[6]Indicator functions will be discussed in ch.3.3 on p.37 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) $[\, x \in U$ and $x \notin V\,]$ or $[\, x \in V$ and $x \notin U\,]$
Hence $x \in (A \triangle B) \triangle C$ means either $x \in (A \triangle B)$ or $x \in C$, i.e.,
either $[\, x \in A, x \notin B$ or $x \in B, x \notin A\,]$ 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 $[\, x \in B, x \notin C$ or $x \in C, x \notin B\,]$, 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.** $\Omega = \biguplus \left[A : A \in \mathfrak{A} \right]$. □

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 $B_n := [\, n^2, (n+1)^2 [\, = \{x \in \mathbb{R} : n^2 \leqq x < (n+1)^2\}$. 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 notation $\mathfrak{P}(\Omega)$ 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 $|X|$ for the number of its elements, and we call $|X|$ the **size** of the set X.
- **b.** For infinite, i.e., not finite sets Y, we define $|Y| := \infty$. □

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

2.2 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. $\square$

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:

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

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. $\square$

We next define 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:

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

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 $n \geqq 2$) is given in prop.**??** on p.**??**. □

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

$\mathbb{N} := \{1, 2, 3, \ldots\}$ denotes the set of **natural numbers**.
$\mathbb{Z} := \{0, \pm1, \pm2, \pm3, \ldots\}$ 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. □

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,
$\mathbb{R}_+ := \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geqq 0\}$ 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,
$\mathbb{R}_{\neq 0} := \{x \in \mathbb{R} : x \neq 0\}$. ☐

Definition 2.16 (Intervals of Numbers [7]). We use the following notation for intervals of real numbers a and b:

$[a, b] := \{x \in \mathbb{R} : a \leqq x \leqq b\}$ is called the **closed interval** with endpoints a and b.

$]a, b[:= \{x \in \mathbb{R} : a < x < b\}$ is called the **open interval** with endpoints a and b.

$[a, b[:= \{x \in \mathbb{R} : a \leqq x < b\}$ and $]a, b] := \{x \in \mathbb{R} : a < x \leqq b\}$ are called **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":

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

Finally we define $[a, b[:=]a, b[:=]a, b] := \emptyset$ for $a \geqq b$ and $[a, b] := \emptyset$ for $a > b$. ☐

Notations 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,

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

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. ☐

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

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

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

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

[7] The following will be generalized in def.**??** on p.**??** to so called ordered integral domains.

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: $$x \vee y = \max(x,y) = \begin{cases} x & \text{if } x \geqq y, \\ y & \text{if } x \leqq y. \end{cases}$$

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

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 $f(x) = 2x$. $\square$

Remark 2.9. For any real number x we have

(2.20) $$\sqrt{x^2} = |x|. \ \square$$

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 $b \geqq 0$

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

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

Example: $\sqrt{9} = +3$, not -3. $\square$

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:*

(2.21) *Triangle Inequality* : $|\boldsymbol{a} + \boldsymbol{b}| \leqq |\boldsymbol{a}| + |\boldsymbol{b}|$

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. ■

2.3 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 $\big(x, \sqrt{x}\big)$ 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 $\big((x, y), \ln(x - y)\big)$ 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

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

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

$$f : C \longrightarrow D$$

as a short way of saying that the function $f(x)$ has domain C and codomain D.
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 $U := \{(x, y) : x, y \in \mathbb{R} \text{ and } 1 \leqq x \leqq 10 \text{ and } y < -2\}$ as the domain and $V := [0, \infty[$ as the codomain for $f_2(x, y) = \ln(x - y)$. These choices are OK because $x - y \geqq 1$ for any $(x, y) \in U$ and hence $ln(x - y) \geqq 0$, 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

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

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

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

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.2.3 (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. $\square$

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 $y \in Y$. $\square$

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. $f:\mathbb{R}\to\mathbb{R};\ x\to e^x$ <u>does not have an inverse</u> $f^{-1}(y)=\ln(y)$ since its domain would have to be the codomain $\mathbb{R}$ of f and $\ln(y)$ is not defined for $y\leqq 0$.

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

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

2.4 Cartesian Products

We next define cartesian products of sets. [8] Those mathematical objects generalize rectangles

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

and quads

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

Definition 2.21 (Cartesian Product)**.** Let X and Y be two sets The set

$$X\times Y\ :=\{(x,y):x\in X,y\in Y\} \tag{2.24}$$

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 as follows: Let $X_1,X_2,\dots,X_n$ be sets. The set

$$X_1\times X_2\cdots\times X_n\ :=\{(x_1,x_2,\dots,x_n):x_j\in X_j \text{ for each } j=1,2,\dots n\} \tag{2.25}$$

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

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 $X\times Y$. $\square$

Example 2.7. The domains given in **a** and **d** of the introduction to ch.2.3 (A First Look at Functions and Sequences) are subsets of the cartesian product

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

.

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

2.5 Sequences and Families

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

Definition 2.22. Let $n_\star$ be an integer and let let there be an item x_j for each integer $j \geqq n_\star$ 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 $(x_n)_{n \geqq n_\star}$ 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 $(u_k)_{k \geqq -2}$ is the sequence of integers $4, 1, 0, 1, 4, 9, 16, \dots$.
The second example is a sequence of sets. If $A_j = [-1 - \frac{1}{j}, 1 + \frac{1}{j}] = \{x \in \mathbb{R} : -1 - \frac{1}{j} \leqq x \leqq 1 + \frac{1}{j}\}$ then $(A_j)_{j \geqq 3}$ 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 $(x_i)_{i \geqq n_\star}$ 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 $(x_n)_{n_\star \leqq n \leqq n^\star}$ 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 $n_\star \leqq n_1 < n_2 < n_3 < \dots$ then we call the corresponding thinned out sequence $(x_{n_j})_{j \in \mathbb{N}}$ a **subsequence** of the sequence $(x_n)_{n \geqq m}$.
If this subset of indices is finite, i.e., we have $n_\star \leqq n_1 < n_2 < \dots < n_K$ for some suitable $K \in \mathbb{N}$ then we call $(x_{n_j})_{j=1}^{K}$ a **finite subsequence** of the original sequence. □

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

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

Here is the formal definition of a family.

Definition 2.23 (Indexed families)**.** Let J and X be nonempty set 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** or simply a **family** in X.
b. J is called the **index set** of the family.
c. For each $\jmath \in J$, $x_\jmath$ is called a **member of the family** $(x_j)_{j \in J}$. □

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 $\mathbb{N}$.

2.6 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 $13, 14, 15, 16$. [9]

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 $k \geqq k_0$ what to do for $k + 1$ if only we know what to do for k, then we know what to do for **any** $k \geqq k_0$! □

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. □

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. [10]

$$\bigcup_{j=1}^{1} A_j := A_1, \qquad \bigcup_{j=1}^{n+1} A_j := \Big(\bigcup_{j=1}^{n} A_j \Big) \cup A_{n+1}, \tag{2.26}$$

$$\bigcap_{j=1}^{1} A_j := A_1, \qquad \bigcap_{j=1}^{n+1} A_j := \Big(\bigcap_{j=1}^{n} A_j \Big) \cap A_{n+1}. \tag{2.27}$$

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

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

[10] 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.32.

computed as follows.

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

Remark 2.11. 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 $\Big(\bigcup_{j=1}^{k} A_j\Big) \cap B = \bigcup_{j=1}^{k} (A_j \cap B)$. 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 $k \geqq k_0$ 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 $k \geqq k_0$ 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** $k \geqq k_0$.

Here is an example which generalizes prop.2.2 on p.11.

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*

$$\Big(\bigcup_{j=1}^{n} A_j\Big) \cap B = \bigcup_{j=1}^{n} (A_j \cap B), \tag{2.28}$$

$$\Big(\bigcap_{j=1}^{n} A_j\Big) \cup B = \bigcap_{j=1}^{n} (A_j \cup B). \tag{2.29}$$

PROOF: We only prove (2.28), and this will be done by induction on n, i.e., the number of sets A_j. The proof of (2.29) is left as exercise 2.11

A. Base case: $k_0 = 1$. The statement $P(1)$ is (2.28) 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:

$$\text{(2.30)} \qquad \textbf{Induction assumption: } P(k): \left(\bigcup_{j=1}^{k} A_j\right) \cap B = \bigcup_{j=1}^{k} (A_j \cap B) \text{ is true for some } k \geqq 1.$$

Under this assumption

$$\text{(2.31)} \qquad \text{we must prove the truth of } P(k+1): \left(\bigcup_{j=1}^{k+1} A_j\right) \cap B = \bigcup_{j=1}^{k+1} (A_j \cap B).$$

The trick is to manipulate $P(k+1)$ in a way that allows us to "plug in" the induction assumption. For (2.31) 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 (2.30) will be used at some point.

$$\begin{aligned}
\left(\bigcup_{j=1}^{k+1} A_j\right) \cap B &= \left(\left(\bigcup_{j=1}^{k} A_j\right) \cup A_{n+1}\right) \cap B && \text{we used (2.26)}\\
&= \left(\left(\bigcup_{j=1}^{k} A_j\right) \cap B\right) \cup (A_{n+1} \cap B) && \text{we used (2.11) on p. 11}\\
&= \bigcup_{j=1}^{k} (A_j \cap B) \cup (A_{n+1} \cap B) && \textbf{we used the induction assumption!}\\
&= \bigcup_{j=1}^{k+1} (A_j \cap B) && \text{we used (2.26)}
\end{aligned}$$

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 (2.31) 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 $n \geqq 2$. Let $a_1, a_2, \dots, a_n \in \mathbb{N}$. Then*

$$\text{(2.32)} \qquad |a_1 + a_2 + \ldots + a_n| \leqq |a_1| + |a_2| + \ldots + |a_n|$$

PROOF: Note that this proposition generalizes prop.2.5 on p.18 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 2.32 was already shown (see (2.21) on p.18).

B. Induction step: Let us assume that 2.32 is true for some $k \geqq 2$. 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

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

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

(2.33) $$|A + a_{k+1}| \leqq |A| + |a_{k+1}| \leqq B + |a_{k+1}|$$

In other words,

(2.34) $$\big|(a_1 + a_2 + \ldots + a_k) + a_{k+1}\big| \leqq B + |a_{k+1}| = \big(|a_1| + |a_2| + \ldots + |a_k|\big) + |a_{k+1}|,$$

and this is (2.32) 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 $k \geqq k_0 = 2$. ■

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 2.12 (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 $n \geqq k_0$? 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 $n \geqq k_0$. □

2.7 Some Preliminaries From Calculus

Remark 2.13. 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 [5] 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., [5] Stewart, J: Single Variable Calculus.

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

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

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., $G = F$. □

2.8 Convexity ⋆

Note that this chapter is starred, hence optional.

Definition 2.24 (Concave-up and convex functions)**.** Let $-\infty \leqq \alpha < \beta \leqq \infty$ 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 $epi(f) := \{(x_1, x_2) \in I \times \mathbb{R} : x_2 \geqq f(x_1)\}$ 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 $S := \{\lambda\vec{a} + (1-\lambda)\vec{b}\} : 0 \leqq \lambda \leqq 1$ is contained in $epi(f)$.

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 nondecreasing. □

Proposition 2.8 (Convexity criterion)**.** *f is convex if and only if the following is true: For any*

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

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

$$f(x_0) \ \leqq\ S(x_0). \tag{2.35}$$

Note that any x_0 between a and b can be written as $x_0 = \lambda a + (1-\lambda)b$ for some $0 \leqq \lambda \leqq 1$ 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 $0 \leqq \lambda \leqq 1$ it is true that

$$f(\lambda a + (1-\lambda)b) \leqq \lambda f(a) + (1-\lambda) f(b). \tag{2.36}$$

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 $f(x_0) \leqq S(x_0)$ for all $a \leqq x_0 \leqq b$. It follows that convexity of f implies (2.35).

PROOF of "⇐": Let (2.35) 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

$$a_2 \geqq f(a_1) \ \text{ and } \ b_2 \geqq f(b_1). \tag{2.37}$$

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

Let $\vec{a}' := (a_1, f(a_1))$. Let $S' := \{\lambda \vec{a}' + (1-\lambda)\vec{b} : 0 \leqq \lambda \leqq 1\}$ 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 $S'' := \{\lambda \vec{a}' + (1-\lambda)\vec{b}' : 0 \leqq \lambda \leqq 1\}$ 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 $a \leqq x_0 \leqq b$ we have $f(x_0) \leqq S''(x_0)$. We conclude from $S(\cdot) \geqq S'(\cdot) \geqq S''(\cdot)$ that $S(x_0) \geqq f(x_0)$, i.e. $(x_0, S(x_0)) \in epi(f)$. As this is true for any $a \leqq x_0 \leqq b$ 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

$$m = \frac{S(b) - S(x_0)}{b - x_0} > \frac{S(b) - f(x_0)}{b - x_0} = \frac{f(b) - f(x_0)}{b - x_0} = f'(\xi) \tag{2.38}$$

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

$$m = \frac{S(x_0) - S(a)}{x_0 - a} < \frac{f(x_0) - S(a)}{x_0 - a} = \frac{f(x_0) - f(a)}{x_0 - a} = f'(\eta) \tag{2.39}$$

for some $a < \eta < x_0$ (according to the mean value theorem for derivatives).
Because f is concave up we have

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

From (2.38) and (2.39) we obtain

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

and we have reached a contradiction. ■

Proposition 2.10 (Sublinear functions are convex). *Let $f : \mathbb{R} \to \mathbb{R}$ be sublinear. Then f is convex.*

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

(2.40) $$p(\lambda x + (1-\lambda)y) \leqq p(\lambda x) + p((1-\lambda)y) = \lambda p(x) + (1-\lambda)p(y).$$

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

2.9 Exercises for Ch.2

2.9.1 Exercises for Sets

Exercise 2.1. Prove (2.12) of prop.2.2 on p.11.

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

2.9.2 Exercises for Proofs by Induction

Exercise 2.11. Use induction on n to prove (2.29) of prop.2.6 on p.24 of this document: Let $A_1, A_2, \dots$ and B be sets. If $n \in \mathbb{N}$ then $\left(\bigcap_{j=1}^{n} A_j\right) \cup B = \bigcap_{j=1}^{n} (A_j \cup B)$. ☐

Exercise 2.12. [11]
Let $K \in \mathbb{N}$ such that $K \geqq 2$ and $n \in \mathbb{Z}_{\geq 0}$. Prove that $K^n > n$. ☐

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

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 $k \in \mathbb{N}$.

(2.41) 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

$$(k+1)^2 + k + 1 = k^2 + 2k + 1 + k + 1 = (k^2 + k) + 2(k+1) \overset{(2.41)}{=} 2j + 2(k+1).$$

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

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 $n \in \mathbb{N}$. ☐

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 $k \in \mathbb{N}$.

(2.42)
Induction assumption: $k^3 + 5k$ is an integer multiple of 6, i.e., $k^3 + 5k = 6j$ for some $j \in \mathbb{Z}$.

[11] 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 $n = 2$).

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

$$\begin{aligned}(k+1)^3 + 5(k+1) &= k^3 + 3k^2 + 3k + 1 + 5k + 5 \ = \ (k^3+5k) + 3(k^2+k) + 6 \\ &= (k^3+5k) + 6i + 6 \ \overset{(2.42)}{=} \ 6(j+i+1).\end{aligned}$$

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

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 $n \in \mathbb{N}$. □

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 **nondecreasing 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 **nonincreasing 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 powerset 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 nondecreasing sequence."
- "Let $U_n \downarrow$ in $\mathfrak{E}$" is shorthand notation for
  "Let $U_n \subseteq \mathfrak{E} \, (n \in \mathbb{N})$" be a nonincreasing 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

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

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

We call $\bigcup\limits_{i \in I} A_i$ the **union** and $\bigcap\limits_{i \in I} A_i$ the **intersection** of the family $(A_i)_{i \in J}$
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 $A = \bigcup\limits_{a \in A} \{a\}$. According to (3.3) this also is true if $A = \emptyset$. $\square$

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

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

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

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 is important for stochastic processes. 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.

- $[t_0, t_1[,\ [t_1, t_2[,\ \dots,\ [t_{n-1}, t_n[$ 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.10). Then the following "duality principle" holds for any indexed family* $(A_\alpha)_{\alpha \in I}$ *of sets:*

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

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*

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

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

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. $\square$

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

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

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

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. $\square$

Note that the range $f(X)$ of f (see (**??**) on p.**??** 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 whatsover. 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 be used at the start of ch.**??** to define the inverse function $f^{-1} : Y \to X$ of f in case this can be done. [12] 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. $f([1,3]) = \{\, x^2 : x \in [1,3] \,\} = \{\, x^2 : 1 \leqq x \leqq 3 \,\} = [1,9]$.
c. $f(]-4,2[\cap [1,3]) = \{\, x^2 : x \in]-4,2[$ and $x \in [1,3] \,\} = \{\, x^2 : 1 \leqq x < 2 \,\} = [1,4[$. $\square$

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 $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$.

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

Example 3.6 (Preimages)**.** Let $f : \mathbb{R} \to \mathbb{R}; \quad f(x) = x^2$.

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

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

Proposition 3.3. *Some simple properties:*

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

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

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

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

(3.14) $\qquad f(X) \;=\; Y \quad \Leftrightarrow \quad$ *f is "surjective" (see df.**??** on p.**??**)*

(3.15) $\qquad f^{-1}(Y) \;=\; X$ *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 $\{a \leqq f \leqq b\} := f^{-1}([a,b]_R),\ \{a < f < b\} := f^{-1}(]a,b[_R)$,
$\{a \leqq f < b\} := f^{-1}([a,b[_R),\ \{a < f \leqq b\} := f^{-1}(]a,b]_R),\ \{f \leqq b\} := f^{-1}(]-\infty,b]_R)$, 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*

$$(3.16) \qquad f^{-1}(\bigcap_{j\in J} B_j) = \bigcap_{j\in J} f^{-1}(B_j)$$
$$(3.17) \qquad f^{-1}(\bigcup_{j\in J} B_j) = \bigcup_{j\in J} f^{-1}(B_j)$$
$$(3.18) \qquad f^{-1}(B^{\complement}) = \left(f^{-1}(B)\right)^{\complement}$$
$$(3.19) \qquad f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2)$$
$$(3.20) \qquad f^{-1}(B_1 \Delta B_2) = f^{-1}(B_1) \Delta f^{-1}(B_2)$$

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*

$$(3.21) \qquad f(\bigcap_{j\in J} A_j) \subseteq \bigcap_{j\in J} f(A_j)$$
$$(3.22) \qquad f(\bigcup_{j\in J} A_j) = \bigcup_{j\in J} f(A_j)$$

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, $\{0\} \subsetneq [0, \infty]$. $\square$

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*

$$(3.23) \qquad (g \circ f)(U) = g\big(f(U)\big) \text{ for all } U \subseteq X.$$
$$(3.24) \qquad (g \circ f)^{-1} = f^{-1} \circ g^{-1}, \text{ i.e., } (g \circ f)^{-1}(W) = f^{-1}\big(g^{-1}(W)\big) \text{ for all } W \subseteq Z.$$

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 "$E[1_A X] = \dots$" 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)**.** Ω be a nonempty set and $A \subseteq \Omega$. Let $1_A : \Omega \to \{0,1\}$ be the function defined as

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

1_A is called the **indicator function** of the set A. [13] □

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

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

Proposition 3.7. *Let A, B, C be subsets of Ω. Then*

$$\mathbb{1}_{A \cup B} \ = \max(\mathbb{1}_A, \mathbb{1}_B), \tag{3.27}$$
$$\mathbb{1}_{A \cap B} \ = \min(\mathbb{1}_A, \mathbb{1}_B), \tag{3.28}$$
$$\mathbb{1}_{A^{\complement}} \ = 1 - \mathbb{1}_A, \tag{3.29}$$
$$\mathbb{1}_{A \triangle B} \ = \mathbb{1}_A + \mathbb{1}_B \mod 2. \tag{3.30}$$

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

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

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

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

We obtained the equivalence in the middle from the fact that modular arithmetic is associative. ■

[13] Some authors call this **characteristic function** of A and some choose to write χ_A or $\mathbb{1}_A$ instead of 1_A.

4 Basic Measure and Probability Theory

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

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

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

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

$$\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{+\infty\} = \{x \in \mathbb{R} : x \geqq 0\} \cup \{+\infty\}$$

is the set of all nonnegative real numbers augmented by the elements ∞ and $-\infty$.
A function $F : X \to Y$ whose codomain Y is a subset of

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$$

is called an **extended real–valued function**. □

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:

$$c \pm \infty = \infty \pm c = \infty, \tag{4.1}$$
$$c \pm (-\infty) = -\infty \pm c = -\infty, \tag{4.2}$$
$$\infty + \infty = \infty, \tag{4.3}$$
$$-\infty - \infty = -\infty, \tag{4.4}$$
$$(\pm\infty) \mp \infty = \textbf{UNDEFINED}. \tag{4.5}$$

Rules for Multiplication:

$$p \cdot (\pm\infty) = (\pm\infty) \cdot p = \pm\infty, \tag{4.6}$$
$$(-p) \cdot (\pm\infty) = (\pm\infty) \cdot (-p) = \mp\infty, \tag{4.7}$$
$$0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0 \quad \text{and} \quad \frac{1}{\infty} = 0, \tag{4.8}$$
$$(\pm\infty) \cdot (\pm\infty) = \infty, \tag{4.9}$$
$$(\pm\infty) \cdot (\mp\infty) = -\infty, \tag{4.10}$$

Be clear about the ramifications of those rules. Rule (4.5) 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 (4.8) 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 $\lim\limits_{n\to\infty} a_n b_n = ab$.
Counterexample: $a_n = n, b_n = \frac{1}{n}$. □

For the following see SCF2 Definition 1.1.1.

Definition 4.2 (σ–algebras)**.** Let Ω be a nonempty set and 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 Ω if it satisfies the following:

(4.11a) $$\emptyset \in \mathfrak{F},$$

(4.11b) $$A \in \mathfrak{F} \quad \Rightarrow \quad A^{\complement} \in \mathfrak{F}$$

(4.11c) $$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \quad \Rightarrow \quad \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{F}$$

- 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. □

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 Variables).

Remark 4.2. If $\mathfrak{F}$ is a σ–algebra then

(4.12a) $$\Omega \in \mathfrak{F}$$

(4.12b) $$A \in \mathfrak{F} \quad \Rightarrow \quad A^{\complement} \in \mathfrak{F}$$

(4.12c) $$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \quad \Rightarrow \quad \text{and} \quad \bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{F}$$

The last assertion is a consequence of De Morgan's laws (Theorem 3.1 on p.34).

> 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. Two trivial σ–algebras:

(1) Given a nonempty set Ω, $\{\emptyset, \Omega\}$ is the smallest possible σ–algebra.
(2) Given a nonempty set Ω, $\{\emptyset, \Omega\}$ its power set 2^{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 4.3. Let $(\Omega, \mathfrak{F})$ be a measurable spaces and let $\mathfrak{E} \subseteq 2^{\Omega}$. We call the σ–algebra

$$\sigma(\mathfrak{E}) \;=\; \bigcap\{\mathfrak{G} : \mathfrak{G} \supseteq \mathfrak{E} \text{ and } \mathfrak{G} \text{ is a } \sigma\text{–algebra for } \Omega\}. \tag{4.13}$$

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

Remark 4.3.

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

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

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

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

In other words, $span(A)$ is the 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 (4.13). □

Proposition 4.2. *Let Ω be a nonempty set. Assume $\mathfrak{E}_1, \mathfrak{E}_2$ are subsets of 2^Ω such that*

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

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

PROOF: [$\star$] Left as an exercise. ■

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

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

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 4.2 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 4.4 (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**.
- We will not worry about what corresponds to the Borel sets when we deal with the extended real numbers $\bar{\mathbb{R}}$, i.e., we add $\pm\infty$. There is such a thing and those extended Borel sets are properly denoted $\mathfrak{B}(\bar{\mathbb{R}})$. Again, I will try not to even mention extended Borel sets.
- Abbreviations: We will 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 $\mathfrak{B}(\mathbb{R}^1)$. □

Remark 4.4. 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 4.5 (Abstract measures)**.** Let $(\Omega, \mathfrak{F})$ be a measurable space.

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

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

$$\mu(\emptyset) = 0 \qquad \text{(positivity)} \tag{4.14}$$

$$A, B \in \mathfrak{F} \text{ and } A \subseteq B \quad \Rightarrow \quad \mu(A) \leqq \mu(B) \qquad \text{(monotony)} \tag{4.15}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \text{ disjoint} \quad \Rightarrow \quad \mu\Big(\biguplus_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n) \qquad (\sigma\text{–additivity}) \tag{4.16}$$

- 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 (4.16) 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.8).

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

Remark 4.5 (σ–algebras are appropriate domains for measures)**.** The σ–additivity of measures is what makes working with them such a pleasure in many ways. You can now express it as follows: Given a disjoint sequence of measurable sets, the measure of the disjoint union is the sum of the measures. Property (4.11c) 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 of measures. □

Remark 4.6.

(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 expectations $E[X]$ of a random variable X that assignments of the form

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

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

(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 $\alpha \geqq 0$ then $\alpha\mu : A \mapsto \alpha\mu(A)$ also is a measure on $\mathfrak{F}$. □

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

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

of Example 4.2 on p.42 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 4.4 (Borel sets) on p.42) 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 $]a_1, b_1] \times]a_2, b_2] \times \cdots \times]a_n, b_n]$, $a_1 < b_1, a_2 < b_2, \dots, a_n < b_n$. □

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 4.6 (Lebesgue measure)**.** Given

- 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

$$\begin{aligned} &\lambda_0^1(]a, b]) := b - a, \\ &\lambda_0^2(]a_1, b_1] \times]a_2, b_2]) := (b_1 - a_1)(b_2 - a_2), \\ &\lambda_0^3(]a_1, b_1] \times]a_2, b_2] \times]a_3, b_3]) := (b_1 - a_1)(b_2 - a_2)(b_3 - a_3), \\ &\lambda_0^n(]a_1, b_1] \times \cdots \times]a_n, b_n]) := (b_1 - a_1)(b_2 - a_2) \dots (b_n - a_n). \end{aligned} \tag{4.17}$$

It can be shown that each of those real–valued functions satisfies the conditions stated in Fact 4.1. [14] Thus λ_0^n uniquely extends from the parallelepipeds to a measure λ^n on the Borel sets of $\mathbb{R}^n$. This measure is called (n-dimensional) **Lebesgue measure**.

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

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 powerset 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. □

Remark 4.7 (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 $A_{N+1} = A_{N+2} = \ldots = 0$:

(4.18)
$$\begin{aligned} &A_1, A_2, \ldots, A_N \in \mathfrak{F} \text{ mutually disjoint} \\ &\quad \Rightarrow \mu(A_1 \uplus A_2 \uplus \ldots \uplus A_N) = \mu(A_1) + \mu(A_2) + \ldots + \mu(A_N) \quad \text{(additivity).} \end{aligned}$$

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

(4.19a) $\mu(A) \geqq 0$ *for all* $A \in \mathfrak{F}$,

(4.19b) $A \subseteq B \Rightarrow \mu(B) = \mu(A) + \mu(B \setminus A)$,

(4.19c) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.

If μ is finite then also

(4.20a) $A \subseteq B \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$,

(4.20b) $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

PROOF: The first property follows from the fact that $\mu(\emptyset) = 0$, $\emptyset \subseteq A$ for all $A \in \mathfrak{F}$ and (4.15.

To prove the second property, observe that $B = A \uplus (B \setminus A)$.

Proving the third property is more complicated because neither A nor B may be a subset of the other. We first note that because $A \setminus B \subseteq A$, $B \setminus A \subseteq A$ and $A \cap B \subseteq A$, $\mu(A \cup B) = \infty$ can only be true if $\mu(A) = \infty$ or $\mu(B) = \infty$. In this case (4.19c) is obviously true. Hence we may assume that $\mu(A \cup B) < \infty$. We have

(4.21a) $A \cup B = (A \cap B) \uplus (B \setminus A) \uplus (A \setminus B)$

(4.21b) $A \cup B = A \uplus (B \setminus A) = B \uplus (A \setminus B)$

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

It follows from (4.21a) that

$$\mu(A \cup B) \;=\; \mu(A \cap B) + \mu(B \setminus A) + \mu(A \setminus B) \tag{4.22}$$

It follows from (4.21b) that

$$2 \cdot \mu(A \cup B) \;=\; \mu(A) + \mu(B \setminus A) + \mu(B) + \mu(A \setminus B) \tag{4.23}$$

We subtract the left and right sides of (4.22) from those of (4.23) 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 44), that one can extend any setfunction 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 4.4. *Let Ω be a countable, nonempty set, i.e., the elements of 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

$$\mu(A) \;=\; \sum_{\omega \in A} \mu_0\{\omega\}, \quad (A \subseteq \Omega). \tag{4.24}$$

PROOF: [⋆] This is a rather easy consequence of 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. μ_0 is then customarily written $p_n \;=\; P\{\omega_n\}$ and called a **probability mass function** (or just a **probability function** in [6] Wackerly, Mendenhall and Scheaffer: Mathematical Statistics with Applications).

For example, if we define $\Omega := \{1, 2, \dots, n\}$ and $\mathfrak{F} := 2^\Omega$ then the Bin(n,p) distribution is the (probability) measure P on the measurable space $(\Omega, \mathfrak{F})$ defined on the singleton events $\{1\}, \{2\}, \dots, \{n\}$ 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 4.6, p.44) on the space $\mathbb{Z}$ of the integers.

Definition 4.7. Let

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

Then the function

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

has according to Proposition 4.4 a unique extension

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

In other words, $\Sigma(A)$ is the size of A, i.e., the number of elements of$A$. We will call this measure the **summation measure**.

In this document a symbol with an arrow on top denotes a vector. So we write, e.g.,

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

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

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

We now can generalize the definition of summation measure to multiple dimensions. Let $n \in \mathbb{N}$ and

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

Then the function

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

has according to Proposition 4.4 a unique extension

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

As in the one–dimensional case, $\Sigma(A)$ is the size of A, i.e., the number of elements of$A$. We will call this measure the n**–dimensional summation measure**. □

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

We mentioned earlier that (see Definition 4.6, p.44) on the space $\mathbb{Z}$ of the integers.

Proposition 4.5 (Continuity properties of measures)**.** *Let* $(\Omega, \mathfrak{F}, \mu)$ *be a measure space.*

(4.27a) *If* $A_n \downarrow A$ *in* $\mathfrak{F}$ *and* $\mu(A_1) < \infty$ *then* $$\lim_{n\infty} \mu(A_n) \ = \ \mu(A) \ = \ \mu\left(\lim_{n\infty} A_n\right),$$

(4.27b) *If* $B_n \uparrow B$ *then* $$\lim_{n\infty} \mu(B_n) \ = \ \mu(B) \ = \ \mu\left(\lim_{n\infty} B_n\right).$$

PROOF: Prove formula (4.27b) first. To do so replace the sequence B_n with a disjoint sequence C_n such that $A = \biguplus_n C_n$. See Proposition 3.2 (Rewrite unions as disjoint unions) on p.34 and use the σ–additivity of μ.

To prove (4.27a), apply the already proven formula (4.27b) to

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

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

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

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

Remark 4.8. The finiteness condition of formula (4.27a) is never an issue with probability measures P since $P(A) \leqq 1$ for all $A \in \mathfrak{F}$. But the unexpected can happen for nonfinite measures such as the one–dimensional summation measure Σ' of Definition 4.7 which is characterized by

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

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 (4.27a) is not valid.

Let $A_k := \{j \in \mathbb{N} : j \geqq k\}$. 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

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

contrary to our assumption $n \in \mathbb{N}$

$$\text{So} \ \ A \ = \ \emptyset, \quad \text{thus} \ \ \Sigma(\cap_n)A_n \ = \ \Sigma(\emptyset) \ = \ 0.$$

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(4.27a) does not hold. □

Proposition 4.6. ⋆

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.34. ■

Proposition 4.7. ⋆

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 $\lim_n \mu_n$ is a measure.

PROOF: Not given here. We only mention that Proposition 4.5 (Continuity properties of measures) on p.47 is essential to show that μ is σ–additive once it has been shown to be (finitely) additive. ■

4.2 Measurable Functions and Random Variables

Introduction 4.1. We all know what a random variable X is: X has a real number as an outcome but 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 $a \leqq Z \leqq b$ can be computed as

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

  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

$$p_j = P\{X = x_1\}, \quad (j = 1, 2, \dots).$$

  An example would be a Bin(n,p)–distributed random variable (see Example 4.4 (Binomial distribution) on p.46) for which $p_j = \binom{n}{j} p^j (1-p)^{n-j}$.

That won't do anymore, and we will try to make some amendments to the above.

- "... and 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 function value (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 X rather than the real numbers (the codomain of X). We have seen in Fact 4.2 on p.45 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.,

$$\{-1 \leqq X \leqq 1\} = \{\omega \in \Omega : -1 \leqq X(\omega) \leqq 1\} = X^{-1}(\omega)$$

and $P\{-1 \leqq X \leqq 1\}$ is not available because $\{-1 \leqq X \leqq 1\} \notin \mathfrak{F}$?
- What events are important, i.e., what are the sets $B \in \mathbb{R}$ such that the preimage $X^{-1}(B)$ (also written $\{X \in B\}$) [15] 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:

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

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 $\mathfrak{F}$. □

Definition 4.8 (Measurable function)**.** Let

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

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** or, simpler, that f **is in** $m(\mathfrak{F}, \mathfrak{F}')$, if

$$f^{-1}(A') \in \mathfrak{F}, \quad \text{for all } A' \in \mathfrak{F}'. \tag{4.28}$$

In the special case that $\Omega' = \mathbb{R}^n$ or $\Omega' = \mathbb{R}$ and $\mathfrak{F}'$ is the Borel σ–algebra we also say that f is $\mathfrak{F}$–**measurable** or that f **is in** $m(\mathfrak{F})$.
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. □

See SCF2 Definition 1.2.1 for the next definition.

Definition 4.9 (Random Variable)**.**

[15] see the **Notational conveniences II** box that follows Proposition 3.3 on p.36)

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.,

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

then we call X a **random variable**. on $(\Omega, \mathfrak{F}, P)$.

Occasionally we allow X to assume the values ∞, and $-\infty$, i.e., X can be an extended real–valued, $\mathfrak{F}$–measurable, function.
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 can shrink the codomain to $(A, 2^A)$ and talk about the random variable
$$X : (\Omega, \mathfrak{F}, P) \longrightarrow (A, 2^A).$$
Here is the reason that we can take the entire powerset 2^A as the σ–algebra of the codomain:

- 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. □

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

Definition 4.10 (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,

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

(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 abbreviate 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 abbreviate 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 4.9. ⋆

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

Remark 4.11.

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

$$f^{-1}(A') \subseteq \mathfrak{F} \text{ for all} A' \in \mathfrak{F}'. \tag{4.30}$$

PROOF: ⋆ Omitted, but not hard when you use Theorem 3.4 (f^{-1} is compatible with all basic set ops) on p.36 ■

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) $\{f \leqq c\} \in \mathfrak{F}$ *for all* $c \in \mathbb{R}$,
(3) $\{f > c\} \in \mathfrak{F}$ *for all* $c \in \mathbb{R}$,
(4) $\{f \geqq c\} \in \mathfrak{F}$ *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 4.4 on p.42. ■

@@MFx-Auth

A

This is the proper spot for Proposition 4.15 in the addenda to this chapter. See p.75.

For the following see Definitions 2.17 and 2.18 on p.16.

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 and g in $m(\mathfrak{F})$ then each of the following also is $(\mathfrak{F}, \mathfrak{B}^1)$–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 $f_1, f_2, \ldots, f_n$.*
- *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

$$f(\omega) \leqq c \quad \Leftrightarrow \quad f_n(\omega) \leqq c \text{ for all } n,$$

thus

$$\{ f \leqq c \} \; = \; \bigcap_{n \in \mathbb{N}} \{ f_n \leqq c \},$$

and this set is $\mathfrak{F}$–measurable as the intersection of the $\mathfrak{F}$–measurable sets $\{f_n \leqq c\}$. The assertion now follows from Corollary 4.1. ■

Example 4.5 (Binomial random variable v.s. binomial distribution)**.** This example continues Example 4.4 (Binomial distribution) on p.46 which was about the binomial distribution Bin(n,p) defined by its probability mass function

$$p_j \; = \; P\{j\} \; = \; \binom{n}{j} p^j \, (1-p)^{n-j}. \tag{4.31}$$

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and let X be 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 Bin(n,p)–distributed random variable. It satisfies formula (4.31), right?

Not exactly! There is a problem with the probability P. In formula (4.31) it occurs as a measure on the measurable space

$$\left(\{0, 1, \dots, n\}, 2^{\{0,1,\dots,n\}}\right)$$

and <u>NOT</u> 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 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 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 Variables). See p.49. 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 it will work for any measurable function $f : (\Omega, \mathfrak{F}, \mu) \to (\Omega', \mathfrak{F}')$: Define a measure μ_f 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 4.8. *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

$$\mu_f : \mathfrak{F}' \to [0, \infty]; A' \mapsto \mu\{f \in A'\} = \mu\{\omega \in \Omega : f(\omega) \in A'\} \tag{4.32}$$

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: ⋆ A triviality if you make use of Proposition3.4 (f^{-1} is compatible with all basic set ops) on p.36. ■

For the following see SCF2 Definition 1.2.3.

Definition 4.11 (Image measure)**.**

(1) We call the measure μ_f of Proposition 4.8 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 (4.32), given by

$$P_X(B) := P\{X \in B\} = P\{\omega \in \Omega : X(\omega) \in B\}, \quad (B \in \mathfrak{B}^1) \tag{4.33}$$

the **probability distribution** or simply the **distribution** of X. SCF2 also calls P_X the **distribution measure** of X. □

Proposition 4.9. *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 $A \in \mathfrak{F}$ which do not belong to $\sigma(f)$ 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.36.

(2) is easy to see from the definition of masurability of a function. ■

Definition 4.12. Let Ω, Ω' be nonempty, $\mathfrak{F}'$ a σ–algebra on Ω', and $f : \Omega \to \Omega'$.

We call the σ–algebra from Proposition 4.9

$$\sigma(f) := \{f^{-1}(A') : A' \in \mathfrak{F}'\} \tag{4.34}$$

the **σ–algebra generated by** f. □

Remark 4.12. 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 4.9.**(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.

In a sense we can think of $\sigma(X)$ as the information one can associate with the random phenomenon X. This is discussed at length in SCF2, ch.2. □

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 $\mathbb{R}^m$. ■

4.3 Integration and Expectations

The following should be read in conjunction with SCF2 ch.1.3: Expectations.

Remark 4.13. We recall that if $f : \mathbb{R} \to \{0, 1\}$ and $g : \mathbb{R}^n \to \{0, 1\}$ are Riemann–integrable and if also the sets $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}^n$ are Riemann–integrable, i.e., the Riemann integrals

$$\int_{-\infty}^{\infty} 1_A(x)\,dx \qquad \text{and} \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 1_B(x_1, x_2, \dots, x_n)\,dx_1 dx_2 \cdots dx_n$$

of the indicator functions $1_A : \mathbb{R} \to \{0, 1\}$ and $1_B : \mathbb{R}^n \to \{0, 1\}$ exist, then we write

$$\int_A f(x)\,dx = \int_{-\infty}^{\infty} f(x) 1_A(x)\,dx, \tag{4.35}$$

$$\int_B g(x_1, \dots, x_n)\,dx_1 \cdots dx_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) 1_B(x_1, \dots, x_n)\,dx_1 \cdots dx_n. \tag{4.36}$$ □

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

$$f(x) = \sum_{j=1}^{k} c_j\, 1_{]a_j, b_j]}(x),$$

then

$$\int_{-\infty}^{\infty} f(x)\,dx = \sum_{j=1}^{k} c_j \int_{-\infty}^{\infty} 1_{]a_j,b_j]}(x) = \sum_{j=1}^{k} c_j \int_{a_j}^{b_j} dx = \sum_{j=1}^{k} c_j(b_j - a_j) = \sum_{j=1}^{k} c_j\,\lambda^1(]a_j,b_j]) \tag{4.37}$$

where λ_1 denotes Lebesgue measure which was introduced in Definition 4.6 on p.44.

B. Things are similar in the multidimensional case. If $g:\mathbb{R}^n \to \mathbb{R}$ has the form

$$g(x) = \sum_{j=1}^{k} c_j\, 1_{]u_{1j},v_{1j}]\times\cdots\times]u_{nj},v_{nj}]}(x), \quad (u_{ij} < v_{ij} \text{ for } i = 1,\dots,n),$$

then

$$\begin{aligned}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\dots,x_n)\,dx_1\cdots dx_n &= \sum_{j=1}^{k} c_j \int_{u_{1j}}^{v_{1j}}\cdots\int_{u_{nj}}^{v_{nj}} dx_1\cdots dx_n\\ &= \sum_{j=1}^{k} c_j(v_{1j} - u_{1j})\cdots(v_{nj} - u_{nj})\\ &= \sum_{j=1}^{k} c_j\,\lambda^n(]u_{1j},v_{1j}]\times\cdots\times]u_{nj},v_{nj}])\,.\end{aligned} \tag{4.38}$$

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

$$f(x) = \sum_{j=1}^{k} c_j\, 1_{]a_j,b_j]}(x), \ (k\in\mathbb{N}),$$

then the expected value $E[f\circ X]$ of the composite function $f\circ X:\omega\mapsto f\big(X(\omega)\big)$ is

$$E[f\circ X] = \sum_{j=1}^{k} c_j \int E\big[1_{]a_j,b_j]}(X)\big] = \sum_{j=1}^{k} c_j P\{X\in]a_j,b_j]\} = \sum_{j=1}^{k} c_j P_X(]a_j,b_j]). \tag{4.39}$$

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 $f = \sum_{j=1}^{k} c_j 1_{A_j}$ which takes finitely many values c_j and we have computed in each case an integral or an expected value of the form $\sum_{j=1}^{k} c_j\mu(A_j)$ for a suitable measure μ. We will now establish a connection between those instances.
□

Definition 4.13 (Integral of a simple function)**.** Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space, $n \in \mathbb{N}$, and $A_1, A_2, A_n \in \mathfrak{F}$ a finite collection of measurable sets. Let $f : \Omega \to \mathbb{R}$ be defined as

$$f := \sum_{j=1}^{n} c_j 1_{A_j}, \quad 0 \leqq c_j < \infty \text{ for } j = 1, \dots, n. \tag{4.40}$$

We call such a function which only assumes finitely many function values a **simple function**. Note that $f \geqq 0$ and f is measurable as the sum of the measurable functions $\omega \mapsto c_j \cdot 1_{A_j}(\omega)$. We call

$$\int f\,d\mu := \int f(\omega)\,d\mu(\omega) := \int f(\omega)\,\mu(d\omega) := \sum_{j=1}^{n} c_j \mu(A_j). \quad \square \tag{4.41}$$

the **integral**, or also the **abstract integral**, of f with respect to μ.

Remark 4.14. ⋆

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

$$\begin{aligned} f &= 3 \cdot 1_{\mathbb{R}} = 3 \cdot 1_{]-\infty,0[} + 3 \cdot 1_{[0,\infty[} \\ &= 1 \cdot 1_{]-\infty,-1[} + 2 \cdot 1_{]-\infty,1[} + 1 \cdot 1_{]-1,\infty[} + 2 \cdot 1_{[1,\infty[}. \end{aligned}$$

Thus the following is important since it ensures that the definition of $\int f d\mu$ consistent:
C. Let the simple, nonnegative, function f have representations

$$f := \sum_{j=1}^{k} c_j 1_{A_j} = \sum_{j=1}^{k'} c'_j 1_{A'_j}.$$

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. $\square$

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.16.

Definition 4.14 (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 $f \geqq 0$, we define

$$\int f\,d\mu := \sup\left\{ \int h\,d\mu \,:\, h \text{ is simple and } 0 \leqq h \leqq f \right\}. \tag{4.42}$$

If not both $\int f^+ d\mu = \infty$ and $\int f^- d\mu = \infty$, we define

$$\int f\,d\mu := \int f^+\,d\mu - \int f^-\,d\mu. \tag{4.43}$$

Again we call $\int f d\mu$ the **integral** or **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**.

We allow the alternate notation

$$\int f\,d\mu = \int f(\omega)\,d\mu(\omega) = \int f(\omega)\,\mu(d\omega). \ \square$$

The definition of integrability has been changed in MF454 version 2021-02-21 as follows.

- **Old version 2021-02-19**: We called any function f μ–integrable as long as $\int f d\mu = \int f^+ d\mu - \int f^- d\mu$ exists, i.e., this expression is not of the form $\infty - \infty$.
- **From now on**: To call f μ–integrable it must satisfy the condition $\int |f| d\mu < \infty$. You will see in part **b** of Theorem 4.3 (Fundamental properties of the abstract integral) on p.59 that this condition is equibalent to both $\int f^+ d\mu < \infty$ and $\int f^- d\mu < \infty$

Remark 4.15. ⋆

Note that there are measurable functions f which are not μ–integrable even though $\int f d\mu$ exists. For example, let

$$f : (\mathbb{R}, \mathfrak{B}^1, \lambda^1) \longrightarrow (\mathbb{R}, \mathfrak{B}^1); \qquad f(x) := x^+ = x\,1_{[0,\infty[}.$$

Here is a formal proof that $\int x^+ d\lambda^1(x) = \infty$. For each $n \in \mathbb{N}$, let $h_n := n \cdot 1_{[n,2n]}$. Then $h_n \leqq f$ and this simple function has integral $\int h_n d\lambda = n \cdot \lambda^1([n, 2n]) = n^2$. Thus

$$\int x^+\,d\lambda^1 = \sup\left\{ \int h\,d\lambda^1 \,:\, h \text{ is simple and } 0 \leqq h \leqq x^+ \right\} \geqq \sup_{n\in\mathbb{N}}\left\{ \int h_n\,d\lambda^1 \right\} = \infty.$$

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 4.15 (Expected value of a 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

$$E[X] := \int X\,dP = \int X(\omega)\,dP(\omega) = \int X(\omega)\,P(d\omega). \qquad (4.44)$$ □

Proposition 4.11. ⋆

Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and $A \in \mathfrak{F}$. Let μ_A be the measure defined in Proposition 4.6 on p.48:

$$\mu_A(A') = \mu(A \cap A')$$

If $f \in m(\mathfrak{F})$ is μ–integrable then $f1_A$ is integrable with respect to both μ and μ_A, and then

$$\int f1_A\,d\mu = \int f1_A\,d\mu_A = \int f\,d\mu_A.$$

PROOF: Not entirely trivial. You first prove this for simple functions h and then use

$$0 \leqq h \leqq f \Leftrightarrow 0 \leqq h1_A \leqq f1_A$$

to prove the general case. ■

The last proposition shows that if f is μ–integrable and $A \in \mathfrak{F}$ then $\int f1_A d\mu$ exists. We are in a position to define the following.

Definition 4.16. Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space, $A \in \mathfrak{F}$.

If f is a measurable function and $\int f1_A d\mu$ exists (is not of the form $\infty - \infty$) then we call

$$\int_A f\,d\mu := \int f \cdot 1_A\,d\mu \qquad (4.45)$$

the **integral** or **abstract integral**, of f over A with respect to μ. We allow the alternate notation

$$\int_A f\,d\mu = \int_A f(\omega)\,d\mu(\omega) = \int_A f(\omega)\,\mu(d\omega).$$

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 4.3 (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 function values $x_0, x_1, \dots, x_n$, then*

$$\int f\,d\mu \;=\; \sum_{k=0}^{n} x_k\,\mu\left(f^{-1}\{x_k\}\right).$$

In particular, if Ω is finite, 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*

$$\int (\alpha X + \beta Y)\,d\mu \;=\; \alpha\int f\,d\mu \;+\; \beta\int g\,d\mu.$$

PROOF: See SCF2, proof of Theorem 1.3.4. ■

And this is the version for probability spaces which you will find as SCF2 Theorem 1.3.4.

Theorem 4.4. *Let X be a random variable on a probability space $(\Omega, \mathfrak{F}, P)$.*

a. *If X takes only finitely many values $x_0, x_1, \dots, x_n$, then*

$$E(X) \;=\; \sum_{k=0}^{n} x_k\,P\{X = x_k\}.$$

In particular, if Ω is finite, then

$$E(X) \;=\; \sum_{\omega\in\Omega} X(\omega)\,P\{\omega\}.$$

b. *(**Integrability**) The random variable X is integrable if and only if*

$$E[X^+] \; < \; \infty \quad \text{and} \quad E[X^-] \; < \; \infty$$

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 may 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*

$$\varphi\big(E(X)\big) \; \leqq \; E\big(\varphi(X)\big).$$

PROOF: See SCF2. ■

@@MFx-Auth

A

This the placeholder for Theorem 4.13 on p.76 in the addenda to this chapter.

The following theorem, [SCF2 Theorem 1.3.8, is specific to Lebesgue measure. It is true in multiple dimensions, but we only state it for the one–dimensional case.

Theorem 4.5. *Connection between Riemann and Lebesgue integrals] Let f be a bounded function, defined on $\mathbb{R}$, and let $a < b$.*

(1) *The Riemann integral $\int\limits_a^b f(x)\,dx$ is defined (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 $\int\limits_a^b f(x)\,dx$ is defined, then f is Borel–measurable (so the Lebesgue integral $\int\limits_{[a,b]} f(x)\,d\lambda^1(x)$ is also defined), and both integrals agree.*

PROOF: ⋆ Beyond the scope of this course. ■

Remark 4.16.

(1) Theorem 4.5**(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) A singleton sets $\{x\}$ in $\mathbb{R}$ have Lebesgue measure zero, thus 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 use the familiar notation $\int_a^b f(x)\,dx$ for Riemann integrals 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 write $\int_B f(x)\,dx$. $\square$

4.4 Convergence of Measurable Functions and Integrals

The following corresponds to SCF2 Chapter 1.4 but note that what is formulated here for arbitrary measure spaces $(\Omega, \mathfrak{F}, \mu)$ is done 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 4.10 on p.51), to the convergence of functions. In this case the property of interest for an $\omega \in \Omega$ 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 4.17 (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 $\mu(A^{\complement}) = 0$ then 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 4.18 (Convergence almost surely)**.**

Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X_n, X a sequence o random variables with domain Ω such that $\lim_{n\to\infty} f_n = f$ μ–a.e. as defined above. We prefer to say that **the sequence** X_n **has limit** X P**–almost surely** and we write

$$\lim_{n\to\infty} X_n = X \ P\text{–a.s.} \qquad \text{or} \qquad X_n \to X \ P\text{–a.s. as } n \to \infty \ \square$$

The following is SCF2 Example 1.4.4.

Example 4.6. 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

$$f_n(x) := \sqrt{\frac{n}{\sqrt{2\pi}}}\, e^{-\frac{nx^2}{2}} \qquad \text{density function of the } N(0,n)\text{–distribution}, \tag{4.46}$$

$$f(x) := \begin{cases} 0 & \text{if } x \neq 0, \\ \infty & \text{if } x = 0. \end{cases} \tag{4.47}$$

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 $\int_{\mathbb{R}} f(x) d\lambda^1(x) = 0$. So when can we switch $\int$ and $\lim_n$? $\square$

Here is another such example.

Example 4.7. 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

$$f_n := 1_{[n,\infty[}, \quad n = 1, 2, 3, \dots, \qquad \text{i.e., } fn(x) = 1 \text{ for } x \leqq n \text{ and zero else.} \tag{4.48}$$

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. $\square$

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 4.6 (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 $m(\mathfrak{F}, \mathfrak{B}^1)$.*
>
> $$\text{If } 0 \leqq f_1 \leqq f_2 \leqq \dots \text{ a.e. and } \lim_{n\to\infty} f_n = f \text{ a.e., then } \lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$
>
> ***(2).*** *Let X and $X_1, X_2, X_3, \dots$ be random variables on a probability space $(\Omega, \mathfrak{F}, P)$.*
>
> $$\text{If } 0 \leqq X_1 \leqq X_2 \leqq \dots \text{ a.s. and } \lim_{n\to\infty} X_n = X \text{ a.s., then } \lim_{n\to\infty} E[X_n] = E[X].$$

PROOF ⋆ : Will not be given. Observe though that **(2)** matches **(1)** in the special case that $\mu(\Omega) = 1$. ■

Remark 4.17. ⋆

Observe that neither Example 4.6 nor Example 4.7 satisfy the condition of the theorem. The functions in both are nonnegative and in example 4.7 they even are monotone but there they are non–increasing rather than non–decreasing. $\square$

Just as useful as the Monotone Convergence Theorem is the following one (SCF2 Theorem 1.4.9.)

Theorem 4.7 (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 $m(\mathfrak{F}, \mathfrak{B}^1)$. Further assume that $g \geqq 0$ and g is integrable, i.e., $\int g d\mu < \infty$.*

$$\text{If } |f_j| \leqq g \text{ a.s. for each } j \text{ and } \lim_{n\to\infty} f_n = g \text{ a.s.}, \quad \text{then } \lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

***A**. Let X, Y and $X_1, X_2, X_3, \dots$ be random variables.*

$$\text{If } |X_j| \leqq Y \text{ a.s. for each } j \text{ and } \lim_{n\to\infty} X_n = X \text{ a.s.}, \quad \text{then } \lim_{n\to\infty} E[X_n] = E[X].$$

PROOF ⋆ : Will not be given. Observe again that **(2)** matches **(1)** in the special case that $\mu(\Omega) = 1$. ■

Understand 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 [1] Bauer, Heinz: Measure and Integration Theory.

Proposition 4.12. ⋆ *Let $(\Omega, \mathfrak{F}, \mu)$ be a probability space and $a < b$ two real numbers. Assume the function $f :]a, b[\times \Omega \to \mathbb{R}$ satisfies the following.*

- **(1)** *For any fixed $a < t < b$ the function $\omega \mapsto f(t, \omega)$ is μ–integrable (and thus by necessity $\mathfrak{F}$–measurable).*
- **(2)** *For any fixed $\omega \in \Omega$ the function $t \mapsto f(t, \omega)$ has a partial derivative*

$$f_t : s \mapsto f_t(s, \omega) = \frac{\partial f}{\partial t}(s, \omega).$$

- **(3)** *There exists a non-negative and μ–integrable function $g : \Omega \to \mathbb{R}$ which dominates $|f_t|$:*

$$|f_t(s, \omega)| \leqq g(\omega) \text{ for all } a < s < b,\ \omega \in \Omega.$$

Then we can differentiate under the integral. More specifically,

$$s \mapsto \int_\Omega f(s, \omega) \, d\mu(\omega) \quad \text{is differentiable for each } \omega,$$

$$\omega \mapsto f_t(s, \omega) \quad \text{is } \mu\text{–integrable for each } a < s < b, \text{ and}$$

$$\int_\Omega f_t(s_0, \omega) \, d\mu(\omega) = \frac{d}{dt} \int_\Omega f(s_0, \omega) \, d\mu(\omega).$$

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)**,

$$\lim_{n\to\infty} h_n(\omega) \;=\; f_t(s_0,\omega) \;\text{ for all } \omega \in \Omega. \tag{4.49}$$

In particular, the function $\omega \mapsto f_t(s_0,\omega$ is measurable as limit of the measurable h_n.

We next show that $|h_n| \leqq g$ 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 **(2)** we thus obtain $|h_n(\omega)| \leqq g(\omega)$. $\omega \mapsto f_t(s_0,\omega)$ thus is μ–integrable. We apply dominated convergence to formula (4.49) and obtain

$$\lim_{n\to\infty} \int_\Omega h_n(\omega)\,d\mu(\omega) \;=\; \int_\Omega f_t(s_0,\omega)\,d\mu(\omega). \tag{4.50}$$

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 (4.50) 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 4.8.

(1). *Let $(\Omega,\mathfrak{F},\mu)$ be a measure space and let $f \geqq 0$ be an extended real–valued, Borel–measurable function on Ω. Then the set function*

$$\nu : \mathfrak{F} \longrightarrow [0,\infty], \qquad \nu(A) \;:=\; \int_A f\,d\mu \tag{4.51}$$

defines a measure on $\mathfrak{F}$.

PROOF:

A. To show that that $1_\emptyset = 0$, thus $f \cdot 1_\emptyset = 0$. Thus

$$\nu(\emptyset) = \int 0\,d\mu \;=\; \mu(A)\cdot 0 \;=\; 0.$$

(We might have had to use the rule $\infty \cdot 0 = 0$ once or even twice!)

B. ν is monotone since $A \subseteq A'$ for measurable A and A' implies $f \cdot 1_A \leqq f \cdot 1_{A'}$, thus

$$\nu(A) = \int f \cdot 1_A \, d\mu \ \leqq \ \int f \cdot 1_{A'} \, d\mu \ = \ \nu(A').$$

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 $B_k := \biguplus_{j \leqq k} A_j$. Then

$$0 \ \leqq \ \sum_{j=1}^{n} f \cdot 1_{A_j} \ = \ f \cdot 1_{B_n} \uparrow f \cdot 1_A,$$

and thus, by monotone convergence,

$$\nu(A) \ = \ \int f \cdot 1_A \, d\mu = \lim_{n \to \infty} \int f \cdot 1_{B_n} . \ = \ \lim_{n \to \infty} \sum_{j=1}^{k} \int f \cdot 1_{A_j} \ = \ \lim_{n \to \infty} \sum_{j=1}^{k} \nu(A_j) \ = \ \sum_{j=1}^{\infty} \nu(A_j) \ \blacksquare$$

4.5 The Standard Machine – Proving Theorems About Integration

Introduction 4.3. The easiest way to prove facts about integration in general and expectations in particular is often to proceed as follows.

Step 1:	prove the statement for indicator functions 1_A.
Step 2:	Use the linearity of $f \mapsto \int f d\mu$ to prove the statement for simple functions.
Step 3:	Approximate measurable $f \geqq 0$ by simple functions $f_n \uparrow f$ and use the Monotone Convergence Theorem to extend the result to such f.
Step 4:	Prove the case for general $f = f^+ - f^-$ by applying step 3 to f^+ and f^-.

Shreve calls this procedure the **standard machine**. $\square$

We proceed according to the standard machine to prove the following generalized version of SCF2 Theorem 1.5.1.

Theorem 4.9. $(\Omega, \mathfrak{F}, \mu)$ *be a measure space and let* $(\Omega', \mathfrak{F}')$ *be a measuragle 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 4.11 on p.54 and given by*

$$\mu_f(A') \ = \ \mu\{f \in A'\} \ = \ \mu\{\omega \in \Omega \ : \ f(\omega) \in A'\}.$$

If $g \geqq 0$ *or* $g \circ f$ *is integrable then*

$$\int g \circ f \, d\mu \ = \int g \, d\mu_f, \quad i.e., \quad \int g\big(f(\omega)\big) \, d\mu(\omega) \ = \ \int g(\omega') \, d\mu_f(\omega'). \tag{4.52}$$

PROOF:
Step 1. Assume that $g = 1_{A'}$ for some $A' \in \mathfrak{F}'$. Note that

$$1_{A'}\big(f(\omega)\big) \ = \ 1 \ \Leftrightarrow \ f(\omega) \in A' \ \Leftrightarrow \ \omega \in f^{-1}(A'),$$

thus

$$\int_\Omega 1_{A'}\big(f(\omega)\big)\,d\mu(\omega) \;=\; \int_\Omega 1_{f^{-1}(A')}(\omega)\,d\mu(\omega) \;=\; \mu f^{-1}(A') \;=\; \mu_f(A') \;=\; \int_{\Omega'} 1_{A'}(\omega')\,d\mu_f(\omega').$$

We have shown the validity of formula (4.52) for $g = 1_{A'}$.

Step 2. Let $g \geqq 0$ be a simple function $g = \sum\limits_{j=1}^{n} c_j 1_{A'_j}$ ($n \in \mathbb{N}, c_j \geqq 0$). It then follows from the linearity of the integral and what we already haven proven in step 1 that

$$\int_\Omega g\circ f\,d\mu \;=\; \sum_{j=1}^{n} c_j \int_\Omega 1_{A'_j}\circ f\,d\mu \;=\; \sum_{j=1}^{n} c_j \int_{\Omega'} 1_{A'_j}\,d\mu_f \;=\; \int_{\Omega'} g\,d\mu_f.$$

Step 3. Assume that g is a nonnegative, $\mathfrak{F}' - -\mathfrak{B}^1$ measurable function. For each nonnegative integer n let

$$B_{j,n} \;:=\; \left\{\frac{j}{2^n} \;\leqq\; g \;<\; \frac{j+1}{2^n}\right\} \quad (j = 0, 1, \dots, 4^n - 1),$$
$$g_n(\omega') \;:=\; \sum_{j=0}^{4^n-1} \frac{j}{2^n}\cdot 1_{B_{j,n}}(\omega').$$

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 $g_n \leqq g_{n+1}$ 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 $\frac{j}{2^n} \leqq g(\omega') < \frac{j+1}{2^n}$, then

$$\omega' \in B_{j,n}, \;\; \text{thus } g_n(x) \;=\; \frac{j}{2^n} \;\leqq\; g(\omega') \;<\; \frac{j+1}{2^n}, \;\; \text{thus } |g_n(\omega') - g(\omega')| \;<\; \frac{j+1}{2^n} \;-\; \frac{j}{2^n} \;=\; \frac{1}{2^n}.$$

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 $f \geqq 0$ 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

$$\int_\Omega g^+\circ f\,d\mu \;=\; \int_{\Omega'} g^+\,d\mu_f, \qquad \int_\Omega g^-\circ f\,d\mu \;=\; \int_{\Omega'} g^-\,d\mu_f, \tag{4.53}$$

We also have

$$\begin{aligned}(g^+\circ f)(\omega) \;&=\; g^+\big(f(\omega)\big) \;=\; \big[g\big(f(\omega)\big)\big]^+ \;=\; (g\circ f)^+(\omega),\\ (g^-\circ f)(\omega) \;&=\; g^-\big(f(\omega)\big) \;=\; \big[g\big(f(\omega)\big)\big]^- \;=\; (g\circ f)^-(\omega).\end{aligned} \tag{4.54}$$

It follows that

$$\begin{aligned}\int_\Omega |g \circ f|\, d\mu \;&=\; \int_\Omega (g \circ f)^+ \, d\mu \;+\; \int_\Omega (g \circ f)^- \, d\mu \\ &\overset{(4.54)}{=} \int_\Omega (g^+ \circ f)\, d\mu \;+\; \int_\Omega (g^- \circ f)\, d\mu \\ &\overset{(4.53)}{=} \int_{\Omega'} g^+ \, d\mu_f \;+\; \int_{\Omega'} g^- \, d\mu_f .\end{aligned}$$

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

$$\int_\Omega g \circ f \, d\mu \;=\; \int_{\Omega'} g^+ \, d\mu_f \;-\; \int_{\Omega'} g^- \, d\mu_f . \; \blacksquare$$

4.6 Equivalent Measures and the Radon–Nikodým Theorem

It is not necessary for you to learn the next definition. It is of a technical nature to ensure that certain important theorems are valid.

Definition 4.19 (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

$$\mu(A_n) \;<\; \infty \quad \text{for all } n, \qquad \text{and} \quad \bigcup_{n \in \mathbb{N}} A_n \;=\; \Omega. \; \square$$

Example 4.8. ⋆

- 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 $A_k \uparrow \Omega$.
- Summation measure Σ (Definition 4.7 on p.46) is σ–finite:
  For $k \in \mathbb{N}$ let $A_k := \{j \in \mathbb{Z} : |j| \leqq k\}$. Then $\Sigma(A_k) = 2k + 1 < \infty$, and $A_k \uparrow \mathbb{Z}$. $\square$

The next definition is an important one to remember.

Definition 4.20 (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 4.19 (σ–finite measure) on p.68), and let $f \geqq 0$ be in $m(\mathfrak{F}, \mathfrak{B}^1)$. If μ, ν, and f satisfy formula (4.51) of Theorem 4.8 on p.65, i.e.,

$$\nu(A) \;=\; \int_A f(\omega)\, d\mu(\omega), \quad \text{for all } A \in \mathfrak{F}, \tag{4.55}$$

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

$$f = \frac{d\nu}{d\mu} \quad \text{or} \quad d\nu = f\,d\mu \quad \text{or} \quad d\nu(\omega) = f(\omega)\,d\mu(\omega). \quad \square \tag{4.56}$$

Remark 4.18. It follows from Theorem 4.13 on p.76 that 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.
One can prove that under the given assumptions which include the σ–finiteness of μ and nonnegativeness of f and $\tilde{f}$ this almost everywhere uniqueness of the Radon–Nikodým derivative remains true, and this allows us to refer to "the" Radon–Nikodým derivative. $\square$

Remark 4.19. [⋆] We explain now why we call the function f in formula (4.55) a derivative. Consider the normal distribution with mean μ and variance σ^2, i.e., the measure ν on $\mathfrak{B}^1$ defined by

$$\nu(]a,b]) = \int_a^b f(x)\,dx = \int_{]a,b]} f\,d\lambda^1, \quad a,b \in \mathbb{R},\ a < b. \tag{4.57}$$

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 (4.57) extends to arbitrary Borel sets (see Fact 4.1 on p.44). In other words, if we write μ for λ^1, then λ^1, ν, and f satisfy formula (4.55), thus

$$f = \frac{d\nu}{d\lambda^1}.$$

Actually ν is already defined by its values on intervals of the form $]-\infty, x]$ since

$$\nu(]a,b]) = \nu(]-\infty,b]) - \nu(]-\infty,a]).$$

This is of course known to us: The $N(\mu,\sigma^2)$ distribution is given by its cumulative distribution function

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \int_{]-\infty,x]} f(x)\,d\lambda^1(x).$$

It follows from the Fundamental Theorem of Calculus that $f(x) = \frac{dF(x)}{dx}$. We have seen that $\int f(x)d\lambda^1(x)$ equals $\int f(x)dx$ for Riemann integrable f, so we take liberty and write dx for $d\lambda$. We have both

$$f(x) = \frac{dF(x)}{dx}, \qquad f(x) = \frac{d\nu(x)}{dx}.$$

This is the reason why a function f that satisfies formula (4.55) 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 (4.57) 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. $\square$

Definition 4.21 (μ–continuous measure). ⋆

Let μ and ν be measures on a given 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

$$\mu(A) = 0 \Rightarrow \nu(A) = 0, \quad \text{for all } A \in \mathfrak{F}. \ \square$$

- We call μ and ν **equivalent measures** and we write $\mu \sim \nu$, if both

$$\mu \ll \nu \quad \text{and} \quad \nu \ll \mu. \ \square \ \square$$

Remark 4.20.

(1) Two measures μ and ν on $(\Omega, \mathfrak{F})$ are equivalent if and only if

$$\mu(A) = 0 \Leftrightarrow \nu(A) = 0, \quad \text{for all } A \in \mathfrak{F}.$$

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. $\square$

Proposition 4.13. *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

$$\mu(A) = 0 \Rightarrow \int f 1_A \, d\mu = 0.$$

It suffices to show that $\int h d\mu = 0$ for all simple functions h that satisfy $0 \leqq h \leqq f 1_A$ since $\int f 1_A d\mu$ is the supremum of all sucn integrals.
Since $f 1_A = 0$ on $A^\complement$ and thus $0 \leqq h \leqq f 1_A = 0$ on $A^\complement$ we obtain $h = h 1_A$.
h has the form $h = \sum_{j=1}^{n} c_j 1_{A_j}$ for suitable $n \in \mathbb{N}$, $c_j \in \mathbb{R}$, and $A_j \in \mathfrak{F}$. Thus

$$\int h \, d\mu = \int h 1_A \, d\mu = \sum_j c_j \int_A 1_{A_j} \, d\mu = \sum_j c_j \mu(A \cap A_j) \leqq \sum_j c_j \mu(A) = 0.$$

The last equation follows from the assumption $\mu(A) = 0$. ■

Theorem 4.10 (Radon–Nikodým Theorem)**.** *Let μ and ν be measures on a given measurable space $(\Omega, \mathfrak{F})$ and assume moreover that the measure μ is σ–finite. Then*

$$\nu \text{ possesses a Radon–Nikodým derivative of } \nu \text{ with respect to } \mu \text{ on } \mathfrak{F} \quad \Leftrightarrow \quad \nu \ll \mu.$$

PROOF: ⋆ The "⇒" direction was proven in Proposition 4.13. The proof of the reverse direction is outside the scope of this lecture. ■

Corollary 4.2. *Let P and $\widetilde{P}$ be equivalent measures on a given measurable space $(\Omega, \mathfrak{F})$. Then both Radon–Nikodým derivatives $\frac{d\widetilde{P}}{dP}$ and $\frac{dP}{d\widetilde{P}}$ exist, and they satisfy the relation*

$$\frac{d\widetilde{P}}{dP} \cdot \frac{dP}{d\widetilde{P}} = 1 \text{ a.e.}$$

PROOF: ⋆ The Radon–Nikodým Theorem guarantees the existence of both $\frac{d\widetilde{P}}{dP}$ and $\frac{dP}{d\widetilde{P}}$. For all $A \in \mathfrak{F}$ we have

$$\textbf{(A)} \qquad \int_A \frac{d\widetilde{P}}{dP} \cdot \frac{dP}{d\widetilde{P}}\, d\widetilde{P} = \int_A \frac{d\widetilde{P}}{dP} \left(\frac{dP}{d\widetilde{P}}\, d\widetilde{P} \right) = \int_A \frac{d\widetilde{P}}{dP}\, dP = \int_A d\widetilde{P} = \widetilde{P}(A).$$

Let $Z := \frac{d\widetilde{P}}{dP} \frac{dP}{d\widetilde{P}}$. Assume to the contrary that we do not have $Z = 1$ a.e. Then

$$P\{Z > 1 + \varepsilon\} > 0 \quad \text{or} \quad P\{Z < 1 - \varepsilon\} > 0 \quad \text{for some suitably small } \varepsilon > 0.$$

We may assume that $P\{Z > 1 + \varepsilon\} > 0$. Then also $\widetilde{P}\{Z > 1 + \varepsilon\} > 0$ since $\widetilde{P} \sim P$. We write A for $\{Z > 1 + \varepsilon\}$ and obtain

$$\widetilde{P}(A) \overset{(\mathbf{A})}{=} \int_A Z\, d\widetilde{P} \geqq (1 + \varepsilon)\widetilde{P}(A) > \widetilde{P}(A).$$

We have reached a contradiction. ■

Remark: Assume as in Corollary 4.2 that P and $\widetilde{P}$ are equivalent measures. We write $Z := \frac{d\widetilde{P}}{dP}$ for convenience. Let $B_0 := \{Z = 0\}$. Then $\widetilde{P}(B_0) = 0$ because

$$\widetilde{P}(B_0) = \int_{B_0} Z\, dP = \int_{B_0} 0\, dP = 0.$$

Since $P \sim \widetilde{P}$ we also have $P(B_0) = 0$.

Let X be an arbitrary, nonnegative, random variable. Then

$$\int XZ\, dP = \int_{B_0} XZ\, dP + \int_{B_0^{\complement}} XZ\, dP = 0 + \int_{B_0^{\complement}} XZ\, dP. = \int_{B_0^{\complement}} X 1_{\{Z \neq 0\}} Z\, dP.$$

The above holds in particular for indicator functions $X = 1_A$ of any $A \in \mathfrak{F}$ and tells us that we may replace Z with $Z 1_{\{Z \neq 0\}}$. 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

$$\frac{d\widetilde{P}}{dP} = 1 \Big/ \frac{dP}{d\widetilde{P}}. \; \square$$

Remark 4.21. There is a more general version of the last corollary.

Let μ, ν, and ρ be three measures on a given measurable space $(\Omega, \mathfrak{F})$. Assume that the measures μ and ν are σ–finite, that $\rho \ll \nu$ and $\nu \ll \mu$. Then $\frac{d\rho}{d\mu}$ exists and satisfies

$$\frac{d\rho}{d\mu} = \frac{d\rho}{d\nu} \cdot \frac{d\nu}{d\mu} \quad \mu\text{–a.e.}$$

For the existence part observe that $\rho \ll \nu$ and $\nu \ll \mu$ implies $\rho \ll \mu$. □

4.7 Independence

All material in this chapter is standard an 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

$$P(A \cap B) = P(A) \cdot P(B) \quad \text{for all } A \in \mathfrak{G},\ B \in \mathfrak{H}.$$

- **(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 and Y are independent $\Leftrightarrow$ $P\{X \in U \text{ and } Y \in V\} = P\{X \in U\} \cdot P\{Y \in V\}$ for all Borel subsets U and V of $\mathbb{R}$.

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

$$P(A_1 \cap A_2 \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n) \quad \text{for all } A_j \in \mathfrak{G}_j, \quad j = 1, \dots n.$$

- **(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 areequivalent 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. □

Definition 4.22 (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 4.11 (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 4.12 (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 (image measure of P under the measurable functions $\omega \mapsto X(\omega), \omega \mapsto Y(\omega), \omega \mapsto \big(X(\omega), X(\omega)\big)$) factors:*

$$P_{X,Y}(A \times B) \;= P_X(A) \cdot P_Y(B) \text{ for all Borel sets } A, B \subseteq \mathbb{R}. \tag{4.58}$$

(3) *The joint cumulative distribution function factors:*

$$F_{X,Y}(a,b) \ = F_X(a) \cdot F_Y(b) \ \text{ for all } a, b \in \mathbb{R}. \tag{4.59}$$

(4) *The joint moment–generating function factors:*

$$E\left[e^{uX+vY}\right] \ = E\left[e^{uX}\right] \cdot E\left[e^{vY}\right] \ \text{ for all } u, v \in \mathbb{R} \tag{4.60}$$

for which the expectations are finite.

(5) *If there is a joint density then it factors:*

$$f_{X,Y}(x,y) \ = f_X(x) \cdot f_Y(y) \ \text{ for all } x, y \in \mathbb{R}. \tag{4.61}$$

The conditions above imply but are not equivalent to the following.

(6) *If there is a joint density then it factors:*

$$E[X \cdot Y] \ = E[X] \cdot E[Y], \ \text{ provided } E[|X \cdot Y|] < \infty. \tag{4.62}$$

PROOF (outline): See the SCF2 text. ■

4.8 Exercises for Ch.4

Exercise 4.1. 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

$$\int f \, d\mu \ = \ \int f \, d\mu'. \ \square$$

4.9 Addenda to Ch.4

Definition:
The following has been added to Definition 4.5 (Abstract measures) on p.43:

- We call any subset N of a set with measure zero a μ**–null set**. Note that N need not be measurable. $\square$

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

Proposition 4.14. [⋆]
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 $A_j : 1 \leqq j \leqq n$, and let $J := \mathbb{N}$ in case of a sequence $A_j : j \in \mathbb{N}$. Then our assumptions can be stated as follows.

$$A_i \cap A_j \ = \ \emptyset \ \text{ for } i \neq j, \quad \biguplus_{j \in J} A_j \ = \ \Omega, \quad \mathfrak{F} \ = \ \sigma\{A_j : j \in J\}. \tag{4.63}$$

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 $A_{n_1} \uplus A_{n_2} \uplus \dots$? ■

@@MFx-Auth

The following definition should be part of Definition 4.9 of a random variable. See p.50.

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 will call them random items.

Definition 4.23 (Random item).

- Let $(\Omega, \mathfrak{F}, P)$ be a probability space, $(\Omega', \mathfrak{F}')$ a measurable space, and let $X : \Omega \to \Omega'$ be $m(\mathfrak{F}, \mathfrak{F}')$. We call such a function a **random item**.

Note that all random variables are random items.

@@MFx-Auth

The following proposition belongs after Corollary 4.1 on p.52.

Proposition 4.15.

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

$$\{f < g\}, \quad \{f \leqq g\}, \quad \{f > g\}, \quad \{f \geqq g\},$$

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 $\{f \leqq g\}$ since

$$\{f \leqq g\} = \bigcap_{n \in \mathbb{N}} \left\{ f < g + \frac{1}{n} \right\}.$$

Lastly, $\{f > g\}$ and $\{f \geqq g\}$ are measurable as complements of the measurable sets $\{f \leqq g\}$ and $\{f < g\}$ ■

@@MFx-Auth

A

The following theorem belongs after Theorem 4.4 on p.60.

Theorem 4.13. *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.*

$$\text{If } \int_A f\,d\mu \leqq \int_A g\,d\mu \text{ for all } A \in \mathfrak{F} \quad \text{then} \quad f \leqq g \ \mu\text{–a.e.} \tag{4.64}$$

$$\text{If } \int_A f\,d\mu = \int_A g\,d\mu \text{ for all } A \in \mathfrak{F} \quad \text{then} \quad f = g \ \mu\text{–a.e.} \tag{4.65}$$

PROOF: ⋆ Let $A := \{f > g\}$. We will prove (4.64) by showing that the assumption $\mu(A) > 0$ leads to the contradiction $\int_A f d\mu > \int_A g d\mu$.

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 4.5 (Continuity properties of measures) on p.47.

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 $\mu(A_n) \geqq \gamma$. Since $f > g + \frac{1}{n}$ on all of A_n,

$$\int_{A_n} f\,d\mu \geqq \int_{A_n} \left(g + \frac{1}{n}\right) d\mu = \int_{A_n} g\,d\mu + \frac{1}{n}\mu(A_n) \geqq \int_{A_n} g\,d\mu + \frac{\gamma}{n} > \int_{A_n} g\,d\mu.$$

We have reached a contradiction, thus (4.64) holds.

Proof of (4.65): Note that, according to the already proven validity of (4.64), the assumption

$$\int_A f\,d\mu = \int_A g\,d\mu \text{ for all } A \in \mathfrak{F} \quad \text{implies} \quad f \leqq g \ \mu\text{–a.e.,} \quad \text{and} \quad g \leqq f \ \mu\text{–a.e.}$$

This proves $f = g$ μ–a.e. ■

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 define its conditional expectation $E[X \mid \mathfrak{G}]$ with respect to a sub–σ–algebra $\mathfrak{G}$ of $\mathfrak{F}$ not as a number but as a $\mathfrak{G}$–measurable random variable (a function of ω!) which satisfies the

$$\textbf{partial averaging property} \qquad \int_G E[X \mid \mathfrak{G}]\, dP \;=\; \int_G X\, dP \;\text{ for all } G \in \mathfrak{G}.$$

This property gets its name from the fact that it implies matching averages

$$\frac{1}{P(G)} \int_G E[X \mid \mathfrak{G}]\, dP \;=\; \frac{1}{P(G)} \int_G X\, dP \;\text{ for all } G \in \mathfrak{G} \text{ with probability } P(G) > 0\,,$$

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 **??** (Conditional Expectations and Their Core Properties in the General Case)

5.1 Functional Dependency of Random Variables

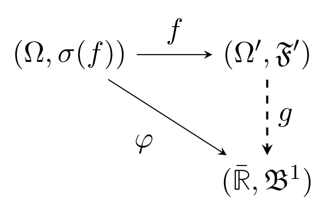
Proposition 5.1 (Doob Factorization Lemma). ⋆

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. Then f transforms Ω into a measurable space $\big(\Omega, \sigma(f)\big)$ by means of the σ–algebra

$$\sigma(f) \;=\; \{f^{-1}(A') : A' \in \mathfrak{F}\}.$$

See Definition 4.12 on p.54 and the proposition preceding it. Further assume that $\varphi : \Omega \to \bar{\mathbb{R}}$ is an extended real valued function with domain Ω. The following then is true:

- ***(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 $f \geqq 0$ then g can be chosen such that $g \geqq 0$.*
- ***(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 trivial 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 "⇒" is done according to the standard machine.

Step 1: φ is a $\sigma(f)$ measurable indicator function, i.e., $f = 1_A$ for some $A \in \sigma(f)$. But any such set A must the preimage $f^{-1}(A')$ for some $A' \in \mathfrak{F}'$. Note that if f is not bijective then A will generally not uniquely determine A'. We define

$$g := 1_{A'},$$

and it is easily verified that $g \circ f = 1_A$.

Step 2: For a nonnegative step function $\varphi := \sum_{j=1}^{k} c_j 1_{A_j}$ $(c_j \geqq 0, A_j \in \sigma(f))$, we define

$$g := \sum_{j=1}^{k} c_j 1_{A'_j},$$

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 $\varphi \geqq 0$ there exiss a sequence of simple functions φ_n such that $\varphi_n \uparrow \varphi$. See the proof of step 3 of Theorem 4.9 on p.66. 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 non–decreasing and thus has a $\mathfrak{F}'$–measurable limit g, and this function satisfies $\varphi = g \circ f$.

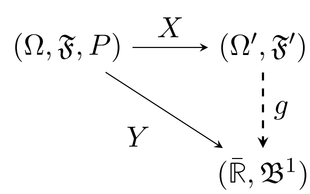
The proof of **(1)** for general g and that of **(3)** will not be given since it is somewhat tedious to consider the case $\infty - \infty$. But note that we also have done the proof of **(2)**. ■

Corollary 5.1. ⋆ *Given are a probability space $(\Omega, \mathfrak{F}, P)$, a measurable space $(\Omega', \mathfrak{F}')$, a random item' X by which we simply mean a $(\mathfrak{F}, \mathfrak{F}')$–measurable function X,* [16] *and a random variable Y on $(\Omega, \mathfrak{F}, P)$. Note that our assumptions imply*

$$\sigma(X) \subseteq \mathfrak{F} \quad \text{and} \quad \sigma(Y) \subseteq \mathfrak{F}$$

so that all probablities $P\{X \in A'\}$ and $P\{Y \in B\}$ exist for all $A' \in \mathfrak{F}'$ and $B \in \mathfrak{B}$.

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 (outline): This is an immediate consequence of the Doob Factorization Lemma, Proposition 5.1, since $\sigma(Y) \subseteq \sigma(X) \Leftrightarrow Y$ is $\big(\sigma(X), \mathfrak{B}^1\big)$–measurable ■

Remark 5.1. Given a probability space $(\Omega, \mathfrak{F}, P)$, a measurable space $(\Omega', \mathfrak{F}')$, and a random item X in $m(\mathfrak{F}, \mathfrak{F}')$, in particular, if $(\Omega', \mathfrak{F}') = (\mathbb{R}, \mathfrak{B}^1)$ and thus X is random variable,

> we can interpret the σ–algebra $\sigma(X)$ as the container of all stochastically relevant information of X

in the sense that 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.

[16]See Definition 4.23 on p.75 in the addenda to ch.sec:basic-meas-prob-theory.

In this context the corollary to the Doob Factorization Lemma says the following:

> If a random variable Y is stochastically known to a random variable X in the sense that its stochastically relevant information $\sigma(Y)$ is part of that of X, i.e., $\sigma(Y) \subseteq \sigma(X)$ then that by itself implies that Y is known to X on an ω by ω basis since the functional dependency $Y = g \circ X$ via the function $\omega' \mapsto g(\omega')$ determines $Y(\omega)$ from X as $g\big(Y(\omega)\big)$. $\square$

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 $sigma$–algebras σX and σY 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

$$\sigma(X) \text{ is } \underline{\text{not}} \text{ contained in } \mathfrak{G}.$$

- **(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 $\mathfrak{G} = \{\emptyset, \Omega\}$?

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 should only expect uniqueness of $X_{\mathfrak{G}}$ up to a set of probability zero. We have at least a partial answer to **(B)**:

$$X_{\mathfrak{G}} \text{ is only determined up to a set of probability zero.}$$

In other words, any random variable $X'_{\mathfrak{G}}$ in $m(\mathfrak{G}, \mathfrak{B}^1)$ which satisfies $X'_{\mathfrak{G}} = X_{\mathfrak{G}}$ P–a.e. 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,

$$G_i \cap G_j = \emptyset \text{ for } i \neq j, \quad \biguplus_{j \in J} G_j = \Omega, \quad \mathfrak{G} = \sigma\{G_j : j \in J\}. \tag{5.1}$$

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 4.14 on p.74. 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 the answer to question **(A)**.

(2) If $|J| = 2$ then $\Omega = G_1 \biguplus G_2$, 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

$$\begin{aligned} X_{\mathfrak{G}}(\omega) \ &:= \begin{cases} \frac{1}{P(G_1)}\, E\left[X 1_{G_1}\right] & \text{if } \omega \in G_1, \\ \frac{1}{P(G_2)}\, E\left[X 1_{G_2}\right] & \text{if } \omega \in G_2. \end{cases} \\ &= \frac{1}{P(G_1)}\, E\left[X 1_{G_1}\right] \cdot 1_{G_1}(\omega) \ + \ \frac{1}{P(G_2)}\, E\left[X 1_{G_2}\right] \cdot 1_{G_2}(\omega) \\ &= \sum_{j=1,2} \frac{1}{P(G_j)}\, E\left[X 1_{G_j}\right] \cdot 1_{G_j}(\omega). \end{aligned}$$

(3) For general J we take the weighted averages on each G_j and splice them into a function of ω:

$$X_{\mathfrak{G}}(\omega) \ := \ \frac{1}{P(G_j)}\, E\left[X 1_{G_j}\right] \text{ if } \omega \in G_j, \quad \text{i.e.,} \quad X_{\mathfrak{G}}(\omega) \ = \ \sum_{j \in J} \frac{1}{P(G_j)}\, E\left[X 1_{G_j}\right] \cdot 1_{G_j}(\omega).$$

We have to amend the equations given in **(2)** and **(3)** to account for the indices j for which $P(G_j) = 0$. 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 seen already that $X_{\mathfrak{G}}$ can be determined at best up to a P–null set $A := \biguplus_{J_0} G_j$ has probability zero as the countable union of P–null sets. Thus we do not change any stochastically properties if we set $X_{\mathfrak{G}}$ to some arbitrary, constant, value, most conveniently zero. In other words, we replace the definition given in **(3)** with

$$X_{\mathfrak{G}}(\omega) \ := \ \sum_{j \in J_1} \frac{1}{P(G_j)}\, E\left[X 1_{G_j}\right] \cdot 1_{G_j}(\omega). \tag{5.2}$$

We now quickly 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 if X is a discrete random variable or if there is a conditional density:

$$\begin{aligned} E[X \mid G] \ &= \ \sum_x x P\{X = x \mid G\} \quad \text{if } X \text{ is discrete, or} \\ E[X \mid G] \ &= \ \int_{-\infty}^{\infty} x f_{X|G}(x)\, dx \quad \text{if there is a conditional density } f_{X|G}(x), \text{ i.e.,} \\ &\qquad\qquad P(A \mid G) \ = \ \int_A f_{X|G}(x)\, dx \text{ for all events } A. \end{aligned}$$

We obtain for indicator functions $X = 1_A (A \in \mathfrak{F})$

$$\begin{aligned} X_{\mathfrak{G}}(\omega) \ &= \ \sum_j \frac{1}{P(G_j)}\, E\left[1_{G_j} 1_A\right] \cdot 1_{G_j}(\omega) \ = \ \sum_j \frac{P(G_j \cap A)}{P(G_j)} \cdot 1_{G_j}(\omega) \\ &= \ \sum_j P(A \mid G_j) \cdot 1_{G_j}(\omega) \ = \ \sum_j E(1_A \mid G_j) \cdot 1_{G_j}(\omega) \ = \ \sum_j E(X \mid G_j) \cdot 1_{G_j}(\omega). \end{aligned}$$

This relationship

$$X_{\mathfrak{G}}(\omega) \;=\; \sum_j E(X \mid G_j)\,\cdot 1_{G_j}(\omega). \tag{5.3}$$

between $X_{\mathfrak{G}}$ and conditional expectations of the form $E[X \mid G_j]$ can be extended by use of the standard machine to arbitrary nonnegative or integrable random variables X

The proposition following this introduction will show that the integral equation

$$\int_G X_{\mathfrak{G}}\, dP \;=\; \int_G X\, dP \tag{5.4}$$

holds for events $G \subseteq \mathfrak{G}$. 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 (5.4). Since this formula yields matching "averages"

$$\frac{1}{P(G)} \int_G X_{\mathfrak{G}}\, dP \;=\frac{1}{P(G)} \int_G X\, dP \tag{5.5}$$

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. $\square$

Proposition 5.2. $\boxed{\star}$ *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 (5.2):*

$$X_{\mathfrak{G}}(\omega) \;:=\; \sum_{j \in J_1} \frac{1}{P(G_j)}\, E\left[X 1_{G_j}\right] \cdot 1_{G_j}(\omega).$$

Then formula (5.4) holds for all $G \in \mathfrak{G}$.

PROOF: We employ the standard machine.

Step 1. If $X = 1_A$ for some $A \in \mathfrak{F}$ then for each $k \in J$,

$$\begin{aligned}
\int_{G_k} X_{\mathfrak{G}}\, dP \;&=\; \sum_{j \in J_1} \frac{1}{P(G_j)} \int_{G_k} E\left[1_A 1_{G_j}\right] \cdot 1_{G_j}\, dP \\
&=\; \sum_{j \in J_1} \frac{1}{P(G_j)} \int_{G_k} P(A \cap G_j) \cdot 1_{G_j}\, dP \\
&=\; \sum_{j \in J_1} \frac{1}{P(G_j)}\, P(A \cap G_j) \cdot P(G_k \cap G_j)\, dP.
\end{aligned}$$

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 $k = j$. Thus all terms in the sum except the one for $j = k$ vanish and we are left with

$$\begin{aligned}\int_{G_k} X_{\mathfrak{G}}\, dP \;&=\; \frac{1}{P(G_k)}\, P(A \cap G_k) \cdot P(G_k)\, dP \;=\; P(A \cap G_k) \\ &= \int_{G_k} 1_A\, dP \;=\; \int_{G_k} X\, dP.\end{aligned}$$

Since all elements of $\mathfrak{G}$ are a finite or infinite union $G_{j_1} \biguplus G_{j_2} \biguplus \cdots$ of the sets G_j this last result extends to

$$\int_G X_{\mathfrak{G}}\, dP \;=\; \int_G X\, dP.$$

for an arbitrary event $G \in \mathfrak{G}$.

Step 2. If $X = \sum\limits_{i=1}^{m} \alpha_i 1_{A_i}$ 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 **Step 1** for each random variable 1_{A_i}, then linearity of the integral,

$$\begin{aligned}\int_G X_{\mathfrak{G}}\, dP \;&=\; \sum_{j \in J_1} \frac{1}{P(G_j)} \int_{G_k} E\left[\sum_{i=1}^{m} \alpha_i 1_{A_i} 1_{G_j}\right] \cdot 1_{G_j}\, dP \\ &= \sum_{i=1}^{m} \alpha_i \left(\sum_{j \in J_1} \frac{1}{P(G_j)} \int_{G_k} E\left[1_{A_i} 1_{G_j}\right] \cdot 1_{G_j}\, dP \right) \\ &= \sum_{i=1}^{m} \alpha_i \int_G 1_{A_i}\, dP \;=\; \int_G \sum_{i=1}^{m} \alpha_i 1_{A_i}\, dP \;=\; \int_G X\, dP.\end{aligned}$$

This proves the proposition in particular 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^- and thus obtain it also for $X = X^+ - X^-$. ■

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 5.1 (Existence Theorem for Conditional Expectations). *Let $(\Omega, \mathfrak{F}, P)$ be a probability space, $\mathfrak{G}$ 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 $\mathfrak{G}$, i.e., $P_{\mathfrak{G}}$ and $\nu_{\mathfrak{G}}$ are the set functions defined as*

$$P_{\mathfrak{G}}(G) \;=\; P(G), \qquad \nu_{\mathfrak{G}}(G) \;=\; \nu(G), \;\; (G \in \mathfrak{G}).$$

*See Definition **??** (Restriction/Extension of a function) on p.**??**. Then $P_{\mathfrak{G}}$ is a probability measure and $\nu_{\mathfrak{G}}$ is a measure on the measurable space $(\Omega, \mathfrak{G})$, $\nu_{\mathfrak{G}} \ll P_{\mathfrak{G}}$. The Radon–Nikodým derivative*

$$E\big[X \mid \mathfrak{G}\big] \;:=\; \frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}$$

is $\mathfrak{G}$–measurable and plays the role of $X_{\mathfrak{G}}$ in formula (5.4) on p.81 in the following sense. $E\big[X \mid \mathfrak{G}\big]$ satisfies

$$\int_G E\big[X \mid \mathfrak{G}\big]\, dP \;=\; \int_G X\, dP \ \text{ for all } G \in \mathfrak{G}. \tag{5.6}$$

***(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 (5.6).

PROOF: $\star$

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.4.13 on p.70) as follows.

$$P_{\mathfrak{G}}(G) \,=\, 0 \;\Rightarrow\; P(G) \,=\, P_{\mathfrak{G}}(G) \,=\, 0 \;\Rightarrow\; \nu(G) \,=\, 0 \;\Rightarrow\; \nu_{\mathfrak{G}}(G) \,=\, \nu(G) \,=\, 0.$$

According to the Radon–Nikodým theorem this suffices to guarantee the existence of the Radon–Nikodým derivative, determined uniquely P–a.s. [17] 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 and $E\big[X \mid \mathfrak{G}\big]$ is ($\mathfrak{G}$–measurable and not only ($\mathfrak{F}$–measurable!

Now we prove formula (5.6). Let $G \in \mathfrak{G}$. Since the function $E\big[X \mid \mathfrak{G}\big]1_G$ is $\mathfrak{G}$–measurable it follows that

$$\int_G E\big[X \mid \mathfrak{G}\big]\, dP \;=\; \int E\big[X \mid \mathfrak{G}\big]1_G\, dP \;=\; \int E\big[X \mid \mathfrak{G}\big]1_G\, dP_{\mathfrak{G}} \;=\; \int_G E\big[X \mid \mathfrak{G}\big]\, dP_{\mathfrak{G}}. \tag{5.7}$$

(See Exercise 4.1 on p.74 for the second equation.) Further,

$$E\big[X \mid \mathfrak{G}\big] \;=\; \frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}, \quad \text{i.e., } \ E\big[X \mid \mathfrak{G}\big]\, dP_{\mathfrak{G}} \;=\; \frac{d\nu_{\mathfrak{G}}}{dP_{\mathfrak{G}}}\, dP_{\mathfrak{G}} \;=\; d\nu_{\mathfrak{G}}. \tag{5.8}$$

We obtain from equations (5.7) and (5.8) that

$$\int_G E\big[X \mid \mathfrak{G}\big]\, dP \;=\; \int_G d\nu_{\mathfrak{G}} \;=\; \nu_{\mathfrak{G}} \;=\; \nu(G) \;=\; \int_G X\, dP$$

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 (5.6) holds for X^+ and X^-. It is a straightforward exercise to show the validity of (5.6) from the linearity of the integral. ■

[17] For the a.s. uniqueness of the Radon–Nikodým derivative see Remark 4.18 on p.69.

Remark 5.2. We state once more that the partial averaging property (5.6) determines the $\mathfrak{G}$–measurable random variable $E\big[X \mid \mathfrak{G}\big]$ P–a.e. in the sense that if X^* is another $\mathfrak{G}$–measurable random variable such that

$$\int_G X dP = \int_G X^* dP \text{ for all } G \in \mathfrak{G}$$

then $P\{X^* \neq E\big[X \mid \mathfrak{G}\big]\} = 0.$ $\square$

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 any**(!)** random variable X^* that satisfies
- **(a)** **(Measurability)** X^* is $\mathfrak{G}$–measurable,
- **(b)** **$\mathfrak{G}$–Partial averaging** or **Partial averaging**

(5.9) $$\int_G X^* \, dP \;=\; \int_G X \, dP \text{ for all } G \in \mathfrak{G},$$

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 the 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 $E\big[X \mid \sigma(X)\big]$ 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.** $\square$

Remark 5.3. 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 term "partial averaging" indicates that only averages

$$\frac{1}{P(G)} \int_G X \, dP, \quad G \in \mathfrak{G} \text{ and } P(G) > 0,$$

are stochastically relevant information for $E\big[X \mid \mathfrak{G}\big]$. Those averages constitute just a part of the averages

$$\frac{1}{P(A)} \int_A X \, dP, \quad A \in \mathfrak{F} \text{ and } P(A) > 0,$$

are 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.79 where $\mathfrak{G}$ was generated by a partitioning sequence $\Omega = G_1 \biguplus G_2 \cdots$. In that case, we have

$$E[X \mid \mathfrak{G}](\omega) \;=\; \sum_{j \in J_1} \frac{1}{P(G_j)}\, E\left[X 1_{G_j}\right] \cdot 1_{G_j}(\omega), \tag{5.10}$$

where J_1 is the set of indices for which $P(G_j > 0)$. See formula 5.2 on p.80. 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 definitely improve this estimate. $\square$

Remark 5.4 (Factored conditional expectation)**.** ⋆ According to Proposition 5.1 (Doob Factorization Lemma) on p.77 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 function

$$E\big[X \mid Z\big] \;=\; g \circ Z, \tag{A}$$

where $z \mapsto g(z)$ is $\mathfrak{B}^1$–$\mathfrak{B}^1$ measurable. Very confusingly it is common to write

$$E[X \mid Z = z] \tag{B}$$

for this function $g(z)$. Thus 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 $Z(\omega)$, becomes

$$E[X \mid Z](\omega) \;=\; E[X \mid Z = Z(\omega)]. \;\square \tag{C}$$

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" which sounds awkward but I would not know to improve upon: the fact that a $\mathfrak{G}$–measurable random variable, i.e., one for which $\mathfrak{G}$ 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 $E[\dots]$.

Theorem 5.2. *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*

$$E[c_1 X + c_2 Y | \mathfrak{G}] \;= c_1 E[X|\mathfrak{G}] + c_2 E[Y|\mathfrak{G}]. \tag{5.11}$$

  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, XY is integrable, and X is $\mathfrak{G}$–measurable, then*

$$E[X \cdot Y | \mathfrak{G}] \;= X \cdot E[Y|\mathfrak{G}]. \tag{5.12}$$

  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 X is an integrable random variable, then*

$$E\big[E[X|\mathfrak{G}|\mathfrak{H}\big] = E[X|\mathfrak{H}]. \tag{5.13}$$

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*

$$E[X|\mathfrak{G}] = E[X]. \tag{5.14}$$

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 2.24 (Concave-up and convex functions) on p.27) and that X is integrable. Then*

$$E[\varphi \circ (X \mid \mathfrak{G}] \geqq \varphi\big(E[X \mid \mathfrak{G}]\big). \tag{5.15}$$

This equation also holds if we assume that X is nonnegative (rather than integrable), although both sides may be $+\infty$.

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*

$$E[\,E[X \mid \mathfrak{G}]\,] = E[X].$$

PROOF: The proof is left as an exercise. ■

Note that

$$E\big[E[X|\mathfrak{G}]\big] = E[X]. \tag{5.16}$$

This equation simply is the partial–averaging property applied to $G = \Omega$.
In other words, $E[X|\mathfrak{G}]$ is an **unbiased estimator** of X.

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 5.3 (Independence Lemma). *Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and let $\mathfrak{G}$ be a sub–σ–algebra of $\mathfrak{F}$. Suppose the random variables $X_1, \dots, X_K$ are $\mathfrak{G}$–measurable and 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 $y_1, \dots, y_L$, and define*

$$g(x_1, \dots, x_K) = Ef(x_1, \dots, x_K, Y_1, \dots, Y_L). \tag{5.17}$$

$$\textit{Then} \qquad E[f(X_1, \dots, X_K, Y_1, \dots, Y_L)|\mathfrak{G}] = g(X_1, \dots, X_K). \tag{5.18}$$

PROOF: See the outline given in the text. ■

6 Financial Models - Part 1

This entire chapter closely follows the book [3] Björk, Thomas: Arbitrage Theory in Continuous Time and we use to a large degree the notation found there.

Definition 6.1 (Adapted stochastic processes – Informal definition)**.** Everything considered in this chapter happens in the context of a once and for all given probability space $(\Omega, \mathfrak{F}, P)$. We must introduce informally a few concepts which will be properly defined in a later chapter.

(1) We consider on $(\Omega, \mathfrak{F}, P)$ a collection of sub–σ–algebras $\mathfrak{F}_t$ of $\mathfrak{F}$ which are associated with the trading times t, and which we interpret as the information aavailable up to time t in a financial market. It is natural to assume that information never decreases as time proceeds, and we assume that, if $t < t'$ then $\mathfrak{F}_t \subseteq \mathfrak{F}_{t'}$. We call $(\mathfrak{F}_t)_t$ the **information filtration** or also simply the **filtration** of the financial market.

(2) Let $(\Omega', \mathfrak{F}')$ be some measurable space. This will usually be $\mathbb{R}$ with its Borel σ–algebra. A **stochastic process** is a collection $X = X_t = X_t(\omega)$, indexed by the trading times t, of $\mathfrak{F}$–$\mathfrak{F}'$–measurable functions $X_t : \Omega \to \Omega'$. We say that X is **adapted to the filtration** $\mathfrak{F}_t$ and we call X an $\mathfrak{F}_t$**–adapted stochastic process** if X_t is actually $\mathfrak{F}_t$–$\mathfrak{F}'$–measurable for each time t.

(c) A stochastic process X_t is a **Markov process** if its future development only depends on its present state and not on any past information: $E[\varphi(X_T)|\mathfrak{F}_t] = E[\varphi(X_T)|X_t]$ for all nonnegative, Borel–measurable functions $x \mapsto \varphi(x)$. $\square$

Remark 6.1. See chapter 7.1 (Stochastic Processes and Filtrations) as follows.

- For the exact definition of a stochastic Process see Definition 7.1 on p.113.
- For the exact definition of a filtration see Definition 7.5 on p.116.
- For the exact definition of an adapted Process see Definition 7.6 on p.117.
- The definition of a Markov process is precise. See Remark 7.9 on p.118. $\square$

6.1 Basic Definitions for Financial Markets

Introduction 6.1. 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 claima**. 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 $X_T - K$ 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. What we just saw is that

$$\mathcal{X}(\omega) \;=\; \Phi\big(S_T(\omega)\big), \quad \text{where } \; \Phi(x) \;=\; (x-K)^+ \;=\; \max(x-K,0).$$

  We will write $\Pi(t;\mathcal{X})$ or $\Pi_t(\mathcal{X})$ for the price process of a contingent claim $\mathcal{X}$. 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? $\square$

Definition 6.2 (Financial Market)**.** A **financial market model**, usually just called a **financial market**, is defined in the context of a filtered is comprised of the following.

- 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, ...
- Trading times $t \geqq 0$ at which the assets $\mathscr{A}^{(j)}$ may be bought or sold. We speak of a **continuous time financial market** if those trading times consist of an interval t_0, T or t_0, ∞ and of a **discrete time financial market** if those trading times consist of a finite or infinite sequence $t_0 < t_1 < t_2 < \dots$ In either case we usually have $t_0 = 0$.
- Unit prices $\vec{S}_t(\omega) \;=\; \big(S_t^{(0)}(\omega), S_t^{(1)}(\omega), \dots, S_t^{(n)}(\omega)\big)$ of the assets $\mathscr{A}^{(j)}$.
- There usually will be a bank account asset. We then reserve slot zero for that asset. We often write B_t rather than $S_t^{(0)}$ to improve readability. $\square$

Remark 6.2. Interest is earned by holdings in a bank account and increases their value as time progresses. We will consider different ways in which interest is earned.

This can be as simple as the case of trading times $t = 0, 1, 2, 3, \dots$ with a fixed interest rate R per unit time. In this case the value of the holdings increases by the factor $1+R$, so it increases between times t and $t+k$ by a factor of $(1+R)^k$.

On the other end of the scale, if the interest rate is stochastic and varying in time, i.e., it is a stochastic process $R_t(\omega)$, then the value of the holdings increases between trading times t and t' by the factor $e^{\int_t^{t'} R_u du}$. $\square$

We list here a few more financial derivatives in addition to the European call.

Definition 6.3.

- 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

$$\Psi\big(S_T(\omega)\big), \quad \text{where } \; \Psi(x) \;=\; (K \,-\, x)^+ \;=\; \max(K-x,0).$$

- 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

$$\Psi\big(S_T(\omega)\big), \quad \text{where } \Psi(x) = x - K. \ \square$$

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 6.1 on p.87 gave the definition of a stochastic process adapted to the filtration $(\mathfrak{F}_t)_t$ which represents the market information up to time t For the definition of a portfolio strategy we replace each $\mathfrak{F}_t$ with the smaller σ–algebra

$$\mathfrak{F}_t^{\vec{S}} := \sigma\big(S_u^{(j)} : u \leqq t,\, 0 \leqq j \leqq n\big)$$

of the information generated, up to time t, by the price processes $S^{(0)}, \dots, S^{(n)}$. We will also utilize the definition of a Markov process given there.

Definition 6.4 (Portfolio strategy)**.**

A **portfolio** or **portfolio strategy** is an $\mathfrak{F}_t^{\vec{S}}$–adapted stochastic process

(6.1) $$\vec{H} = \vec{H}_t(\omega) = \big(H_t^{(0)}(\omega), H_t^{(1)}(\omega), \dots, H_t^{(n)}(\omega)\big)$$

which denotes the holdings (quantity) $H_t^{(j)}$ someone has in asset $\mathscr{A}^{(j)}$. Negative values indicate that this quantity is not owned but owed. We speak of a **Markovian portfolio** if $\vec{H}$ is a Markov process, i.e., it depends on current stock price $\vec{S}_t$ only and not also on its past, i.e., on all of $\mathfrak{F}_t^{\vec{S}}$.
We say that $\vec{H}$ has a **long position** in $\mathscr{A}^{(j)}$ if $H_t^{(j)} > 0$. We say that $\vec{H}$ has a **short position** in this asset if $H_t^{(j)} < 0$. $\square$

Definition 6.5 (Portfolio value in continuous time)**.** Assume that we have a portfolio $\vec{H}$ in a continuous time financial market.

The **portfolio value** associated with $\vec{H}$ is the stochastic process

$$V_t^H := \vec{H}_t \bullet \vec{S}_t = \sum_{j=0}^{n} H_t^{(j)} S_t^{(j)} = H_t^{(0)} S_t^{(0)} + H_t^{(1)} S_t^{(1)} + \cdots + H_t^{(n)} S_t^{(n)}. \square \tag{6.2}$$

Things are more complicated to formulate in the case of discrete trading times.

Definition 6.6 (Portfolio value in discrete time)**.**

Assume $\vec{H}$ is a portfolio in a discrete time financial market with a finite or infinite sequence of trading times $0 \leqq t_0 < t_1 < \cdots$. The **portfolio value** of $\vec{H}$ is the stochastic process

$$\begin{aligned} V_{t_0}^H &:= \vec{H}_{t_1} \bullet \vec{S}_{t_0} = \sum_{j=0}^{n} H_{t_1}^{(j)} S_{t_0}^{(j)}, \\ V_{t_k}^H &:= \vec{H}_{t_k} \bullet \vec{S}_{t_k} = \sum_{j=0}^{n} H_{t_k}^{(j)} S_{t_k}^{(j)} \text{ if } k \neq 0. \end{aligned} \tag{6.3}$$

Remark 6.3. Let us examine why portfolio value is defined differently in discrete and continuous trading models.

A. The continuous case.

We interpret, for each trading time t, $\vec{H}_t$ as the holdings (number of shares) in asset $\mathscr{A}^{(j)}$ at that precise 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, $\sum_{j=0}^{n} H_t^{(j)} S_t^{(j)}$, is the value of the entire portfolio at time t.

B. The discrete case.

B1. 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 value of those $\mathscr{A}^{(j)}$–holdings at that time t_k is based on the quantity $\vec{H}_{t_k}$ traded at the previous trading time t_{k-1} and on "today's" asset price $S_{t_k}^{(j)}$. This value thus is

$$\text{quantity} \times \text{price} = H_{t_k}^{(j)} \cdot S_{t_k}^{(j)},$$

and $V_{t_k}^H$ is the sum $\sum_{j=0}^{n} H_t^{(j)} S_t^{(j)}$ of those holdings.

B2. The case $k = 0$.

As in the case **B1**, $\vec{H}_{t_1}$ is interpreted as the holdings in asset $\mathscr{A}^{(j)}$ established at the intital trading time t_0 (and held constant until the future time t_1). The value of those $\mathscr{A}^{(j)}$–holdings at time t_0 is based on the quantity $\vec{H}_{t_1}$ traded "today" at time t_0 and on "today's" asset price $S_{t_0}^{(j)}$. This value thus is

$$\text{quantity} \times \text{price} \;=\; H_{t_1}^{(j)} \cdot S_{t_0}^{(j)},$$

and $V_{t_0}^H$ is the sum $\sum\limits_{j=0}^{n} H_t^{(j)} S_t^{(j)}$ of those holdings. □

Example 6.1. 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 then IBM stock contributes -3337.767 dollars to the value V_t^H of that portfolio. □

We stated earlier that money that is tied up in a riskless asset (zero coupon bond or bank account) will appreciate between times $t_1 < t_2$ by the amount $e^{\int_{t_0}^{t} R_s ds}$ where the process $R = R_t(\omega)$ is the interest rate at time t. We can turn this around and think of how much we are willing to pay at time t_1 for such a riskless asset if it pays the amount $Z_{t_2}(\omega)$ at time t_2. The answer is that we will discount that price to the amount

$$Z_{t_1} \;=\; e^{-\int_{t_1}^{t_2} R_s ds}\, Z_{t_1}$$

since this amount accrues, when invested at t_1 in a riskless asset, to the amount Z_{t_2}.

Definition 6.7 (Discount process)**.**

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 the process

(6.4) $$\text{Int}_t \;:=\; \exp\left[\int_0^t R_s\, ds\right]$$

represents the interest accrued between times 0 and t, i.e., an investment B_0 in a bank account at time zero will have accrued to $B_t = B_0\, \text{int}_t$ at time t. We call this process the **money market account price process** of and we call

(6.5) $$D_t \;:=\; \exp\left[-\int_0^t R_s\, ds\right]$$

the **discount process** of $\mathscr{A}^{(0)}$. □

Note that we have

(6.6) $$D_t \;=\; \frac{1}{\text{Int}_t} \;=\; \frac{B_0}{B_t}, \quad \text{assuming that Int}_t \neq 0.$$

The term money market account price process for Int_t has been adopted from SCF2 Chapter 6.5. It represents the value at time t of one currency unit which was invested in the riskless asset at time zero and continuously rolled over at the interest rate R_u, $0 \leqq u \leqq t$. □

Definition 6.8 (Self–financing Portfolio Strategy)**.** A portfolio strategy is a **self–financing portfolio strategy** 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. □

Remark 6.4. The above definition is not very mathematical and we should make precise its meaning. That is done by means of a so called **Budget equation** which will look completely different in the case of discrete time trading models such as the multiperiod binomial asset model (Chapter 6.2.2) than will be the case in continuous time trading models such as the Black–Scholes market (Chapter 9). □

Remark 6.5. In calculus quite a bit of time is spent on how to model reality using differentiation and integration. We interpret

$$f'(x) \;=\; \frac{dy}{dx} \;=\; \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{6.7}$$

as the quotient of an "infinitesimally small" change $dy(x)$ in function value $y = y(x)$ and an "infinitesimally small" change dx in argument x. This allows us to build a bridge between continuous and discrete time trading. We amend what was said in remark 6.3 about the continuous case as follows. Rather than viewing $\vec{H}_t$ as

"the holdings in asset $\mathscr{A}^{(j)}$ at that precise time t"

we have the following two choices.

(1) we interpret
  - $\vec{H}_t$ = the holdings in asset $\mathscr{A}^{(j)}$ purchased at time $t - dt$ at the asset price $S^{(j)}_{t-dt}$, i.e., the trade took place "just before" time t and we paid the price $S^{(j)}_{t-dt}$ in force at that time for each unit of the asset.

(2) we interpret
  - $\vec{H}_{t+dt}$ = the holdings in asset $\mathscr{A}^{(j)}$ purchased at time t at the asset price $S^{(j)}_t$, i.e., we use $t + dt$ which stands for the time "just after" t to label the positions traded at time t and purchased at the price $S^{(j)}_t$ in force at time t for each unit of the asset. □

We are now ready to elaborate, also in the continuous case, on Definition 6.8 of a self–financing portfolio (p.92) where it was stated that such a portfolio strategy must satisfy

(1) 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.
(2) There is no infusion of external money to purchase additional shares.

This will be done by means of a budget equation.

Definition 6.9 (Budget Equation)**.** A **Budget equation** for a portfolio states that exactly the value resulting from the holdings established at the last trading time before the present trading time will be used to establish the new holdings. This will look different for discrete time markets and continuous time markets.

Budget equation for a discrete time financial market:

$$\sum_{j=0}^{n} H_{t_{k+1}}^{(j)} S_{t_k}^{(j)} = V_{t_k}^{H} = \sum_{j=0}^{n} H_{t_k}^{(j)} S_{t_k}^{(j)} \quad \text{for } t_k > t_0. \tag{6.8}$$

In the case of continuous time trading we write, according to choice **(2)** of Remark 6.5, $H_{t+dt}^{(j)}$ for the holdings traded at time t:

Budget equation for a continuous time financial market:

$$\sum_{j=0}^{n} H_{t+dt}^{(j)} S_{t}^{(j)} = V_{t}^{H} = \sum_{j=0}^{n} H_{t}^{(j)} S_{t}^{(j)} \quad \text{for } t > t_0. \tag{6.9}$$

Note that time t_0 has been excluded because there is no reinvestment of the money that comes from a previous trade. □

Remark 6.6. Formula (6.9) for the continuous time budget equation is preliminary. We need knowledge about multidimensional stochastic calculus to transform it into

$$dV_t^H = \vec{H}_t \bullet d\vec{S}_t = \sum_{i=1}^{N} h_i(t)\, dS_i(t). \ \square$$

Proposition 6.1 (Self–financing criterion)**.**

A portfolio strategy $\vec{H}$ is self–financing if and only if it satisfies the budget equation of Definition 6.9.

PROOF: This follows from the observation that both budget equations simply state that precisely the value V_t^H of the entire portfolio is reinvested in that portfolio. Thus $\vec{H}$ is "rebalanced" at each time of trade without any inflow or outflow of funds. ■

Definition 6.10 (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, $\vec{H}_t$ is self–financing with a value process V_t^H which satisfies

(6.10) $$V_0^H = 0,$$

(6.11) $$P\{V_T^H \geqq 0\} = 1,$$

(6.12) $$P\{V_T^H > 0\} > 0. \quad \square$$

Note that the above is equivalent to replacing T with some $0 < t \leqq T$ since we can invest the positive amount V_t^H entirely into the riskless bond and (asuming a nonnegative interest rate) have at least that much profit at time T.

We are designing a model and will make some simplifying assumptions even though they may be unrealistic in the real world.

Assumption 6.1. The market adheres to the following:

- The processes $S_t^{(j)}$ and $H_t^{(j)}$ can equal any real number.
- There is no bid–ask spread: The trading house will not charge more to sell an asset to you than the price at which it is willing to 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 a bond).

The following condition is so central 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. $\square$

Definition 6.11 (Contingent Claim)**.** A **contingent claim** (**financial derivative**) is a $\mathfrak{F}_T^S$–measurable random variable $\mathcal{X}(\omega)$. We call $\mathcal{X}$ a **simple claims** if there is a function $s \mapsto \Phi(s)$ of stock price s such that

$$\mathcal{X} = \Phi \circ S_T.$$

We occasionally refer to Φ as the **contract function** of that claim. $\square$

Definition 6.12 (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 **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 6.7. We stress that part of the definition of a replicating portfolio is the condition that it be self–financing. □

Part of Assumption 6.1 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 6.1 (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

$$\Pi(\mathcal{X}) = V^H, \quad \textit{i.e.,} \quad \Pi_t(\mathcal{X}) = V_t^H, \textit{ for all trading times } t.$$

PROOF:

The case $t = T$ is immediate: We mentioned already in the introduction 6.1 to Chapter 6.1 on Basic Definitions for Financial Markets (see p.87) 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 $0 \leqq t_0 < T$ 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 us the proceeds to purchase a replicating portfolio $\vec{H}_{t_0}$ at its value, $V_{t_0}^H$.
3. We invest 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 $\Delta' \geqq \Delta$ at time $t = T$. The specific value of Δ' will depend on the interest rate process.
5. $\vec{H}_{t_0}$ 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 at time $t = t_0$.
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 $V^H_{t_0}$.
2. $t = t_0$: We us the proceeds to purchase a claim $\mathcal{X}$ at a price of $\Pi_{t_0}(\mathcal{X})$.
3. We invest the difference $\Delta := V^h_{t_0} - \Pi_{t_0}(\mathcal{X})$ in the riskless asset.
4. That investment will grow to Δ' at time $t = T$.
5. $\mathcal{X}$ will be worth V^H_T 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 at time $t = t_0$.
6. We have made a profit of Δ' without investing any of our own money. ■

6.2 The Binomial Asset Model

The binomial model is characterized by the following assumptions.

Assumption 6.2 (Binomial Asset Model)**.** Trading in this model only happens at times $t = 0, 1, 2, \dots$. Thus it is a discrete time financial market in the sense of Definition 6.2 (Financial Market) on p.88. There are only two assets.

(1) $\mathscr{A}^{(0)}$ is a bond/bank account. We denote its 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), thus

$$B_1 = (1+R)B_0, \ \dots, \ B_n = (1+R)B_{n-1} = (1+R)^n B_0. \tag{6.13}$$

(2) $\mathscr{A}^{(1)}$ is a stock. No superscripts are needed and 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

$$S_n = S_{n-1} \cdot Z_n = \begin{cases} u \cdot S_{n-1}, & \text{with probability } p_u > 0, \\ d \cdot S_{n-1}, & \text{with probability } p_d > 0\,, \end{cases} \tag{6.14}$$

$$\text{Here} \qquad Z_n := \begin{cases} u, & \text{with probability } p_u > 0, \\ d, & \text{with probability } p_d > 0\,. \end{cases} \tag{6.15}$$

is an iid sequence of Bernoulli trials 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 6.8 (Portfolio Strategy for the binomial model)**.**
According to Definition 6.4 (Portfolio Strategy) on p.89
a portfolio strategy for the binomial asset model is a process

$$\vec{H}_t(\omega) = \left(H^{(0)}_t(\omega), H^{(1)}_t(\omega)\right), \quad t = 1, 2, \dots, T \tag{6.16}$$

which denotes the holdings $H^{(j)}_t$ of an investor in $\mathscr{A}^{(0)}$ and $\mathscr{A}^{(1)}$ during the interval $[t-1, t]$. Negative values indicate that this quantity is not owned but owed.

Its portfolio value is

(6.17)
$$\begin{aligned} V_0^H &:= H_1^{(0)} S_0^{(0)} \;+\; H_1^{(1)} S_0^{(1)}, \\ V_t^H &:= H_t^{(0)} S_t^{(0)} \;+\; H_t^{(1)} S_t^{(1)} \text{ if t > 0. } \square \end{aligned}$$

We next specify the budget equation that must be satisfied by a self–financing portfolio. See Definition 6.9 (Budget Equation) on p.93.

Proposition 6.2 (Budget equation in the binomial asset model)**.** *A portfolio strategy*

$$\vec{H}_t(\omega) \;=\; \big(H_t^{(0)}(\omega), H_t^{(1)}(\omega)\big), \quad t = 1, 2, \dots, T$$

for binomial asset model (see Remark 6.8 on p.96) is self–financing if and only if the following condition holds for all $t = 0, \dots, T-1$

Budget equation:

(6.18)
$$H_t^{(0)}(1+R) \;+\; H_t^{(1)} S_t \;= H_{t+1}^{(0)} \;+\; H_{t+1}^{(1)} S_t.$$

PROOF: This is the self–financing criterion (Proposition 6.1 on p.93) since a bank account position established at time $t-1$ has increased, due to interest compounding, by a factor $1+R$. ■

If we use the notation of [3] Björk, Thomas: Arbitrage Theory in Continuous Time and write x_t for $H_t^{(0)}$, y_t for $H_t^{(1)}$ then the budget equation (6.18) reads

$$x_t(1+R) + y_t S_t \;= x_{t+1} + y_{t+1} S_t. \;\square$$

6.2.1 The One Period Model

In the one period model there are only two times $t = 0$ and $t = 1$. A portfolio $\vec{H}_0 = (H_0^{(0)}, H_0^{(1)})$ is purchased at $t = 0$.
We follow the notation of [3] Björk, Thomas: Arbitrage Theory in Continuous Time and write x for $H_0^{(0)}$, y for $H_0^{(1)}$. According to assumption 6.2, 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 6.3. *The model above is free of arbitrage if and only if the following conditions hold:*

(6.19)
$$d \;< (1+R) \;<\; u.$$

PROOF that if (6.19) does not hold then there will be arbitrage portfolios:
It is clear that there are arbitrage portfolios if $d \geqq 1+R$ since we can 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.

If $u \leqq 1 + R$ then we sell that stock short 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 6.1. See p.112. ■

We focus just on the stock price process $S = (S_0, S_1)$. 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 $A^{\complement} = \{S_1 = sd\}$ and $\sigma(S_1) = \{\emptyset, \Omega, A, A^{\complement}\}$ and then also $\sigma(S_0, S_1) = \sigma(S_1) = \{\emptyset, \Omega, A, A^{\complement}\}$.

We thus have completely determined the filtration $(\mathfrak{F}_t^S)_{t=0,1}$ generated by S as

$$\mathfrak{F}_0^S = \{\emptyset, \Omega\}, \qquad \mathfrak{F}_1^S = \{\emptyset, \Omega, A, A^{\complement}\}.$$

Let us restrict our probability space $(\Omega, \mathfrak{F}, P)$ to the events known by S, i.e., we downsize $\mathfrak{F}$ to $\mathfrak{F} := \sigma(S_0, S_1) = \mathfrak{F}_1^S$. 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 q_u such that $1 + R$ is the convex combination

$$1 + R = (1 - q_u)d + q_u u = q_u u + q_d d \quad (\text{define } q_d := 1 - q_u). \tag{6.20}$$

This pair of numbers, q_u and q_d, defines a probability measure Q on $(\Omega, \mathfrak{F})$ via

$$Q(\emptyset) := 0, \ Q(\Omega) := 1, \ Q(A) := q_u, \ Q(A^{\complement}) := q_d = 1 - q_u. \tag{6.21}$$

To summarize, absence of arbitrage allows us to define a probability measure Q on the information σ–algebra $\sigma(S_0, S_1) = \mathfrak{F}_1^S = \mathfrak{F}$ of stock price S. Q is equivalent to P (see Exercise 6.2 on p.112).

We have seen in formula (6.13) on p.96 that the interest factor by which a hank account holding increases between times zero and n is $(1 + R)^n$. we can turn this around and see that an asset worth V_n at time has to be discounted to $\frac{1}{(1+R)^n} V_n$ if one wants to determine how many units of the riskless asset $\mathscr{A}^{(0)}$ are needed today to generate that amount V_n at time n. We see that the discount process in the binomial model is given by

$$D_0 = 1, \ D_1 = \frac{1}{1 + R}, \ ; \dots, \ D_n = \frac{1}{(1 + R)^n}, \ \dots \tag{6.22}$$

Proposition 6.4. *The measure Q defined by q_u (and $q_d = 1 - q_u$) of formula (6.20) on $\mathfrak{F}_1^S$ satisfies the following.*

(a) *We obtain the present stock price from its price in the future by discounting that one and taking its expectation with respect to the measure Q:*

$$S_0 = \frac{1}{1 + R} \cdot E^Q[S_1], \tag{6.23}$$

(b) *The discounted stock price $M_n = D_n S_n, n = 0, 1$, is an $\mathfrak{F}_n^S$–martingale.*

PROOF of **(a)**: Let E^Q denote expectation with respect to the measure Q. Then

$$\begin{aligned} E^Q[S_1] &= (su)Q\{S_1 = su\} + (sd)Q\{S_1 = sd\} = (su)q_u + (sd)q_d \\ &= s\big(uq_u + dq_d\big) \overset{(6.20)}{=} s(1+R) = S_0(1+R). \end{aligned}$$

PROOF of **(b)**: Left as an exercise. ■

We give some definitions in the sequel which will be restated later in a more general context.

Definition 6.13 (Martingale Measure)**.** We call a probability measure Q that satisfies **(a)** and thus also **(b)** of Proposition 6.4 a **martingale measure**. We also call it a **risk–neutral measure**, since, using it for taking expectations, it has the following property. On average, when we account for the riskless ("risk–neutral") growth by discounting S_1, this discounted value must equal the (known) present value of the asset. □

We now compute the probabilities q_u and q_d which determine the martingale measure Q.

Proposition 6.5. *The martingale probabilities q_u and q_d of formula* (6.20) *on p.98 can be explicitly computed as*

$$q_u = \frac{(1+R)-d}{u-d}, \qquad q_d = \frac{u-(1+R)}{u-d}. \tag{6.24}$$

PROOF: Trivial. ■

Definition 6.14 (Contingent Claim)**.** A **contingent claim** (**financial derivative**) is a $\mathfrak{F}_1^S$–measurable random variable $\mathcal{X}(\omega)$.

Note that $\mathfrak{F}_1^S = \sigma(S_0, S_1) = \sigma(S_1)$ since $S_0 = s =$ const. Thus, by Doob's factorization lemma, there is a function $x \mapsto \Phi(x)$ of stock price x such that

$$\mathcal{X} = \Phi \circ S_1.$$

We occasionally refer to Φ as the **contract function** of that claim.

In a more general setting it will not always be true that $\mathcal{X}$ can be written as a function of stock price at expiration time. We then call contingent claims with a contract function Φ **simple claims**. □

To find an answer to the question how, in the one period model, a derivative $\mathcal{X}$ due at time $t = 1$ should be priced today, we introduce 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, and those holdings

$$\vec{H}_{t_j} = \big(H^0_{t_j}, H^1_{t_j}, \dots, H^n_{t_j}\big)$$

remain constant until t_{j+1}. In the finite 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^H_T takes place at $t = 0$.
- There are only two assets, the risk free asset with prices $B_t = B_0, B_1$, and the 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 we know today what our holdings are today

We recall our assumption that the market is efficient and that there is no arbitrage.

In a complete market all contingent claim can be hedged by a suitable portfolio. thus we know how to correctly price those claims at any point in time: We use the value of the corresponding hedge at that time. For this reason it is important to know conditions under which a market is complete. For the one period model we have the following.

Proposition 6.6. *If the one period binomial model is free of arbitrage then it is complete.*

PROOF: Let $\mathcal{X}$ be an arbitrary claim with contract function Φ, i.e.,

$$\mathcal{X} = \Phi \circ S_1$$

We claim that that the portfolio $\vec{H}_0 = (x, y)$, given by

$$\begin{aligned} x &= \frac{1}{1+R} \cdot \frac{u\Phi(sd) - d\Phi(su)}{u-d}, \\ y &= \frac{1}{s} \cdot \frac{\Phi(su) - \Phi(sd)}{u-d}. \end{aligned} \tag{6.25}$$

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 (6.13) and (6.14) on p.96 that, since $S_0 = \text{const} = s$, and since money market investments will increase by a factor $1 = R$, the portfolio $\vec{H}_0 = (x, y)$ 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

$$\begin{aligned} (1+R)x + suy &= \Phi(su), \\ (1+R)x + sdy &= \Phi(sd). \end{aligned}$$

This is a linear system of equations with determinant $x(1+R)sy \cdot (d-u)$ which is not zero since $d < u$. Thus there is a unique solution (x, y). It is easy to see that

$$\begin{aligned} x &= \frac{1}{1+R} \cdot \frac{u\Phi(sd) - d\Phi(su)}{u-d}, \\ y &= \frac{1}{s} \cdot \frac{\Phi(su) - \Phi(sd)}{u-d}. \quad \blacksquare \end{aligned} \tag{6.26}$$

We have seen in Proposition 6.4 on p.98 that discounted stock price is a martingale with respect to risk–neutral measure Q. The next proposition states that the same is true for (arbitrage free) pricing of contingent claims.

Proposition 6.7. *In the one period binomial model the discounted, arbitrage free, price process $D_t\Pi_t(\mathcal{X})$ of a contigent claim $\mathcal{X}$ is a Q–martingale. In particular, we have risk–neutral valuation*

$$\Pi_0(\mathcal{X}) \;=\; \frac{1}{1+R}\cdot E^Q[\mathcal{X}]. \tag{6.27}$$

PROOF: Let $\vec{H}$ be a hedging portfolio for $\mathcal{X}$. Since trading only takes place at $t=0$, $\vec{H}$ is determined by $(x,y) := \vec{H}_0$ Moreover,

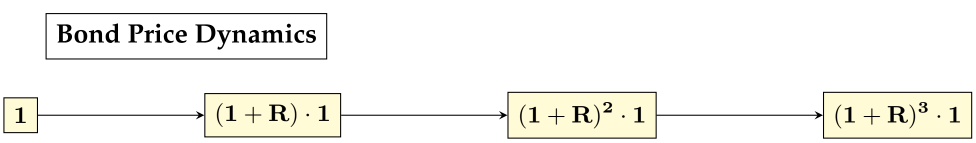
$$\Pi_0(\mathcal{X}) \;=\; V_0^H \;=\; x\cdot 1 \,+\, y\cdot s$$

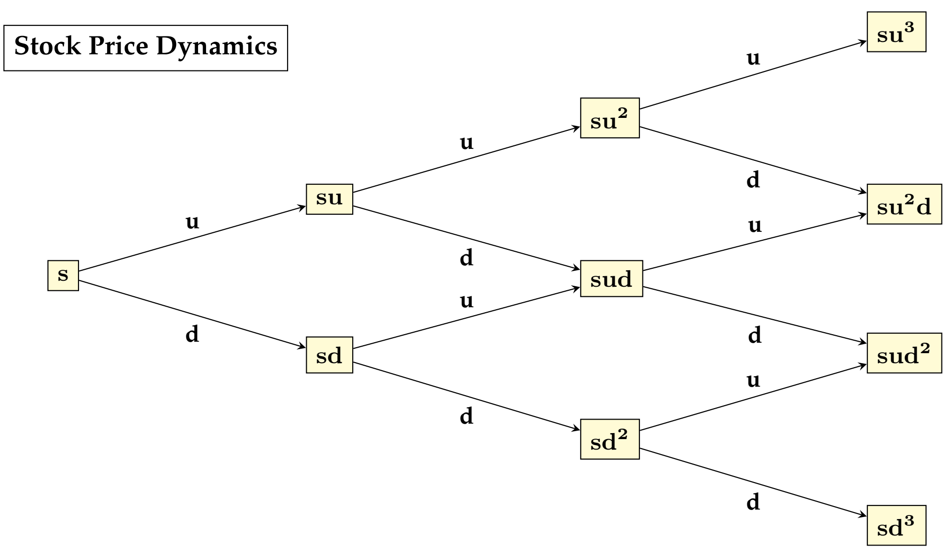
We use the expressions (6.26) (p.100) for x and y and afterwards the expressions (6.24) on p.100 for the martingale probabilities q_u and q_d. We obtain

$$\begin{aligned}\Pi(0;\mathcal{X}) \;&=\; \frac{1}{1+R}\cdot\Big[\frac{(1+R)-d}{u-d}\,\Phi(su) \;+\; \frac{u-(1+R)}{u-d}\,\Phi(sd)\Big]\\ &=\; \frac{1}{1+R}\cdot\big(\Phi(su)\cdot q_u \;+\; \Phi(sd)\cdot q_d\big) \;=\; \frac{1}{1+R}\,E^Q\left[\Phi\circ S_1\right] \;=\; \frac{1}{1+R}\,E^Q\left[\mathcal{X}\right].\end{aligned}$$ ■

6.2.2 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 6.2 for the binomial model that the dynamics that govern the development of the price B_t of the risk–less 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.





6.1 (Figure)**. Stock price dynamics**

Notations 6.1. We look at a vertical slice of the diagram in Figure 6.1 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 $\mathfrak{N}_{t_0,k}$ is the node that was reached from the start since 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 $su^k d^{t_0-k}$. 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 $\Pi(\mathfrak{N}_{t_0,k})$ for this arbitrage free price belonging to that node, i.e., to the stock price $su^k d^{t_0-k}$. We will see in Theorem 6.2 on p.103 that in an arbitrage free market every simple claim has such prices for every node in the tree. □

Some definitions such as that of a portfolio strategy and an arbitrage portfolio were already given in a generality sufficient for the multiperiod binomial model, but those for martingale measures were only established for the one period model and need to be generalized.

We next specialize Definition 6.10 (Arbitrage Portfolio) on p.93 to the multiperiod binomial model.

Definition 6.15 (Arbitrage portfolio in the multiperiod binomial model)**.** An **arbitrage portfolio** is a self-financing portfolio H with the properties

$$V_0^H = 0, \qquad P\{V_T^H \geqq 0\} = 1, \qquad P\{V_T^H > 0\} > 0. \ \square$$

Proposition 6.8. *The multiperiod model is free of arbitrage if and only if*

$$d < (1+R) < u. \tag{6.28}$$

PROOF: Same as for the one period case (Proposition 6.3 on p.97). ■

We remind the reader of Assumption 6.1 on p.94 about efficient market behavior.

- The binomial model is free of arbitrage and we thus assume that

$$d < (1+R) < u. \ \square$$

We next adapt Definition 6.13 (Martingale Measure) on p.99 to the multiperiod model, remembering from Proposition 6.4 which precedes it that a martingale measure was characterized by making the discounted stock price a martingale.

Definition 6.16 (Martingale Measure)**.** We call a probability measure Q that satisfies for all trading times $t = 0, 1, 2, \dots, T-1$ and for all possible values s' of S_t the relation

$$s' = \frac{1}{1+R} \cdot E^Q[S_{t+1} | S_t = s'],$$
$$\text{i.e., } S_t = \frac{1}{1+R} \cdot E^Q[S_{t+1} | S_t],$$

a **martingale measure** or also a **risk–neutral measure**. □

Remark:
Stock price S_t in the multiperiod model is clearly Markov since we either have $S_{t+1} = uS_t$ or $S_{t+1} = dS_t$, and thus S_{t+1} does not depend on stock price before t. It follows from the alternate characterization of the Markov property in Remark 7.9 on p.118 that if Y is a random variable that only depends on stock price information $S_t, S_{t+1}, S_{t+2}, \dots$ then

$$E^Q[Y \mid \mathfrak{F}^S_{t'}] \;=\; E^Q[Y \mid S_{t'}], \quad \text{for all } s' < t.$$

In particular, since $Y := S_t$ only depends on such information, it follows that

$$E^Q[S_t \mid \mathfrak{F}^S_{t'}] \;=\; E^Q[S_t \mid S_{t'}], \quad \text{for all } s' < t.$$

In particular it is true for martingale measure Q that discounted stock price is a Q–martingale, since

$$\frac{1}{1+R} \cdot E^Q[S_{t+1} \mid \mathfrak{F}_t] \;=\; \frac{1}{1+R} \cdot E^Q[S_{t+1} \mid S_t] \;=\; S_t. \;\square$$

Proposition 6.9. *In the multiperiod model that does not admit arbitrage there is a unique martingale measure Q. As in the one period model it is defined by the two "martingale probabilities"*

$$q_u \;=\; \frac{(1+R)-d}{u-d},$$
$$q_d \;=\; \frac{u-(1+R)}{u-d}.$$

PROOF: Same as the proof given for the one period model. See prop.6.4 on p.98 ■

Definition 6.17 (Contingent Claim)**.** A **contingent claim** (**financial derivative**) is a $\mathfrak{F}^S_T$–measurable random variable $\mathcal{X}(\omega)$. We call $\mathcal{X}$ a **simple claims** if there is a function $s \mapsto \Phi(s)$ of stock price s such that

$$\mathcal{X} = \Phi \circ S_1.$$

We occasionally refer to Φ as the **contract function** of that claim. □

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 6.2.

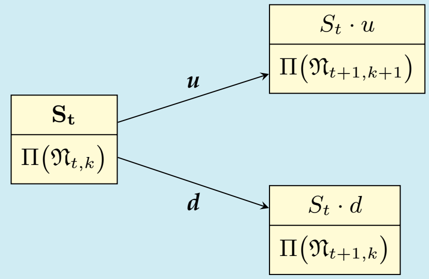
(1) *The discounted option price $\frac{1}{(1+R)^{T-t)}}\Pi_t(\mathcal{X})$ is an $\mathfrak{F}^S_t$–martingale.*

(2) *The option price is computed at time $0 \leqq t \leqq T$ for a stock price of $S_t(\omega) = su^k d^{t-k}$, attained by k upward moves and $t-k$ downward moves, as*

$$\Pi_t(\mathcal{X}) \;=\; \frac{1}{(1+R)^{T-t}} E^Q\big[\Phi(S_T) \mid S_t = su^k d^{t-k}\big]. \tag{6.29}$$

(3) *Every simple claim can be hedged. The portfolio quantities x_{t+1}, y_{t+1} for the node $\mathfrak{N}_{t,k}$ (remember: x_t, y_t = purchases at time $t-1$!) in the tree excerpt shown below are as follows.*

(6.30)
$$x_{t+1} = \frac{1}{1+R} \cdot \frac{u\Pi\big(\mathfrak{N}_{t+1,k}\big) - d\Pi\big(\mathfrak{N}_{t+1,k+1}\big)}{u-d}\,,$$
$$y_{t+1} = \frac{1}{s} \cdot \frac{\Pi\big(\mathfrak{N}_{t+1,k+1}\big) - \Pi\big(\mathfrak{N}_{t+1,k}\big)}{u-d}\,.$$



PROOF (outline): ⋆

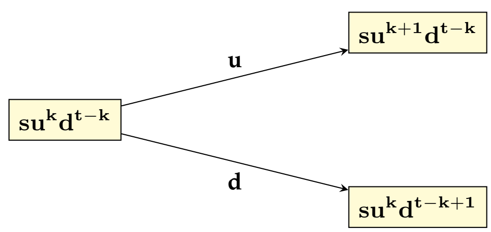
For the following recall the notation we introduced for the nodes of the binomial tree displayed in Figure 6.1 (Stock price dynamics) on p.102. Fix a time $0 \leqq t < T$ and assume that the arbitrage free claim price are known for all nodes at time $t+1$. We can consider those prices as the contract function Φ' of a new contingent claim

$$\mathcal{X}' = \Phi'(s') \text{ where } s' = sd^{t+1},\, sud^t,\, su^2d^{t-1}, \dots, su^td,\, su^{t+1}$$

runs through the stock prices that can be attained at time $t+1$.

Fix $0 \leqq k \leqq t$ 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 by either an upward move or a downward move in stock price are $\mathfrak{N}_{t+1,k+1}$ and $\mathfrak{N}_{t+1,k}$.

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 6.1 on p.102 we have

$$\Pi\big(\mathfrak{N}_{t+1,k+1}\big) = \Phi'(su^{k+1}d^{t-k}), \quad \text{and} \quad \Pi\big(\mathfrak{N}_{t+1,k}\big) = \Phi'(su^kd^{t-k+1}).$$

We apply the risk–neutral valuation formula (6.27) of Proposition 6.7 on p.101 to this one- period tree with the new contract function Φ' and 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

(6.31)
$$\begin{aligned}\Pi\big(\mathfrak{N}_{t,k}\big) &= \Pi_t(\mathcal{X}) = \Pi_0(\mathcal{X}') = \frac{1}{1+R} \cdot E^Q[\mathcal{X}'] \\ &= \frac{1}{1+R}\big(q_u \cdot \Phi'(su^{k+1}d^{t-k}) + q_d \cdot \Phi'(su^kd^{t-k+1})\big) \\ &= \frac{1}{1+R}\big(q_u \cdot \Pi\big(\mathfrak{N}_{t+1,k+1}\big) + q_d \cdot \Pi\big(\mathfrak{N}_{t+1,k}\big)\big).\end{aligned}$$

Now we see how to do a proof by induction.

The base case $T = 1$ is valid by direct application of the one period model.

Our (strong) induction assumption is that the proposition is valid for all simple claims $\mathcal{X}''$ in multi-period models with expiration time $T'' < T$.

We have seen above how to obtain from the claim $\mathcal{X} = \Phi(S_T)$ the option prices for the nodes corresponding to time $t = T-1$. As we did above we construct from them a new claim

$$\mathcal{X}'' = \Phi''(s'') \quad \text{where} \quad s'' = sd^T,\ sud^{T-1},\ su^2d^{T-2},\ \dots,\ su^{T-1}d,\ su^T.$$

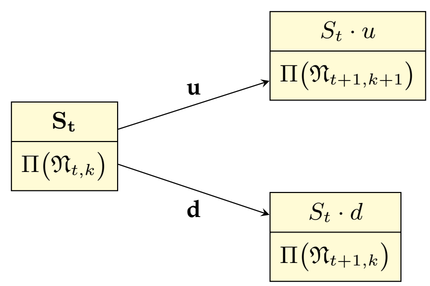
According to our induction assumption we obtain from this the arbitrage free option price for a stock price of $S_t(\omega) = su^kd^{t-k}$ attained by k upward moves and $t-k$ downward moves as

$$\Pi_t(\mathcal{X}'') = \frac{1}{(1+R)^{T-1-t}} E^Q\big[\Phi''(S_{T-1}) \mid S_t = su^kd^{t-k}\big]. \tag{6.32}$$

We combine this and formula6.32 and obtain (details being omitted, but we make use of the fact that the claim is a function of stock price and that stock price is a Markov process, hence conditioning the future on $\mathfrak{F}_t$ means the same as doing that conditioning on S_t)

$$\begin{aligned}\Pi\big(\mathfrak{N}_{t,k}\big) &= \frac{1}{(1+R)^{T-1-t}} \frac{1}{(1+R)} E^Q\big[E^Q\big[\Phi(S_T) \mid \mathfrak{F}_{T-1}\big] \mid \mathfrak{F}_t\big] \\ &= \frac{1}{(1+R)^{T-t}} E^Q\big[\Phi(S_T) \mid \mathfrak{F}_t\big]\end{aligned}$$

The above proves parts 1 and 2 of the theorem. The last part follows from the formulas 6.25 given in Proposition 6.6 on 100 for a replicating portfolio $\vec{H}_0 = (x, y)$ in the one period case. The corresponding tree is as shown on the right when we display stock price in the upper half and option price in the lower half of each node.

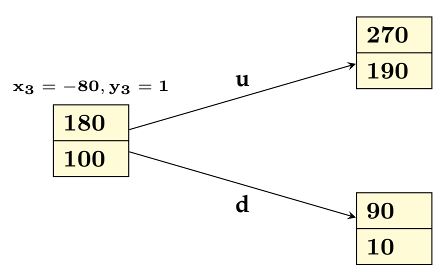


In other words the portfolio quantities for the node $\Pi\big(\mathfrak{N}_{t,k}\big)$ are given by

$$\begin{aligned}x_{t+1} &= \frac{1}{1+R} \cdot \frac{u\Pi\big(\mathfrak{N}_{t+1,k}\big) - d\Pi\big(\mathfrak{N}_{t+1,k+1}\big)}{u-d}, \\ y_{t+1} &= \frac{1}{s} \cdot \frac{\Pi\big(\mathfrak{N}_{t+1,k+1}\big) - \Pi\big(\mathfrak{N}_{t+1,k}\big)}{u-d}.\end{aligned}$$

Since this computation can be done for all nodes before expiration time in the tree we know how to set up the complete portfolia strategy $\vec{H}_t$ for $t = 0, 1, \dots, T-1$. ■

In the following we will draw trees which look like the one to the right. (We did so already in the proof of Theorem 6.2.) 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 correspontding replicating portfolio that correspond to that node.



The following example is taken from chapter 2 of [3] Björk, Thomas: Arbitrage Theory in Continuous Time.

Example 6.2. 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 $q_u = q_d = 0.5$.
Figure 6.1 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).

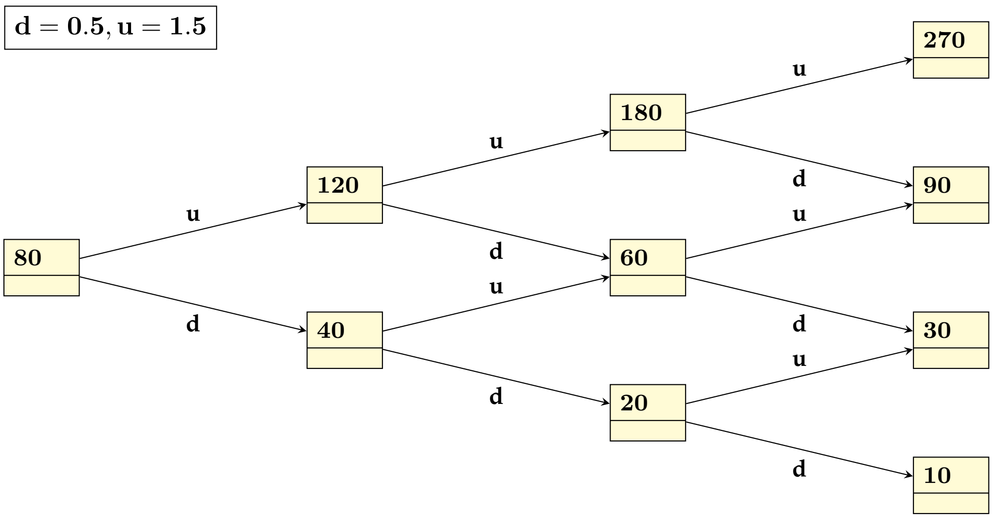


Figure 6.1: Stock prices.

The claim we want to price is a European call with a strike price of $K = \$80.00$, with an expiration date of $T = 3$.
This is a simple clain $\mathcal{X} = \Phi(S_T)$ 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 6.2 shows the updated tree.
We know from formula (6.31) on p.104 how to compute a claims price from those of the two child nodes to the right. With the notations introduced in Notations 6.1 on p.102,

$$\Pi\big(\mathfrak{N}_{t,k}\big) = \frac{1}{1+R}\big(q_u \cdot \Pi(\mathfrak{N}_{t+1,k+1}) + q_d \cdot \Pi(\mathfrak{N}_{t+1,k})\big).$$

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

$$\Pi\big(\mathfrak{N}_{2,2}\big) = \frac{1}{1+0}\big(0.5 \cdot 190) + 0.5 \cdot 10\big) = 100.$$

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

$$\Pi\big(\mathfrak{N}_{2,1}\big) = \frac{1}{1+0}\big(0.5 \cdot 10) + 0.5 \cdot 0\big) = 5.$$

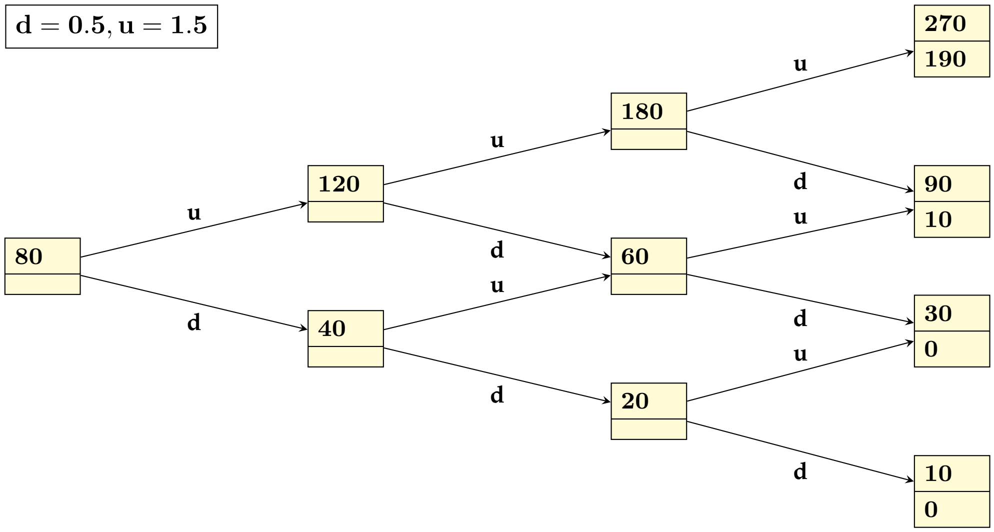


Figure 6.2: Stock prices and contract function values.

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

$$\Pi\big(\mathfrak{N}_{1,1}\big) \;=\; \frac{1}{1+0}\big(0.5 \cdot 100) \;+\; 0.5 \cdot 0.5\big) \;=\; 52.5.$$

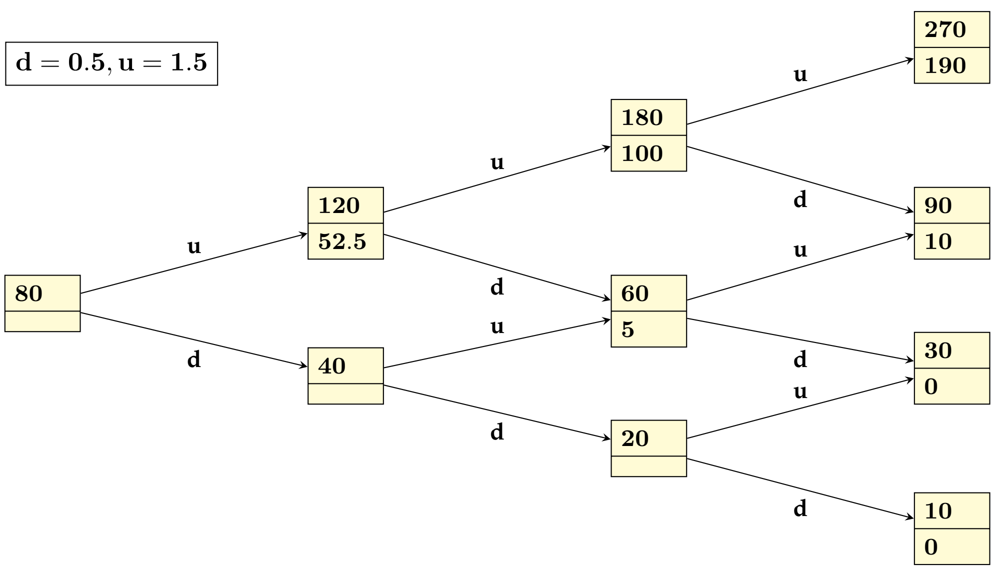


Figure 6.3: Stock prices and contract function values.

Figure 6.3 shows the tree with those additional values.
We compute the arbitrage free option prices for the remaining three nodes in this order:

$$\Pi\big(\mathfrak{N}_{2,0}\big),\ \Pi\big(\mathfrak{N}_{1,0}\big),\ \Pi\big(\mathfrak{N}_{0,0}\big).$$

The completed tree is shown in Figure 6.4.

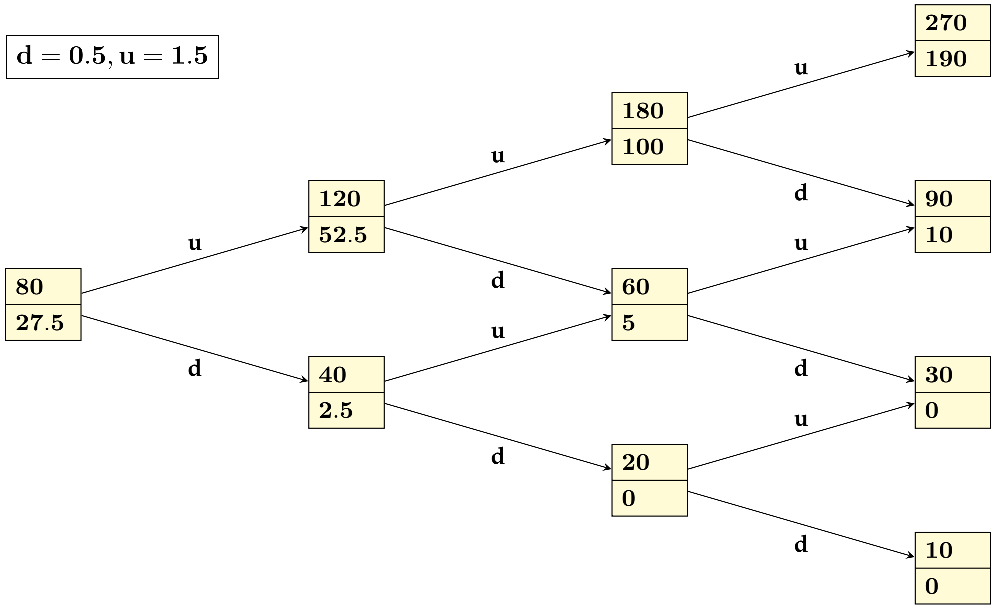
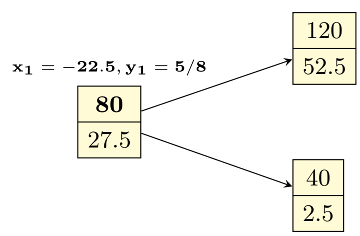


Figure 6.4: Completed tree with all option prices.

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 (6.30) of Theorem 6.2 on p.103 and obtain $x_1 = -22.5, y_1 = \frac{5}{8}$ since

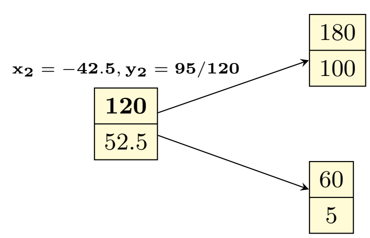


$x_1 = \frac{1}{1+0} \cdot \frac{1.5 \cdot 2.5 - 0.5 \cdot 52.5}{1.5 - 0.5} = \frac{3 \cdot 5 - 1 \cdot 105}{4} = -\frac{90}{4} = -22.5,$

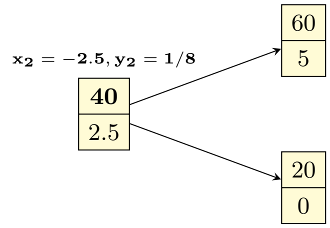
$y_1 = \frac{1}{s} \cdot \frac{\Phi(u) - \Phi(d)}{u - d} = \frac{1}{80} \cdot \frac{52.5 - 2.5}{1.5 - 0.5} = \frac{50}{80} = \frac{5}{8}.$

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 (6.30) 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 (6.30) 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 6.5.

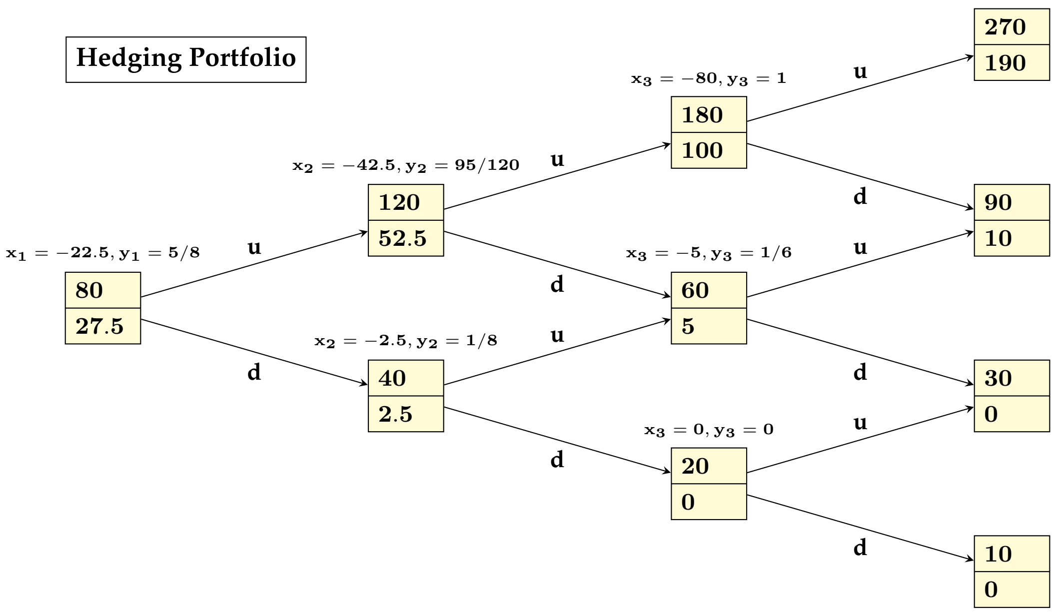


Figure 6.5: Hedging portfolio holdings.

This concludes the example. □

Remark 6.9. 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 q_u, q_d are constant for the entire tree since they only depend on u, d, and R. In this example they are

$$q_u = \frac{(1+R) - d}{u - d} = \frac{\frac{3}{2} - 1}{\frac{3}{2}, -\frac{1}{2}} = \frac{\frac{1}{2}}{1} = \frac{1}{2}, \qquad q_d = 1 - q_u = \frac{1}{2}$$

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

$$\Pi(2;\mathcal{X}) \;=\; \frac{1}{1+R}\Big[\, q_u \cdot \Pi(3;\mathcal{X}) \text{ of upward node } + \; q_d \cdot \Pi(3;\mathcal{X}) \text{ of downward node} \Big].$$

This method can be employed for any binomial tree, for arbitrarily many periods.

Step 3: 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 $\Pi(t-1;\boldsymbol{N})$:

$$\Pi(t-1;\boldsymbol{N}) \;=\; \frac{1}{1+R}\Big[\, q_u \cdot \Pi(t;\boldsymbol{N}_u) \;+\; q_d \cdot \Pi(t;\boldsymbol{N}_u) \Big]. \;\square$$

We mention again that this entire chapter 6 (Financial Models - Part 1) closely follows the book [3] Björk, Thomas: Arbitrage Theory in Continuous Time.

Notations 6.2. We will write

$$V\big(\mathfrak{N}_{t,k}\big) \;\; (0 \leqq t \leqq T),$$

for the value process of the replicating portfolio strategy, determined in Theorem 6.2 on p.103 by the formulas (6.30), when computed for the node $\mathfrak{N}_{t,k}$ of the binomial tree. $\square$

Proposition 6.10. *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 $\Pi_t(\mathcal{X})$ and $\vec{H}_t$ with their tree node equivalents, $\Pi\big(\mathfrak{N}_{t,k}\big)$ and $V\big(\mathfrak{N}_{t,k}\big)$, we have the following.*
The replicating portfolio is determined by the recursive formulas

$$\begin{aligned} V\big(\mathfrak{N}_{t,k}\big) \;&= \frac{1}{1+R}\big(q_u V\big(\mathfrak{N}_{t+1,k+1}\big) \;+\; q_d V\big(\mathfrak{N}_{t+1,k}\big)\big),\\ V\big(\mathfrak{N}_{T,k}\big) \;&= \Phi(s u^k d^{T-k}). \end{aligned} \tag{6.33}$$

Here q_u and q_d are the martingale probabilities from Proposition 6.9 on p.103, given by

$$q_u \;=\; \frac{(1+R)-d}{u-d}\,, \qquad\qquad q_d \;=\; \frac{u-(1+R)}{u-d}\,. \tag{6.34}$$

Further, the hedging portfolio quantities x_{t+1}, y_{t+1} for the node $\mathfrak{N}_{t,k}$ are

$$\begin{aligned} x_{t+1} \;&= \frac{1}{1+R}\cdot\frac{uV\big(\mathfrak{N}_{t+1,k}\big) - dV\big(\mathfrak{N}_{t+1,k+1}\big)}{u-d}\,,\\ y_{t+1} \;&= \frac{1}{s}\cdot\frac{V\big(\mathfrak{N}_{t+1,k+1}\big) - V\big(\mathfrak{N}_{t+1,k}\big)}{u-d}\,. \end{aligned}$$

In particular, the arbitrage free price of the claim at $t = 0$ is given by $V_0(0)$.

PROOF: This is just a rehash of Proposition 6.9 and Theorem 6.2 together with the pricing principle, Theorem **??** on p.**??**, which states that

$$V\left(\mathfrak{N}_{t,k}\right) \;=\; \Pi\left(\mathfrak{N}_{t,k}\right) \;\text{ for all nodes }\; \mathfrak{N}_{t,k} \;\text{ in the binomial tree.}$$

■

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 q_u under the risk–neutral measure (martingale measure) Q 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 6.11 (Arbitrage free price at time zero)**.** *The arbitrage free price at $t = 0$ of a simple claim $\mathcal{X}$ at time T is*

$$\Pi(0;\mathcal{X}) \;= \frac{1}{(1+R)^T}\cdot E^Q[\mathcal{X}], \tag{6.35}$$

where Q denotes the martingale measure, or more explicitly

$$\Pi(0;\mathcal{X}) \;= \frac{1}{(1+R)^T}\cdot \sum_{k=0}^{T}\binom{T}{k} q_u^k q_d^{T-k}\Phi(su^k d^{T-k}). \tag{6.36}$$

PROOF: (6.35) follows directly from the algorithm above.

To prove (6.36) we proceed as follows.

For $t = 1, 2, \dots, T$ let $Z_t(\omega) := 1$ if the stock price moves up between $t-1$ and t, and let $Z_t(\omega) := 0$ otherwise. We assume that we live in the risk–neutral world, so the probabilities for $Z_t = 1$ and $Z_t = 0$ are

$$P^Q\{Z_t = 1\} \;=\; q_u; \qquad P^Q\{Z_t = 0\} \;=\; q_d.$$

Let $K \;:=\; Z_1 + Z_2 + \cdots + Z_T$. Then K has a binom$(T; q_u)$ distribution since it tracks the number of successes (up–moves) of the iid Bernoulli trials $Z_1, \dots, Z_T$.

Since $K = k \;\Leftrightarrow\; S_T = su^k d^{T-k}$ and $\mathcal{X} = \Phi(S_T)$ by def.**??** (Contingent Claim) on p.**??** it follows that

$$\mathcal{X} \;=\; \Phi(S_T) \;=\; \Phi(su^K d^{T-K}).$$

For $k = 0, 1, \dots, T$ let $\psi(k) := \Phi(su^k d^{T-k})$. It follows for the expected value of $\mathcal{X}$ under risk–eutral measure that

$$\begin{aligned}
E^Q[\mathcal{X}] \;&= E^Q[\Phi(su^K d^{T-K})] \;=\; E^Q[\psi(K)]\\
&= \sum_{k=0}^{T}\psi(k)P^Q\{K=k\}\\
&= \sum_{k=0}^{T}\psi(k)\;\text{binom}(T;q_u)(k)\\
&= \sum_{k=0}^{T}\psi(k)\,\binom{T}{k}\,q_u^k\,q_d^{T-k}\\
&= \sum_{k=0}^{T}\Phi(su^k d^{T-k})\,\binom{T}{k}\,q_u^k\,q_d^{T-k}.
\end{aligned}$$

We apply (6.35) to the above and obtain (6.36). ■

We end this section by proving absence of arbitrage.

Proposition 6.12 (No arbitrage criterion)**.** *The market that belongs to the multiperiod binomial model is free of arbitrage if and only if*

$$d < 1 + R < u.$$

PROOF: The necessity follows from the corresponding one period result. Assume that the condition is satisfied. We want to prove absence of arbitrage, so let us assume that H (a potential arbitrage portfolio) is a self-financing portfolio satisfying the conditions

$$P\{V_T^H \geqq 0\} = 1, \quad P\{V_T^H > 0\} > 0.$$

We first show that $E^Q V_T^H > 0$ for such a portfolio. For $n \in \mathbb{N}$ let $\{A_n := V_T^H \geqq \frac{1}{n}\}$. Since $Q \sim P$, $P\{V_T^H > 0\} > 0$ implies $c := Q\{V_T^H > 0\} > 0$.
Since $A_n \uparrow \{V_T^H > 0\}$ implies $Q(A_n) \uparrow Q\{V_T^H > 0\}$ we get $Q(A_n) > \frac{c}{2}$ for some n. Thus

$$E^Q V_T^H = \int_\Omega V_T^H \, dQ \geqq \int_{A_n} V_T^H \, dQ \geqq \int_{A_n} \frac{1}{n} \, dQ = \frac{Q(A_n)}{n} \geqq \frac{c}{2n} > 0.$$

It follows from risk neutral valuation that

$$V_0^H = \frac{1}{(1+R)^T} \cdot E^Q[V_T^H] > 0.$$

This violates the condition $V_0^H = 0$ a.s. of an arbitrage portfolio H. ■

6.3 Exercises for Ch.6

Exercise 6.1. Prove the following part of Proposition 6.3 on p.97 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 6.2. We asserted that the probability measure Q defined by (6.21) on p.98 is equivalent to P on $\sigma(S_0, S_1)$. Prove it. □

7 Brownian Motion

7.1 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 $t_0 \leqq t \leqq T$ or $t_0 \leqq t < T$, where $0 \leqq t_0 < T \leqq \infty$. For example, $X(t,\omega)$ can the price of a stock at some time future time t which is unknown to us, and ω captures that uncertainty.

Definition 7.1 (Stochastic Process)**.** A **stochastic process** X on a probability space $(\Omega, \mathfrak{F}, P)$, often just called a **process**, is a collection of random items X_t which take values $X_t(\omega)$ in a measurable space $(\Omega', \mathfrak{F}')$. Being a random item, 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 $[\,t_0, \infty[$ and is interpreted as time. Usually the start time t_0 will be zero, and the end time T will be the time of expiration of a financial derivative.

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.

$$X \;=\; X_t \;=\; X(t) \;=\; (X_t)_t \;=\; \big(X(t)\big)_{t_0 \le t \le T} \;=\; X_t(\omega) \;=\; X(t,\omega) \;=\; \ldots$$

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 radom item X_t actually is a random variable. However we will also deal with vector valued stochastic processes

$$\vec{X} \;=\; \vec{X}_t \;=\; [X^1(t), X^2(t), \ldots, X^N(t)].$$

We will 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.

We will also sometimes use the notation $X(t,\cdot)$ if we want to emphasize X as the random variable which results when we look at the potential outcomes at a fixed time t. □

We will introduce some more terminlogy for random items indexed by time which do not qualify as stochastic processes in the sense of Definition 7.1 (Stochastic Process) on p.113 because the time index does not live in a contiguous interval.

Definition 7.2. 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 items 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 suitable as index set for a stochastic process, i.e., I is a contiguous interval, then we also refer to X as a **continuous time stochastic process**. with start time k_0.

(b) If I is an infinite, contiguous sequence of integers $0 \leqq k_0, k_0+1, k_0+2, \dots$ then we call X a **stochastic sequence**. with start time k_0.

(c) If I is an infinite sequence of real numbers $0 \leqq t_0 < t_1 < t_2 < \cdots$ or a finite sequence of real numbers $0 \leqq t_0 < t_1 < t_2 \leqq t_n = T$ then 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.

(d) If the index set of the form $I = 1, 2, \dots n$ and we interpret $X_1, \dots X_n$ as the coordinate values of an n–tuple then prefer to write

$$\vec{X} = (X_1, \dots X_n) \quad \text{or} \quad X(\vec{\omega}) = \big(X_1(\omega), \dots X_n(\omega)\big)$$

and call this a **random vector**. $\square$

Remark 7.1. 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 can classify collections of random items with an index set $I \subseteq [0, \infty[$ as
- continuous time stochastic processes,
- discrete time stochastic processes. $\square$

Before we can proceed we must discuss the information associated with a stochastic process. We recall Proposition 4.9 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 can define

$$\sigma(f, g) \ := \ \sigma\{A \subseteq \Omega : A = f^{-1}(A') \text{ for some } A' \in \mathfrak{F}' \text{ or } A = f^{-1}(A'') \text{ for some } A'' \in \mathfrak{F}''\},$$

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 7.3. 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 indexing those functions f_i by "time"! We call the σ–algebra

$$\sigma(f_i : i \in I) \ := \sigma\{\, A \subseteq \Omega \ : \ A = f_i^{-1}(A_i) \text{ for some } i \in I \text{ and } A_i \in \mathfrak{F}_i \,\} \tag{7.1}$$

generated by all preimages of measurable sets the **σ–Algebra generated by the family of functions** f_i $\square$

Remark 7.2. This last definition can be applied to the special case of a collection of random items $X_i, i \in I$ 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

$$\sigma(X_i : i \in I) \ = \sigma\big\{\, A \subseteq \Omega \ : \ A = \{X_i \in A_i\} \text{ for some } i \in I \text{ and } A_i \in \mathfrak{F}_i \,\big\}. \tag{7.2}$$

Note that since each X_i is a random item, each preimage $\{X_i \in A_i\}$ belongs to $\mathfrak{F}$, thus

$$\sigma(X_i : i \in I) \subseteq \mathfrak{F}.$$

> we can interpret $\sigma(X_i : i \in I)$ as the container of all stochastically relevant information of X

See Remark 5.1 on p.78. □

We are now back to stochastic processes and index sets I which can be interpreted as time intervals. As we just have seen we can associate with each stochastic process $X = (X_t), t \in I$ the collection $(\sigma(X_t))\,, t \in I$. However we are more interested in the stochastically relevant information not of X_t but that of the entire past of the process X up to time t which is stored in $\sigma\{X_s : s \leqq t\}$. This leads us to the definition of a filtration.

Definition 7.4 (Filtration for a process X_t)**.** For a continuous time or discrete time stochastic process X with index set I we define for $t \in I$,

$$\mathfrak{F}_t^X := \sigma\{X_s : s \in I, s \leqq t\} \tag{7.3}$$

As made plausible in Section 5.1 (Functional Dependency of Random Variables) in the context of a single random variable, we can think of $\mathfrak{F}_t^X$ as the stochastically relevant information of the process X for all times $s \leqq t$. 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. □

If this next remark just confuses you then you are advised to just **skip it!**

Remark 7.3. ⋆ For a fixed time t a ($\mathfrak{F}$–measurable) event A is an element of $\mathfrak{F}_t^X$ if there is a functional dependency of 1_A on the trajectory $X(s), s \in I, s \leqq t$, More generally a random variable Z is $\mathfrak{F}_t^X$–measurable if there is a functional dependency on the trajectory $X(s), s \in I, s \leqq t$. [18] □

It is very important that you understand the next example without trying to apply any mathematical reasoning.

Example 7.1 (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 $A = \{2.78 < X_s \leqq 3.14,\ \text{for } 5 \leqq s < 7\}$. then we have $A \in \mathfrak{F}_7^X$ but not $A \in \mathfrak{F}_{6.999}^X$ since observing the process X_s up to time $t = 6.999$ and seing that $2.78 < X_s \leqq 3.14$ for $5 \leqq s \leqq 6.999$ does not determine whether or not $2.78 < X_7 \leqq 3.14$.

(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.

[18] We are not mathematically exact here since we have do not have "product σ–algebras" available as a tool to appropriately generalize Doob's factorization lemma to families of random items $(X_t)_t$ in place of just a single X_t.

(3) Assume that X has continuous trajectories $s \mapsto X_s(\omega)$ so that the Riemann integral $Z(\omega) = \int_0^T X_u(\omega)du$ 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 is enough to understand the behavior of $\int_0^T X_u(\omega)du$. But note that Z is not $m(\mathfrak{F}_{T-\delta}^X)$ for any $\delta > 0$, no matter how small.

(4) Assume that X has continuous trajectories $s \mapsto X_s(\omega)$ and let

$$\tau(\omega \;:=\; \inf\{s \geqq 0 : X_s(\omega) \geqq 20\},$$

i.e., the random time τ denotes the first time that the trajectory enters the interval $[20, \infty[$. Then the event $\{\tau \leqq 8.5\}$ is in $\mathfrak{F}_{8.5}$ since

$$\tau(\omega) \leqq 8.5 \;\Leftrightarrow\; X_s(\omega) \geqq 20 \;\text{ for some } s \leqq 8.5,$$

and this clearly is determined by the behavior of $X_s(\omega)$ for $0 \leqq s \leqq 8.5$.

(4a) More generally assume again that X has continuous trajectories. Let γ be an arbitrary real number. Let

$$\tau(\omega \;:=\; \inf\{s \geqq 0 : X_s(\omega) \geqq \gamma\}$$

be the time of first entry into $[\gamma, \infty[$. Then $\{\tau \leqq t\}$ is in $\mathfrak{F}_t$ for any $t > 0$ since

$$\tau(\omega) \leqq t \;\Leftrightarrow\; X_s(\omega) \geqq \gamma \;\text{ for some } s \leqq t.$$

(5) Assume that X has continuous trajectories $s \mapsto X_s(\omega)$ and let

$$\rho(\omega \;:=\; \sup\{s \geqq 0 : X_s(\omega) \geqq 20\},$$

i.e., the random time ρ denotes the <u>last</u> time that the trajectory is inside the interval $[20, \infty[$. Then the event $\{\rho \leqq t\}$ is <u>not</u> in $\mathfrak{F}_t$ for any $t > 0$ since we cannot predict at time t the future behavior of the trajectory. (Why did I exclude the case $t = 0$?)

Remark 7.4. 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,

$$\text{if } s < t \quad \text{then } \mathfrak{F}_s^X \;\subseteq\; \mathfrak{F}_t^X.$$

This property by itself is so useful that we want to encapsulate it in its own definition. □

Definition 7.5 (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 $(\Omega, \mathfrak{F}, (\mathfrak{F}_t), P)$ 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 $X_s, s \in I, s \leqq t$ are determined by the member $\mathfrak{F}_t$ of a filtration $(\mathfrak{F}_t)_{t\in I}$.

Definition 7.6 (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 $X(s), s \in I, s \leqq t$ is determined by the information in $\mathfrak{F}_t$ for each time t, i.e., if

$$X_s \text{ is } \mathfrak{F}_t\text{–measurable for each } s \in I, s \leqq t,$$

then we say that X is **adapted to the filtration** $\mathfrak{F}_t$, □

Remark 7.5. 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 $W^{(1)}, \dots, W^{(m)}$ called **Brownian motions** or **Wiener processes**. By this we mean that each stock price $S^{(j)}$ is adapted to the filtration defined by

$$\mathfrak{F}_t \ := \ \sigma\big(W_s^{(j)} : 1 \leqq j \leqq m, s \leqq t, s \in I\big) \ \text{ for each } t \in I,$$

i.e., to the filtration generated by those Brownian motions.

Conditional expectations 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. □

Key properties of Brownian Motion will be that this process is both a martingale and a Markov process and that stock prices are Markov processes. The next two definitions which are about those two concepts are standard. They are straight copies from the SCF2 text.

Definition 7.7 (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. We call X

- **(a)** a **martingale** if $E[X_t \mid \mathfrak{F}_s] = X_s$ a.s., for all $s \leqq t$ such that $s, t, \in I$,
- **(b)** a **submartingale** if $E[X_t \mid \mathfrak{F}_s] \geqq X_s$ a.s., for all $s \leqq t$ such that $s, t, \in I$,
- **(c)** a **supermartingale** if $E[X_t \mid \mathfrak{F}_s] \leqq X_s$ a.s., for all $s \leqq t$ s.t. $s, t, \in I$. □

Remark 7.6. 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 $\Leftrightarrow$ $E[X_{n+1} \mid \mathfrak{F}_n] \geqq X_n$ a.s., for all $n \in \mathbb{N}$,
- **(c)** X is a supermartingale $\Leftrightarrow$ $E[X_{n+1} \mid \mathfrak{F}_n] \leqq X_n$ a.s., for all $n \in \mathbb{N}$. □

Remark 7.7.

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 7.1. *A martingale X satisfies $E[X_s] = E[X_t]$ for any $s, t \in I$.*

PROOF: The proof is left as an exercise. ■

Definition 7.8 (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 $\mathfrak{F}$, and 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 $0 \leqq s \leqq t \leqq T$ 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

$$E[f_t(X_t) \mid \mathfrak{F}_s] = f_s(X_s). \tag{7.4}$$

Then we call X a **Markov process** (with respect to the filtration $(\mathfrak{F}_t)_{t\in[0,T]}$. □

Remark 7.8. It is customary in this definition of a Markov process to consider the family $(f_u)_{u\in[0,T]}$ of functions with argument x as a single function $(u, x) \mapsto f(u, x)$. Definition 7.8 thus is equivalent to the following.

The process X is a Markov process if and only if the following is satisfied.
Let $0 \leqq t \leqq T$, and let f be an arbitrary, nonnegative, Borel–measurable function $x \mapsto f(x)$.
Then there is a function $(u, x) \mapsto f(u, x), 0 \leqq u \leqq t$, such that

$$E[f(t, X_t)|\mathfrak{F}_s] = f(s, X_s) \quad \text{for all } 0 \leqq s \leqq t. \tag{7.5}$$ □

Remark 7.9. 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 X is Markov:

The process X is a Markov process if and only if the following is satisfied.

Let $0 \leqq t \leqq T$, and let φ be an arbitrary, nonnegative, Borel–measurable function $x \mapsto \varphi(x)$. Then

$$E[\varphi(X_T)|\mathfrak{F}_t] = E[\varphi(X_T)|X_t]. \tag{7.6}$$

The interpretation is as follows: [19]

The future development of a Markov process does not depend on the past, only on the present.

The equivalence of (7.4) and (7.6) is not hard to see.

First assume that (7.4) holds true. Let φ be nonnegative and Borel–measurable. By assumption there is a function f_t that satisfies

$$E[\varphi(X_T) \mid \mathfrak{F}_t] = f_t(X_t).$$

[19] https://en.wikipedia.org/wiki/Markov_property

Since the right–hand side is a function of X_t the same must be true for the left–hand side, i.e., $E[\varphi(X_T) \mid \mathfrak{F}_t]$ is $\sigma(X_t)$–measurable. Thus

$$E[\varphi(X_T) \mid \mathfrak{F}_t] = E\big[E[\varphi(X_T) \mid \mathfrak{F}_t] \mid X_t\big] = E[\varphi(X_T) \mid X_t].$$

Here we have used "taking out what is known" followed by the Iterated Conditioning property. See Theorem 5.2 on p.85.

Now assume that (7.6) is satisfied. Let f_t be nonnegative and Borel–measurable and $s < t$. 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 factorization 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 (7.4). □

Remark 7.10. 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 $0 \leqq s \leqq t$" becomes "for all $s, t \in I$ such that $s \leqq t$".

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 (7.4), (7.5), (7.6), then it is customary to speak of a **Markov chain** rather than a time discrete Markov process. □

Example 7.2. Here are two informal examples of Markov chains.

(1) The random sequence $X(\omega) = X_n, n = 0, 1, 2, 3, \dots$, is defined as follows. We assume that $X_0 = n_0$ for some fixed $n_0 \in \mathbb{Z}$, and

$$X_n = \begin{cases} X_{n-1} + 1 & \text{with probability } 0 < p < 1, \\ X_{n-1} - 1 & \text{with probability } 1 - p. \end{cases}$$

Cleary, this sequence satisfies (7.6) 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 $S = S_n(\omega)$ of a stock at times $n = 0, 1, 2, 3, \dots$ develops according to the following rules: $S_0 = s_0$ for some fixed real number s_0, and

$$S_n \;=\; \begin{cases} u \cdot S_{n-1} & \text{with probability } 0 < p < 1, \\ d \cdot S_{n-1} & \text{with probability } 1-p, \end{cases}$$

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 (7.6) 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 6 (Financial Models - Part 1). □

7.2 Digression: Product Measures ⋆

This optional section is very skeletal and its only purpose is to justify certain properties $\int_a^b \int_\Omega X_s(\omega)\, d\omega\, ds$.

Definition 7.9 (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

(7.7) $$\mathfrak{F}_1 \otimes \mathfrak{F}_2 \;:=\; \sigma\{\, A_1 \times A_2 \;:\; A_1 \in \mathfrak{F}_1, A_1 \in \mathfrak{F}_1 \,\},$$

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

(7.8) $$A_1 \times A_2 \mapsto \mu(A_1)\, \nu(A_2)$$

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**, □

Example 7.3. 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 7.1 (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 $\mathfrak{F}_1$–measurable for each fixed ω_1.

If $f \geqq 0$ or f is $\mu \times \nu$–integrable then

$$\int_{A_1 \times A_2} f \, d\mu \times \nu = \int_{A_1} \left(\int_{A_2} f(\omega_1, \omega_2) \, d\nu(\omega_2) \right) d\mu(\omega_1) = \int_{A_2} \left(\int_{A_1} f(\omega_1, \omega_2) \, d\mu(\omega_1) \right) d\nu(\omega_2). \tag{7.9}$$

In particular switching the order of integration yields the same result.

Remark 7.11. ⋆

- 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 7.1 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:

$$\int_A f d\mu \times \nu = \int_{\Omega_1} \left(\int_{A_{\omega_1}} f \, d\nu \right) d\mu = \int_{\Omega_2} \left(\int_{A_{\omega_2}} f \, d\mu \right) d\nu. \quad \square$$

- We are interested in the case of an extended real valued continuous time stochastic process $X = X(t, \omega), t \in I$ which we assume $\mathfrak{B}(I) \otimes \mathfrak{F}$–measurable. Recall that expectations are integrals dP. Thus Fubini-Tonelli says that for $[a, b[\subseteq I$,

$$\int_{[a,b[\times\Omega} X d\lambda^1 \times P = \int_a^b E[X_t] \, dt = E\left[\int_a^b X_t \, dt\right]$$

7.3 Basic Properties of Brownian Motion

Definition 7.10 (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) $P\{t \mapsto W_t$ is continuous for ALL $t\} = 1$.
(4) Let $0 \leqq s < t < \infty$. Then the increment $W_t - W_s$ is independent of the σ–algebra $\mathfrak{F}_s$.
(5) Let $0 \leqq s < t < \infty$. Then $W_t - W_s \sim \mathcal{N}(0, t-s)$, i.e., $W_t - W_s$ is normal with

$$E[W_t - W_s] = 0, \qquad \text{Var}[W_t - W_s] = t - s. \quad \square \tag{7.10}$$

Remark 7.12. 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 $\mathfrak{F}_t^W$ defined as

$$\mathfrak{F}_t^W = \sigma(W_s : 0 \leqq s \leqq t).$$

In that case independence of the increments can be verified by showing that

(4') For any finite selection of times $0 \leqq t_0 < t_1 < \cdots < t_m < \infty$ the increments $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \ldots, W_{t_m} - W_{t_{m-1}}$ are independent.

In this case we simply speak of Brownian motion without mentioning the filtration $\mathfrak{F}_t^W$.
$\square$

A proof acceptable to mathematicians that definition 7.10 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**.

Definition 7.11 (Moment–generating function)**.** Let X be a random variable on a probability space $(\Omega, \mathfrak{F}, P)$. If u is a real number 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. If $\vec{X} = (X_1, \ldots, X_n)$ is a random vector on $(\Omega, \mathfrak{F}, P)$ and $\vec{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ 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,

$$\text{if } \vec{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n,\ \vec{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n, \quad \text{then } \vec{a} \bullet \vec{b} = \sum_{j=1}^n a_j b_j$$

denotes the standard inner product of $\mathbb{R}^n$

We can thus associate with any random variable X and random vector $\vec{X}$ the functions

(7.11) $$\Phi_X : \mathbb{R} \longrightarrow [0,\infty], \quad \text{defined as} \quad \Phi_X(u) = E\left[e^{uX}\right].$$

(7.12) $$\Phi_{\vec{X}} : \mathbb{R}^n \longrightarrow [0,\infty], \quad \text{defined as} \quad \Phi_{\vec{X}}(u) = E\left[e^{\vec{u}\bullet\vec{X}(\omega)}\right].$$

We call Φ_X (resp., $\Phi_{\vec{X}}$), the **moment–generating function** of X (resp., of $\vec{X}$). In the multi-dimensional case we also call $\Phi_{\vec{X}}$) the **joint moment–generating function** □

Proposition 7.2.

Let Z be a normal random variable with mean μ and variance σ^2 on a probability space $(\Omega, \mathfrak{F}, P)$. Then its moment–generating function is

(7.13) $$\Phi_Z(u) = e^{\mu u + \frac{1}{2}\sigma^2 u^2}.$$

PROOF: I was not able to locate the proof in [6] 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 7.3. *Let $W_t, 0 \leqq t < \infty$ be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. If $s, t \in [0, \infty[$ then*

(7.14) $$E[W_t] = 0,$$

(7.15) $$Cov[W_s, W_t] = E[W_s W_t] = \min(s,t).$$

PROOF: See SCF2, ch.3.3.2 ■

Proposition 7.4. ⋆

Let $W_t, 0 \leqq t < \infty$ be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$. Let $0 \leqq t_0 < t_1 < \cdots < t_m$. Then the covariance matrix for the m–dimensional random vector $\left(W_{t_1}, W_{t_2} \dots, W_{t_m}\right)$ is

(7.16) $$\begin{bmatrix} E[W_{t_1}W_{t_1}] & E[W_{t_1}W_{t_2}] & \dots & E[W_{t_1}W_{t_m}] \\ E[W_{t_2}W_{t_1}] & E[W_{t_2}W_{t_2}] & \dots & E[W_{t_2}W_{t_m}] \\ \vdots & \vdots & \ddots & \vdots \\ E[W_{t_m}W_{t_1}] & E[W_{t_m}W_{t_2}]\,; & \dots & E[W_{t_m}W_{t_m}] \end{bmatrix} = \begin{bmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \dots & t_m \end{bmatrix}$$

Moreover the moment–generating function for $\left(W_{t_1}, W_{t_2} \dots, W_{t_m}\right)$ is

(7.17) $$\begin{aligned} \varphi(u_1, \dots, u_m) &= E\left[\exp\left\{u_m W_{t_m} + u_{m-1}W_{t_{m-1}} + \dots + u_1 W_{t_1}\right\}\right] \\ &= \exp\left\{\frac{1}{2}(u_1 + u_2 + u_m)^2 t_1 + \frac{1}{2}(u_2 + u_3 + u_m)^2(t_2 - t_1) + \cdots \right. \\ &\qquad \left. \cdots + \frac{1}{2}(u_{m-1} + u_m)^2(t_{m-1} - t_{m-2}) + \frac{1}{2}u_m^2(t_m - t_{m-1})\right\}. \end{aligned}$$

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 7.2 (SCF2 Theorem 3.3.2 – Characterizations of Brownian motion). ⋆ *Let $(\Omega, \mathfrak{F}, P)$ be a probability space with a process $W_t, 0 \leqq t < \infty$ such that $W_0(\omega) = 0$ and the assignment $t \mapsto W_t(\omega)$ defines a continuous function of t for each $\omega \in \Omega$.*
Then we have equivalence

$$\textbf{(1)} \Leftrightarrow \textbf{(2)} \Leftrightarrow \textbf{(3)}$$

of the following:

(1) *For all $0 = t_0 < t_1 < \cdots < t_m$, 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 $0 = t_0 < t_1 < \cdots < t_m$, 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 (7.16).*

(3) *For all $0 = t_0 < t_1 < \cdots < t_m$, the random variables $W_{t_1}, W_{t_2} \dots, W_{t_m}$ have the joint moment–generating function (7.17).*

If any of **(1)**, **(2)**, **(3)**, *holds (and hence they all hold), then $W_t, 0 \leqq t < \infty$ is a Brownian motion with respect to $\mathfrak{F}^W_t$.*

PROOF: ■

The following is SCF2 Theorem 3.3.4.

Theorem 7.3 (Brownian motion is a martingale). *Let $W = W_t, t \geqq 0$, 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 $0 \leqq s \leqq t$, we have

$$\begin{aligned} E[W_t \,|\, \mathfrak{F}_s] &= E[\big(W_t - W_s\big) + W_s \,|\, \mathfrak{F}_s] \;=\; E[\big(W_t - W_s\big) \,|\, \mathfrak{F}_s] \;+\; E[W_s \,|\, \mathfrak{F}_s] \\ &= E[W_t - W_s] \;+\; W_s \;=\; W_s. \end{aligned}$$

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. ■

7.4 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 7.4. 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 all accept that

$$\text{(7.18)} \qquad \|\vec{x}\|_2 \ := \ \sqrt{\sum_{j=1}^{n} {x_j}^2} \qquad \text{and} \qquad d_2(\vec{x}, \vec{y}) := \|\vec{x} \ - \ \vec{y}\|_2$$

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) This is not quite as plausible but we might also be willing to accept

$$\text{(7.19)} \qquad \|\vec{x}\|_1 \ := \ \sum_{j=1}^{n} |x_j| \qquad \text{and} \qquad d_1(\vec{x}, \vec{y}) := \|\vec{x} \ - \ \vec{y}\|_1$$

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,

$$\text{(7.20)} \qquad \|f\|_{L^1} \ := \ \int_a^b |f(t)| \, dt \qquad \text{and} \qquad d_{L^1}(f, g) := \|f \ - \ g\|_{L^1}.$$

(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.

$$\text{(7.21)} \qquad \|f\|_{L^2} \ := \ \sqrt{\int_a^b f(t)^2 \, dt} \qquad \text{and} \qquad d_{L^2}(f, g) := \|f \ - \ g\|_{L^2}.$$

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 can 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 choese the density φ to be zero outside $[a,b]$. So now we define for $f, g : \mathbb{R} \to \mathbb{R}$, size and difference as follows.

$$\text{(7.22)} \qquad \|f\|_{L^2} \ := \ \sqrt{\int_{-\infty}^{\infty} f(t)^2 \varphi(t) \, dt} \qquad \text{and} \qquad d_{L^2}(f, g) := \|f \ - \ g\|_{L^2}.$$

This last example allows us 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.

$$\|f\|_{L^2} := \sqrt{\int_\Omega f(\omega)^2 \, d\mu(\omega)} \quad \text{and} \quad d_{L^2}(f,g) := \|f - g\|_{L^2}. \; \square \tag{7.23}$$

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 7.12 (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.

(7.24a)	$\|x\| \geqq 0$ for all $x \in V$ and $\|0\| = 0$	**positive semidefiniteness**
(7.24b)	$\|\alpha x\| = \|\alpha\| \cdot \|x\|$ for all $x \in V$, $\alpha \in \mathbb{R}$	**absolute homogeneity**
(7.24c)	$\|x + y\| \leqq \|x\| + \|y\|$ 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 7.13 (Pseudometric spaces)**.** Let X be an arbitrary, nonempty set.
A **pseudometric** on X is a real–valued function of two arguments

$$d(\cdot,\cdot) : X \times X \to \mathbb{R}, \qquad (x,y) \mapsto d(x,y)$$

with the following three properties:

(7.25a)	$d(x,y) \geqq 0$ and $d(x,x) = 0$ for all $x, y \in X$	**positive semidefiniteness**
(7.25b)	$d(x,y) = d(y,x)$ for all $x, y \in X$	**symmetry**
(7.25c)	$d(x,z) \leqq d(x,y) + d(y,z)$ 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 $d(x,y) < \varepsilon$. $\square$

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"

$$\|\vec{x}\| = 0 \;\Leftrightarrow\; \vec{x} = 0 \quad \text{and} \quad d(\vec{x},\vec{y}) = 0 \;\Leftrightarrow\; \vec{x} = \vec{y}. \tag{7.26}$$

On the other hand, regardless whether we interpret $\int \ldots dt$ 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, 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.

(7.27)
$$\int |f|\, d\mu = 0 \quad \Leftrightarrow \quad \int f^2\, d\mu = 0 \quad \Leftrightarrow \quad f = 0 \ \mu\text{–a.e.}$$

and thus

(7.28)
$$\int |f-g|\, d\mu = 0 \quad \Leftrightarrow \quad \int (f-g)^2\, d\mu = 0 \quad \Leftrightarrow \quad f = g \ \mu\text{–a.e.}$$

There is another difference but it is of more of a technical nature. We never have to worry about $\|\vec{x}\| = \infty$ or $d(\vec{x},\vec{y}) = \infty$. In contrast to this we have, 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 7.26 is so important that it leads to the following definitions which are a lot more important than those of seminorms and pseudometrics.

Definition 7.14 (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.

(7.29a)	$\|x\| \geqq 0$ for all $x \in V$ and $\|x\| = 0 \Leftrightarrow x = 0$	**positive definiteness**
(7.29b)	$\|\alpha x\| = \|\alpha\| \cdot \|x\|$ for all $x \in V$, $\alpha \in \mathbb{R}$	**absolute homogeneity**
(7.29c)	$\|x+y\| \leqq \|x\| + \|y\|$ for all $x, y \in V$	**triangle inequality**

The pair $(V, \|\cdot\|)$ is called a **normed vector space** $\square$

Definition 7.15 (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:

(7.30a)	$d(x,y) \geqq 0$ for all $x, y \in X$ and $d(x,y) = 0 \Leftrightarrow x = y$	**positive definiteness**
(7.30b)	$d(x,y) = d(y,x)$ for all $x, y \in X$	**symmetry**
(7.30c)	$d(x,z) \leqq d(x,y) + d(y,z)$ 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'll write X for short if it is clear which metric we are talking about. ☐

7.5 Quadratic Variation of Brownian Motion

Notations 7.1. In the following the letter Π will not denote the pricing function of a contingent claim as it did in Chapter 6 (Financial Models - Part 1), but a **partition**

$$\Pi = \{ t_0, t_1, \dots, t_n \}, \quad \text{where } 0 = t_0, < t_1, < \cdots < t_n = T$$

is interpreted as a set of times for a stochastic process with index set $I = [0, T]$ for some fixed $T > 0$. The step sizes $t_j - t_{j-1}$ are not assumed to be of equal size. We denote by

$$\|\Pi\| := \max\{t_{j+1} - t_j : j = 0, \dots, n-1\}.$$

the maximum step size (difference of neighboring times) of the partition. We will refer to $\|\Pi\|$ as the **mesh** of Π. ☐

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 7.16 (Quadratic Variation)**.** Let $f : [0, T] \to \mathbb{R}$ be a (Borel measurable) function of time t. We call

$$[f, f](T) := \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} [f(t_{j+1}) - f(t_j)]^2, \tag{7.31}$$

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

$$\int_a^b f(t)\,dt = \lim_{\|\Pi\| \to 0} \sum_{j=0}^{n-1} [f(t_j^*)(t_j - t_{j-1}), \ \ t_{j-1} \leqq t \leqq t_j,$$

in the definiton of the Riemann integral. In other words, the limit is taken along partitions $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ in such a way that the mesh becomes smaller and smaller. ☐

Remark 7.13 (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 $X_t, 0 \leqq t \leqq T$.

We will often write $[X, X]_T$ and $[X, X]_T(\omega)$ rather than $[X, X](T)$ and $[X, X](T, \omega)$. ☐

Remark 7.14. Let $f : [0,T] \to \mathbb{R}$ be a (Borel measurable) function with a continuous derivative. 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 $\omega \in \Omega$,

$$[W,W]_T(\omega) \;=\; T \quad \text{for all } T \geqq 0.$$

He actually proves a lot less:

Theorem 7.4. *Let W be a Brownian motion. For $T > 0$ and a partition $\Pi = \{ t_0, t_1, \dots, t_n \}$ of $[0,T]$, let*

$$Q_\Pi \;:=\; \sum_{j=0}^{n-1} (W_{t_{j+1}} - W_{t_j})^2.$$

Then

$$\lim_{\|\Pi\| \to 0} E\big[(Q_\Pi - T)^2\big] \;=\; 0.$$

PROOF: See the proof of SCF2 Theorem 3.4.3. ■

Remark 7.15 (About SCF2 Remarks 3.4.4). and 3.4.5]

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 have been able to 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. □

7.6 Brownian Motion as a Markov Process

Theorem 7.5. *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 $0 \leqq s \leqq t \leqq T$ and $f_t : \mathbb{R} \to [0, \infty,\; x \mapsto f_t(x)$ Borel–measurable. According to Definition 7.8 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

$$E[f_t(W_t) \mid \mathfrak{F}_s] \;=\; f_s(W_s). \tag{7.32}$$

It can be shown that

(7.33) $$f_s : \mathbb{R} \longrightarrow \mathbb{R}, \qquad x \mapsto E\left[f_t(x + W_t - W_s)\right]$$

is the sought after function. For the proof see SCF2 ch.3.5. ■

We will show that Brownian motion has a transition density as defined next.

Definition 7.17. ⋆

Let $X = X_t, 0 \leqq t < \infty$ 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

(7.34) $$p :]0, \infty \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}; (\tau, x, y) \mapsto p(\tau, x, y)$$

which satisfies, for every nonnegative Borel measurable function $f : \mathbb{R} \to \mathbb{R}$ the relation,

(7.35) $$E[f(X_t) \mid \mathfrak{F}_s] = \int_{-\infty}^{\infty} f(y)\, p(\tau, X_s, y)\, dy.$$

We call $p(\tau, x, y)$ the **transition density** for X. □

Remark 7.16. ⋆

Let $B \subseteq \mathbb{R}$ be a Borel subset and $f(x) := 1B(x)$. Then (7.35) becomes

(7.36) $$P\{X_{s+\tau} \in B \mid \mathfrak{F}_s\}(\omega) = E[1_B(X_t) \mid \mathfrak{F}_s](\omega) = \int_B p(\tau, X_s(\omega), y)\, dy.$$

We recall from Remark 7.9 on p.118 that the expressions above are $\sigma(X_s)$–measurable. This can also be seen directly since the random variable $\omega \mapsto \int_{\mathbb{R}} f(y)\, p(\tau, X_s, y)\, dy$ 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. This allows us to apply Doob factorization to $P\{\cdots \mid \mathfrak{F}_s\}$ just as we did in Remark 5.4 on p.85 to $E\{\cdots \mid \mathfrak{F}_s\}$. There is 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 in the case that discrete random variables or random variables with joint density functions are involved. Formula (7.36) becomes, if $X_s(\omega) = x$,

(7.37) $$P\{X_{s+\tau} \in B \mid X_s = x\} = \int_B p(\tau, x, y)\, dy.$$

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 assuming that it's 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 as they were the trajectories X_τ that had originally started at time zero in x. □

Proposition 7.5 (Transition density for Brownian motion)**.**

$$p(\tau, x, y) \;=\; \frac{1}{\sqrt{2\pi\tau}}\, e^{-\frac{(y-x)^2}{2\tau}}.$$

PROOF: The proof is given as part of SCF2 Theorem 3.5.1. ■

7.7 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 7.18 (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

$$S_t \;:= S_0\, \exp\left[\sigma W_t \;+\; \left(\alpha - \frac{1}{2}\sigma^2\right) t\right]. \tag{7.38}$$

geometric Brownian motion or also **GBM**. We will see in Example 8.1 on p.138 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 7.19 (Exponential martingale)**.**

Let $W = W_t, t \geqq 0$, 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 $Z = Z_t, t \geqq 0$, defined as

$$Z_t \;:= \exp\left[\sigma W_t \;-\; \frac{1}{2}\sigma^2 t\right], \tag{7.39}$$

the level σ **exponential martingale** of W. □

Z_t derives its name from the following theorem (SCF2 Theorem 3.6.1).

Theorem 7.6. *Let $W = W_t, t \geqq 0$, 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. ■

7.8 Exercises for Ch.7

8 One–Dimensional Stochastic Calculus

8.1 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 set, e.g., $[0, \infty[$ or $[t_0, T$ for some start time $0 \leqq t_0 < T$.

The next definition is from SCF2 ch.4.2.1.

Definition 8.1 (Simple Process)**.** Let
$\Pi := \{ t_0, t_1, \dots, t_n \}$, where $0 = t_0, < t_1, < \dots < t_n = T$ be a partition of $[0, T]$. for some fixed $T > 0$, with mesh $\|\Pi\| := \max \{ t_{j+1} - t_j : j = 0, \dots, n-1 \}$. 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. □

The next definition is from SCF2 ch.4.2.1.

Definition 8.2 (Itô Integral of a Simple Process)**.** Let $\Pi := \{ t_0, t_1, \dots, t_n \}$, where $0 = t_0, < t_1, < \dots < t_n = T$ be a partition of $[0, T]$ and let Z be a simple process on Ω which has constant trajectories on each partitioning interval $[t_j, t_{j+1}[$. We call

$$\int_0^t \Delta_u dW_u := \sum_{j=0}^{k-1} \Delta(t_j)[W(t_{j+1}) - W(t_j)] + \Delta(t_k)[W_t - W(t_k)]. \tag{8.1}$$

the **Itô integral** of Z with respect to W. □

Theorem 8.1 (SCF2 Theorem 4.2.1)**.** *The Itô integral* $\int_0^t \Delta_u dW_u$ *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

$$E[I_t] = 0 \text{ for all } t \geqq 0. \quad \text{Thus } Var[I_t] = E[I_t^2].$$

The next theorem shows how to evaluate $E[I_t^2]$.

Theorem 8.2 (SCF2 Theorem 4.2.2 - Itô isometry)**.** *The Itô integral defined by* (8.1) *on p.132 satisfies*

$$E[I_t^2] = E[\int_0^t \Delta_u^2 du]. \tag{8.2}$$

PROOF: See SCF2. ■

Theorem 8.3 (SCF2 Theorem 4.2.3). *The quadratic variation $[I,I]_t$ up to time t of the Itô integral $I_t = \int\limits_0^t Z_u dW_u$ is*

$$[I,I]_t = \int_0^t Z_u^2 du. \tag{8.3}$$

PROOF: See SCF2. ■

Remark 8.1. 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

$$d\int\limits_0^t Z_u\,dW_u = Z_t\,dW_t \tag{A}$$

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.

This makes a lot of sense for $Z_t = 1$. If we take the partition $\Pi = \{0,t\}$ then Definition 8.2 (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**) gives us exactly that. □

Remark 8.2. We write the Itô integral $I_t = \int\limits_0^t Z_u dW_u$ as a differential

$$d\,I_t = d\int\limits_0^t Z_u\,dW_u = Z_t\,dW_t.$$

We square both sides of this equation and obtain

$$dI_t\,dI_t = Z_t^2\,dW_t\,dW_t = Z_t^2\,dt. \ \square$$

8.2 The Itô Integral for General Processes

Definition 8.3 (L^2 convergence of stochastic processes). ⋆

We specialize the last observation from Example 7.4 on 124 to the following.

Given is a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ where $T > 0$ and $0 \leqq t \leqq T$. We can consider a stochastic process Y as a measurable function of two variables, (8.4) $$Y \,:\, ([0,T]\times\Omega, \mathfrak{B}^1\otimes\mathfrak{F}) \longrightarrow (\mathbb{R},\mathfrak{B}^1); \qquad (t,\omega)\mapsto Y_t(\omega).$$

See Definition 7.9 (Product spaces and product measures of two factors) on p.120 for the definition of $\mathfrak{B}^1\otimes\mathfrak{F}_t$.

> **A not very important note on the measurability of stochastic processes.** We have been making this measurability assumption and will keep doing so in this course without mentioning it explicitly, together with the following one for processes assumed adapted to the filtration $(\mathfrak{F}_t)_t$:
> Let t be fixed such that $0 < t \leqq T$. Then $(u,\omega)\mapsto X_u(\omega)$, considered as a function with domain $[0,t]\times\Omega$, is assumed to be $\mathfrak{B}^1\otimes\mathfrak{F}_t$–measurable for all $0\leqq u\leqq T$.

We will keep glossing over such technical arguments and are only mentioning it here because we integrate stochastic processes with respect to a product measure.

We apply formula (7.23) of Example 7.4 on 124 to the setting above and define L^2**–size** $\|Y\|_{L^2}$. and L^2**–distance** $d(Y,Y')$. of such stochastic processes as follows.

$$\|Y\|_{L^2} := \sqrt{\int_{[0,T]\times\Omega} Y_t(\omega)^2\, d(\lambda^1\times P)(t,\omega)}, \qquad \text{and} \qquad d_{L^2}(Y,Y') := \|Y - Y'\|_{L^2}.$$

If $Y \geqq 0$ or $\|Y\|_{L^2} < \infty$ then The Fubini–Tonelli (Theorem 7.1 on p.121) applies, and we obtain when writing $E[\dots]$ instead of $\int_\Omega \dots dP$ and using Riemann integral notation $\int_0^T \dots dt$ instead of Lebesgue integral notation $\int_{[0,T]}\dots d\lambda$,

(8.5) $$\|Y\|_{L^2} = \sqrt{E\left[\int_0^T Z_t^2\,dt\right]}.$$

Let $X, X^{(1)}, X^{(2)}, X^{(3)}, \cdots$ adapted, stochastic processes on Ω which satisfy $\|X\|_{L^2} < \infty$ and $\|X^{(n)}\|_{L^2} < \infty$ for all n. We say that the sequence $X^{(n)}$ **converges in $\mathbf{L^2}$** to X, and we write

(8.6) $$L^2\!-\lim_{n\to\infty} X^{(n)} = X, \;\text{ if }\; \lim_{n\to\infty} d_{L^2}(X^{(n)},X) = 0, \;\text{ i.e., }\; \lim_{n\to\infty} E\left[\int_0^T (X_t^{(n)} - X_t)^2\,dt\right] = 0. \;\square$$

Remark 8.3. We emphasize that $\|X\|_{L^2}$ is a pseudonorm only, not a genuine norm, and $d_{L^2}(Y,Y')$ is a pseudometric only, not a genuine metric. If a sequence of processes $X^{(n)}$ converges in L^2 to a process X and X' is another process such that the set

$$A := \{\omega\in\Omega : X(\cdot,\omega) \text{ and } X'(\cdot,\omega) \text{ differ on a set of Lebesgue measure} \neq \text{zero}\}$$

has probability zero then we also have $L^2\!-\lim\limits_{n\to\infty} X^{(n)} = X'$.

Here is an example. $\Omega=[0,1]$, $\mathfrak{F}=\mathfrak{F}_1=\mathfrak{B}^1$ for $0\leqq t\leqq 1$,

$$X_t(\omega) = \begin{cases} 1 & \text{if } t=\omega,\\ 0 & \text{if } t\neq\omega,\end{cases} \qquad X_t^{(n)}(\omega) = X'_t(\omega) = 0 \text{ for all } 0\leqq t,\omega\leqq 1 \text{ and } n\in\mathbb{N}.$$

Then $L^2\!-\lim\limits_{n\to\infty} X^{(n)} = X = X'\,(=0)$ even though there is no ω for which the trajectories $X(\cdot,\omega)$ and $X'(\cdot,\omega)$ are identical. $\square$

Fact 8.1. *Let $T > 0$ and $Z_u, 0 \leqq t \leqq T$, be an adapted stochastic process that is* ***square–integrable****, i.e.,*

$$\|Z\|_{L^2} = \sqrt{E\left[\int_0^T Z_t^2\,dt\right]} < \infty. \tag{8.7}$$

Then the following is true.

(a) *One can find a sequence $Z(n)$ of simple processes that are also square–integrable such that $L^2\text{–}\lim\limits_{n\to\infty} Z^{(n)} = Z$ (see formula (8.6)).*

(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.,*

$$\lim_{n\to\infty} E\left[\int_0^T (I_u - \Phi_u)^2\,du\right] = 0.$$

(c) *We write*

$$\int\limits_0^t Z_u^{(n)}\,dW_u := \Phi_t$$

for the process $\Phi = \Phi_t$ described in ***(b)*** *and call it the* ***Itô integral*** *of Z with respect to W.* □

Theorem 8.4 (SCF2 Theorem 4.3.1 - Itô isometry)**.** *The process $I_t := \int\limits_0^t Z_u^{(n)} dW_u$ defined in Fact 8.1 for square integrable and adapted Z satisfies the following.*

a. ***(Continuity)*** *As a function of the upper limit of integmtion t, the paths of I_t are continuous.*

b. ***(Adaptivity)*** *For each t, I_t is $\mathfrak{F}_t$–measurable.*

c. ***(Linearity)*** *If $I_t = \int\limits_0^t \Delta_u dW_u$ and $J_t = \int\limits_0^t \Gamma_u dW_u$,*

then $I_t \pm J_t = \int\limits_0^t \Delta_u dW_u \pm \int\limits_0^t \Gamma_u dW_u$;

furthermore, for every constant c, $cI_t = c\int\limits_0^t \Delta_u dW_u$.

d. ***(Martingale)*** *I_t is a martingale.*

e. ***(Itô isometry)*** *$E[I_t^2] = E\int\limits_0^t \Delta_u^2 du$.*

f. ***(Quadratic variation)*** *$[I, I]_t = \int\limits_0^t \Delta_u^2 du$.*

PROOF: Not given. ■

8.3 The Itô Formula for Functions of Brownian Motion

Theorem 8.5 (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 $T \geqq 0$,

$$f(T, W_T) - f(0, W(0)) = \int_0^T f_t(t, W_t)dt + \int_0^T f_x(t, W_t)dW_t + \frac{1}{2}\int_0^T f_{xx}(t, W_t)dt. \tag{8.8}$$

PROOF: See SCF2 for a sketch. ■

8.4 The Itô Formula for Functions of an Itô Process

Definition 8.4 (SCF2 Definition 4.4.3 - Itô process)**.** Let $W_t, t \geqq 0$, be a Brownian motion, and let $\mathfrak{F}_t, t \geqq 0$, be an associated filtration.

An **Itô process** on $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ is a stochastic process

$$X_t = x + \int_0^t \Delta_u dW_u + \int_0^t \Theta_u du, \tag{8.9}$$

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 (8.9). Furthermore we say that **(A)** and **(B)** express (8.9) in differential notation, and that (8.9) expresses **(A)** and **(B)** as an **integral equation**. □

Remark:

(1). The phrase ".... which we also equivalently express as ...." is to be taken literally: We do not mathematically distinguish between the integral equation (**??**) and the associated set of stochastic differential **(A)** plus initial condition **(B)**. They mean exactly the same thing.

(2). We bury into this footnote [20] a technical remark taken literally from SCF2. □

Lemma 8.1 (SCF2 Lemma 4.4.4)**.** *The quadratic variation of the Itô process* (8.9) *is*

$$[X, X]_t = \int_0^t \Delta_u^2 du. \tag{8.10}$$

PROOF: See SCF2 for a sketch. ■

Definition 8.5 (SCF2 Definition 4.4.5)**.** Given are an Itô process

$$X_t = X_0 + \int_0^t \Delta_u dW_u + \int_0^t \Theta_u du,$$

[20]**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 (8.9) are defined and the Itô integral is a martingale. We shall always make such integrability assumptions, but we do not always explicitly state them.

on $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ and an adapted process $\Gamma_t, t \geqq 0$. We define [21]

$$\int_0^t \Gamma_u dX_u := \int_0^t \Gamma_u \Delta_u dW_u + \int_0^t \Gamma_u \Theta_u du. \ \square \tag{8.11}$$

Theorem 8.5 (Itô–Doeblin formula for Brownian motion) on p.135. 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 8.6 (SCF2 Theorem 4.4.6 - Itô–Doeblin formula for an Itô process)**.** *Let $X_t, t \geqq 0$ be an Itô process as described in Definition 8.4 on p.136, 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 $T \geqq 0$,*

$$\begin{aligned}
f(T, X_T) &= f(0, X_0) + \int_0^T f_t(t, X_t)dt + \int_0^T f_x(t, X_t)dX_t \\
&\quad + \frac{1}{2}\int_0^T f_{xx}(t, X_t)dX[t,t] \\
&= f(0, X_0) + \int_0^T f_t(t, X_t)dt + \int_0^T f_x(t, X_t)\Delta_t dW_t \\
&\quad + \int_0^T f_x(t, X_t)\Theta_t dt + \frac{1}{2}\int_0^T f_{xx}(t, X_t)\Delta_t^2 dt.
\end{aligned} \tag{8.12}$$

PROOF: See SCF2. ■

Remark 8.4.

Itô formula for an Itô process in differential notation:

$$df(t, X_t) = f_t(t, X_t)\, dt + f_x(t, X_t)\, dX_t + \frac{1}{2} f_{xx}(t, X_t)\, dX_t dX_t. \tag{8.13}$$

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.

$$\begin{aligned}
dX_t\, dX_t &= (\Delta_t dW_t + \Theta_t dt)\,(\Delta_t dW_t + \Theta_t dt) \\
&= Delta_t^2 dW_t dW_t + 2\Delta_t\Theta_t dW_t dt + \Theta_t^2 dt dt = \Delta_t^2 dt
\end{aligned}$$

We make these substitutions in (8.13) and group the dt terms:

[21] 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 (8.11) are defined.

(8.14)
$$\begin{aligned} df\big(t, X_t\big) &= f_x\big(t, X_t\big)\Delta_t\, dW_t \\ &+ \Big(f_t\big(t, X_t\big) + f_x\big(t, X_t\big)\Theta_t + \frac{1}{2} f_{xx}\big(t, X_t\big)\Delta_t^2\Big)\, dt. \ \square \end{aligned}$$

Example 8.1 (Generalized Geometric Brownian Motion)**.** Definition 7.18 on p.131 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

(8.15)
$$X_t = \exp\left[\int_0^t \sigma_u dW_u + \int_0^t \left(\alpha_u - \frac{1}{2}\sigma_u^2\right) du\right],$$

where α_t and σ_t are adapted processes. Then X is an Itô process with differential

(8.16)
$$dX_t = \sigma_t dW_t + \left(\alpha_t - \frac{1}{2}\sigma_t^2\right) dt, \quad X_0 = 0.$$

From the multiplication table we obtain its squared differential

(8.17)
$$dX_t dX_t = \sigma_t^2 dW_t dW_t = \sigma_t^2 dt.$$

Let $S_0 \in]0, \infty$ (i.e., S_0 is deterministic), and $f(x) := S_0 e^x$. Since f does 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

(8.18)
$$S_t := S_0 e^{X_t} = S_0 \exp\left[\int_0^t \sigma_s dW_s + \int_0^t \left(\alpha_s - \frac{1}{2}\sigma_s^2\right) ds\right],$$

Since $S_t = f(X_t)$ an application of the Itô formula yields

(8.19)
$$\begin{aligned} dS_t &= df\big(X_t\big) = f'\big(X_t\big)\, dX_t + \frac{1}{2} f''\big(X_t\big)\, dX_t dX_t \\ &= S_0 e^{X_t}\, dX_t + \frac{1}{2} S_0 e^{X_t}\, dX_t dX_t = S_t\, dX_t + \frac{1}{2} S_t dX_t dX_t \end{aligned}$$

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 the random process itself and also the differential dW_t of a Brownian motion. $\square$

Remark 8.5. It follows from formulas (8.16) and (8.17) that

$$\begin{aligned} S_t dX_t &\overset{(8.16)}{=} \sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt \,-\, \frac{1}{2}\sigma_t^2 S_t\, dt \\ &\overset{(8.17)}{=} \sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt \,-\, \frac{1}{2} S_t\, dX_t dX_t, \end{aligned}$$

We plug this expression for $S_t dX_t$ into the last equation of (8.19) and obtain

$$\begin{aligned} dS_t \;&=\; \left(\sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt \,-\, \frac{1}{2} S_t\, dX_t dX_t\right) \,+\, \frac{1}{2}\, S_t dX_t dX_t \\ &= \sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt. \end{aligned}$$

$$\begin{aligned} dS_t \;&=\; \left(\sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt \,-\, \frac{1}{2} S_t\, dX_t dX_t\right) \,+\, \frac{1}{2}\, S_t dX_t dX_t \\ &= \sigma_t S_t\, dW_t \,+\, \alpha_t S_t\, dt. \end{aligned}$$

That last formula

(8.20) $$dS_t \;= \alpha_t S_t\, dt \,+\, \sigma_t S_t\, dW_t$$

describes the dynamics of the process S_t. Interpreted as the price of a stock it expresses that the 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 an excellent wasy to model the price evolution of a risky asset 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 (8.18) becomes ordinary GBM

(8.21) $$S_t \;= S_0\, \exp\left\{\sigma W_t \,+\, \left(\alpha - \frac{1}{2}\sigma^2\right)\, t\right\}.$$

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 7.19 on p.131. 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 8.7 (SCF2 Theorem 4.4.9 - Itô integral of a deterministic integrand)**.** *Let $W_s, s \geqq 0$, be a Brownian motion and let Δ_s be a nonrandom function of time. Define $I_t = \int_0^t \Delta_s dW_s$. For each $t \geqq 0$, the random variable I_t is normally distributed with expected value zero and variance $\int_0^t \Delta_s^2 ds$.*

PROOF: See SCF2. ■

Example 8.2 (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 $W = W_t$. Assume that the interest rate $R = R_t(\omega)$ in a market economy is modeled by the SDE

$$dR_t = \big(\alpha - \beta R_t\big)\, dt + \sigma\, dW_t, \tag{8.22}$$

$\alpha, \beta, \sigma \in]0, \infty[$ are positive and deterministic constants. We call this the **Vasicek model**.
The solution to this SDE is

$$R_t = e^{-\beta t} R_0 + \frac{\alpha}{\beta}(1 - e^{-\beta t}) + \sigma e^{-\beta t} \int_0^t e^{\beta s} dW_s. \tag{8.23}$$

For a proof see SCF2. □

Example 8.3 (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

$$dR_t = \big(\alpha - \beta R_t\big) dt + \sigma \sqrt{R_t}\, dW_t, \tag{8.24}$$

$\alpha, \beta, \sigma \in]0, \infty[$ are positive and deterministic constants. We call this the **Cox–Ingersoll–Ross model**.

$$E[R_t] = e^{-\beta t} R_0 + \frac{\alpha}{\beta}(1 - e^{-\beta t}). \tag{8.25}$$

This is the same expectation as in the Vasicek model.

$$Var[R_t] = \frac{\sigma^2}{\beta} R_0 (e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha \sigma^2}{2\beta^2}(1 - 2e^{-\beta t} + e^{-2\beta t}). \tag{8.26}$$

In particular,

$$\lim_{t\to\infty} Var[R_t] = \frac{\alpha\sigma^2}{2\beta^2}. \ \square$$

□

For a proof see SCF2. □

8.5 Exercises for Ch.8

Exercise 8.1. 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 8.2 (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 8.3. 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 Black–Scholes Model Part I: The PDE

Introduction 9.1. This chapter is based on the finance application oriented aspects of GBM (geometric Brownian motion) that were briefly mentioned in Remark 8.5 about generalized GBM (p.139) and replicating portfolios for a continent claim given in Chapter 6.2 (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 (zero coupon bond or money market account). We will 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. □

9.1 Formulation of the Black–Scholes Model

Notations 9.1. 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 [3] Björk, Thomas: Arbitrage Theory in Continuous Time which I used in Chapter 6 (Financial Models - Part 1) 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. In SCF2 it is always 1.
$\vec{H}_t$	N/A	portfolio (# of shares) vector for all assets.
$x_t = H_t^{(0)}$	N/A	# of shares (= dollar value) of the riskless asset.
$y_t = H_t^{(1)}$	Δ_t	# of shares of the risky asset.
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 rather than $H_t^{(0)}$ for the number of shares (dollars) in the bank account (B = **B**ank account) and H_t^S (S = **S**tock) rather than $H_t^{(1)}$ for the number of shares in the risky asset.

The portfolio value process thus will be occasionally written as follows.

$$V_t \;=\; H_t^B \;+\; H_t^S S_t. \;\Box$$

Definition 9.1 (Black–Scholes Market Model)**.**
The **Black–Scholes market model** consists of a time $T > 0$, a risk free asset (bank account) with price process $B = B_t, 0 \leqq t \leqq T$, a risky asset with price process $S = S_t, 0 \leqq t \leqq T$, 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.

(9.1) $$dB_t = rB_t\,dt;\; B_0 = 1;$$

(9.2) $$dS_t = \alpha S_t\,dt + \sigma S_t\,dW_t;\quad S_0 \in [0,\infty[;\, \alpha,\sigma \in]0,\infty[\,,$$

(9.3) $$\mathcal{X} = \Phi(S_T) \quad \text{(simple contingent claim)},$$

- $c : [0,T] \times [0,\infty[\quad (t,x) \mapsto c(t,x)$ twice continuously differentiable such that

(9.4) $$\Pi(t;\mathcal{X}) = c(t,S_t) \quad \text{(price process of } \mathcal{X})$$

- The market is efficient: No arbitrage portfolios. □

Remark 9.1.
(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 rate r per unit time.
(2) Formula (9.2) states that S_t is GBM with constant, instantaneous mean rate of return α and constant volatility σ. See Remark 8.5 on p.139. There are more general models (Definition 12.1 on p.170) in which the constants α and σ are replaced by measurable functions $\alpha(t,x), \sigma(t,x)$ of time and the price of the risky asset:

$$dS_t = \alpha(t,S_t)S_t\,dt + \sigma(t,S_t)S_t\,dW_t;\quad S_0 \in [0,\infty[.$$

(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) <u>c</u>all is used when deriving the PDE for that function is derived. Note that this function must satisfy the terminal condition

(9.5) $$c(T,S_T) = \Pi(T;\mathcal{X}) = \Phi(S_T).$$

(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.
(4) You should recall from Assumption 6.1 on p.94 that we have always assumed that the market is free of arbitrage, in addition to some other assumptions such as complete liquidity, no transaction costs and no bid–ask spread. □

9.2 Discounted Values of Option Price and Hedging Portfolio

Proposition 9.1. *The budget equation for a self–financing portfolio strategy in a Black–Scholes market is*

(9.6) $$dX_t = \Delta_t dS_t + r\big(X_t - \Delta_t S_t\big)\,dt.$$

Further we have the following equation for the portfolio value dynamics.

$$
\begin{aligned}
dX_t &= rX_t dt + \Delta_t(\alpha - r)S_t dt + \Delta_t \sigma S_t\, dW_t. \\
&= r(X_t - \Delta_t S_t)dt + r\Delta_t S_t dt + (\alpha - r)\Delta_t S_t dt + \Delta_t \sigma S_t\, dW_t.
\end{aligned}
\tag{9.7}
$$

PROOF: See SCF2, Chapter 4.5.1 (Evolution of Portfolio Value). ■

Remark 9.2. Formula (9.7) signifies that a portfolio value change dX_t is composed of

- **a.** An average underlying rate of return r on the bank account value $X_t - \Delta_t S_t$,
- **b.** An average underlying rate of return $r + (\alpha - r) = \alpha$ on the risky asset investment in height of $\Delta_t S_t$. Since people will not take a greater risk investing in a stock than putting money in the bank we should expect that $\alpha \geqq r$, thus $(\alpha - r)$ is a risk premium for investing in the stock.
- **c.** A volatility term $\Delta_t \sigma S_t dW_t$ which is proportional to the size $\Delta_t \sigma S_t$ of the stock investment. □

Remark 9.3. Budget equations in the continuous case will be formulated by means of "stochastic differentials" once we have the necessary tools from stochastic calculus, and they will look quite different from formula (6.9). See Remark 6.6 on p.93 which follows the definition of a continuous trading budget equation. □

Proposition 9.2. *The discounted portfolio value $d\big(e^{-rt}X_t\big)$ satisfies*

$$
\begin{aligned}
d\big(e^{-rt}X_t\big) &= \Delta_t(\alpha - r)\, e^{-rt}S_t\, dt + \Delta_t \sigma e^{-rt} S_t\, dW_t \\
&= \Delta_t\, d\big(e^{-rt}S_t\big).
\end{aligned}
\tag{9.8}
$$

PROOF: See SCF2, Chapter 4.5.1 (Evolution of Portfolio Value). ■

Remark 9.4. Formula (9.8) shows that change in the discounted portfolio value has notthing 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 9.1 of the Black–Scholes Market Model such that $\Pi(t;\mathcal{X}) = c(t, S_t)$.

Proposition 9.3. *The price dynamics of the contingent claim are*

$$
dc\big(t, S_t\big) = \left[c_t\big(t,S_t\big) + \alpha S_t\, c_x\big(t,S_t\big) + \frac{1}{2}\sigma^2 S_t^2\, c_{xx}\big(t,S_t\big)\right] dt + \sigma S_t\, c_x\big(t,S_t\big)\, dW_t.
\tag{9.9}
$$

Those of the discounted option price $e^{-rt}c\big(t,S_t\big)$ are

$$
\begin{aligned}
d\big(e^{-rt}c(S_t)\big) &= e^{-rt}\Big[-rc\big(t,S_t\big) + c_t\big(t,S_t\big) + \alpha S_t\, c_x\big(t,S_t\big) + \frac{1}{2}\sigma^2 S_t^2\, c_{xx}\big(t,S_t\big)\Big]\, dt \\
&\quad + e^{-rt}\sigma S_t\, c_x\big(t,S_t\big)\, dW_t.
\end{aligned}
\tag{9.10}
$$

PROOF: See SCF2, Chapter 4.5.2 (Evolution of Option Value). ■

9.3 The Pricing Principle in the Black–Scholes Market

According to the pricing principle (Theorem 6.1 on p.95) an arbitrage free price $c(t, S_t)$ of the contingent claim requires that a replicating portfolio with value process X_t satisfies

$$c(t, S_t) \;=\; X_t, \text{ for all trading times } t.$$

This is equivalent to $e^{-rt}X_t = e^{-rt}c(t, S_t)$ for all t. In terms of differentials:

$$\text{(9.11)}\qquad \begin{aligned} d\big(e^{-rt}X_t\big) &= d\big(e^{-rt}c(t, S_t)\big) \text{ for all } t,\\ X_0 &= c\big(0, S_0\big) \end{aligned}$$

We thus may equate the right hand sides of the formulas (9.8) and (9.10). We obtain after canceling the factor e^{-rt} everywhere and omitting the argument (t, S_t) of the function c and its derivatives c_t, c_x, c_{xx},

$$\text{(9.12)}\qquad \begin{aligned} &\Delta_t\sigma S_t\, dW_t + \Delta_t(\alpha - r)S_t\, dt\\ =\; &\sigma S_t\, c_x\, dW_t + \Big[-rc \,+\, c_t \,+\, \alpha S_t\, c_x \,+\, \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx}\Big]\, dt. \end{aligned}$$

Since evolution with respect to dt is fundamentally different of that with respect to dWt it is allowed to separately equate first the dW_t terms and then the dt terms of formula (9.12). We first equate the dW_t terms and obtain after canceling $\sigma e^{-rt}S_t$ the

delta–hedging rule:

$$\text{(9.13)}\qquad \Delta_t \;= c_x\big(t, S_t\big) \;\text{ for all } t \in [0, T[.$$

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 9.2 (Delta (Greek))**.** For a contingent claim $\mathcal{X}$ in the Black–Scholes market with the function $(t, x) \mapsto c(t, x)$ yielding the price process $\Pi_t(\mathcal{X}) = c(t, S_t)$ we call the partial derivative of $c(t, x)$ with respect to stock price x,

$$\text{(9.14)}\qquad \text{delta} \;:= \frac{\partial c}{\partial x},$$

the **delta** of the claim. Delta is one of the so called **greeks** of the claim. □

We next equate the dt terms of formula (9.12). We just proved that $\Delta_t = c_x(t, S_t)$, thus

$$c_x(\alpha - r)S_t \;=\; -rc \,+\, c_t \,+\, \alpha S_t\, c_x \,+\, \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx}.$$

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

$$rc \ = \ c_t \ + \ rc_x S_t \ + \ \frac{1}{2}\,\sigma^2 S_t^2\, c_{xx}. \tag{9.15}$$

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 (9.5).

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

$$c_t(t, x) + rx\, c_x\big(t, x\big) \ + \ \frac{1}{2}\,\sigma^2 x^2 c_{xx}(t, x) \ = r\,c\big(t, x\big), \ x \geqq 0, \tag{9.16}$$

subject to the terminal condition

$$c(T, x) \ = \Phi(S_T). \tag{9.17}$$

We use the equations $V_t^H = X_t = c(t, S_t$ and $\Delta_t = c_x(t, S_t)$ to express the hedging portfolio for the claim $\mathcal{X}$ purely in terms of the claim pricing function.

$$\vec{H}_t \ = \big(\, c(t, S_t) \ - \ c_x(t, S_t),\, c_x(t, S_t)\,\big) \tag{9.18}$$

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 risky asset.

Remark 9.5. 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 $0 \leqq t \leqq T$ satisfies the PDE just for $0 \leqq t \leqq T$ and $x \geqq 0$ then we are fine snce continuity of $t \mapsto c(t, S_t)$ and $t \mapsto X_t$ for $0 \leqq t \leqq T$ implies that the hedge equation $X_t = c(t, S_t)$ extends from $0 \leqq t < T$ to $t = T$, and the boundary condition $c(T, x) = \Phi(x)$ yields $X_T = \Phi(X_T)$.

To summarize, it is enough to show that the Black–Scholes PDE holds for all $x \geqq 0$ and $t \in [0, T[$ □

9.4 The Black–Scholes PDE for a European Call

The Black–Scholes PDE (9.16) on p.146 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 9.6. 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 $x = 0$.

$$c(t,0) \;=\; 0 \;\text{ for all } t \in [0,T]. \tag{9.19}$$

This is true for the following reason. Formula (9.16) 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., $\text{const} = y(0) = c(0,0)$. Thus

$$c(t,0) \;=\; c(0,0)\, e^{rt} \quad \text{for all } 0 \leqq t \leqq T. \tag{\textbf{A}}$$

$K \geqq 0$, 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 9.6 above but also the growth condition

$$\lim_{x\to\infty} \big(c(t,x) \;-\; (x - e^{r(T-t)}K)\big) \;= 0 \;\text{ for all } t \in [0,T]. \tag{9.20}$$

Since $x - e^{r(T-t)}K$ 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 6.3 on p.88 that this was the contract function for a forward contract with strike price K. $\square$

Without proof for now:

Theorem 9.1. *The solution to the Black–Scholes partial differential equation* (9.16) *with terminal condition* (9.17)*, zero stock price condition* (9.19)*, and growth condition* (9.20) *is*

$$c(t,x) \;= xN\big(d_+(T-t,x)\big) \;-\; Ke^{-r(T-t)}N\big(d_-(T-t,x)\big), \quad 0 \leqq t < T,\; x > 0, \tag{9.21}$$

where

$$d_\pm(\tau,x) \;= \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} \;+\; \left(r \pm \frac{\sigma^2}{2}\right)\tau\right], \tag{9.22}$$

and N is the cumulative standard normal distribution

$$N(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{z^2}{2}}\, dz = \frac{1}{\sqrt{2\pi}} \int_{-y}^{\infty} e^{-\frac{z^2}{2}}\, dz. \tag{9.23}$$

PROOF: Will be given later, in Theorem **??** on p.**??**. ■

Remark 9.7. 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 (9.21) becomes

$$\text{BSM}(\tau, x; K, r, \sigma) = xN\big(d_+(\tau, x)\big) - Ke^{-r\tau}N\big(d_-(\tau, x)\big), \tag{9.24}$$

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 9.8. 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 □

9.5 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 $r \geqq 0$ 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 $\frac{\partial P}{\partial r}$.

Definition 9.3 (Björk Def.9.4: Greeks)**.**
The following derivatives are part of what is known as the **Greeks** of the function F.

$$\Delta = \frac{\partial F}{\partial x} \qquad \textbf{delta} \tag{9.25}$$

$$\Gamma = \frac{\partial^2 F}{\partial x^2} \qquad \textbf{gamma} \tag{9.26}$$

$$\rho = \frac{\partial F}{\partial r} \qquad \textbf{rho} \tag{9.27}$$

$$\Theta = \frac{\partial F}{\partial t} \qquad \textbf{theta} \tag{9.28}$$

$$\nu = \frac{\partial F}{\partial \sigma} \qquad \textbf{vega} \ \square \tag{9.29}$$

Remark 9.9. 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 contiguous 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. $\square$

Having stated that the Greeks are defined for all simple claims, the following formulas are specific for the pricing function $c(t,x)$ of a European call.

Proposition 9.4. *The following is true for the Greeks of a European call.*

$$\textbf{delta} = c_x(t,x) \ = \ N\big(d_+(T-t,x)\big), \tag{9.30}$$

$$\textbf{gamma} = c_{xx}(t,x) \ = \ \frac{1}{\sigma x\sqrt{T-t}}\, N'\big(d_+(T-t,x)\big), \tag{9.31}$$

$$\textbf{theta} = c_t(t,x) \ = \ -rK\, e^{-r(T-t)} N\big(d_-(T-t,x)\big) \ - \ \frac{\sigma x}{2\sqrt{T-t}}\, N'\big(d_+(T-t,x)\big). \tag{9.32}$$

Because both the cumulative distribution function $N(x)$ density $N'(x)$ of a standard normal random variable 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 9.5. *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 bank account, and H_t^S denotes the*

number of shares held in the risky asset (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 $S_t = x$.

$$V_t^H = F(t,x), \tag{9.33}$$
$$H_t^B = F(t,x) - x \cdot F_x(t,x), \tag{9.34}$$
$$H_t^S = F_x(t,x). \tag{9.35}$$

PROOF: Formula (9.33) 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 (9.35) is the delta hedging rule which states the number of shares in the underlying (risky) asset is the derivative of the pricing function F with respect to stock price, evaluated at $x = S_t$.

Formula (9.34) 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.

$$H_t^B = V_t^H - S_t \cdot V_t^S, \qquad \text{i.e.,} \quad H_t^B = F(t,x) - x \cdot F_x(t,x)). \ \blacksquare$$

Remark 9.10. 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 9.4. 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 $S_t = x$, $H_t^B = c(t,x) - xc_x(t,x) = Ke^{-r(T-t)}N(d_-)$ in the money market account. See formula (9.30). □

We defined in Definition 6.3 on p.88 a forward contract as a simple claim with contract function $\Phi(x) = x - K$.

Proposition 9.6. *We write $f(t,x)$ for the pricing function of the forward contract. It is computed as follows.*

$$f(t,x) = x - e^{-r(T-t)}K. \tag{9.36}$$

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.,

$$H_0^B = -e^{-rT}K, \qquad H_0^S = 1.$$

This is 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 the portfolio value is

$$V_t^H = -e^{rt} \cdot e^{-rT} K + S_t = S_t - e^{-r(T-t)} K.$$

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

$$f(t, S_t) = V_t^H = S_t - e^{-r(T-t)} K \quad \text{for all } 0 \leqq t \leqq T. \quad \blacksquare$$

Associated with a forward contract is its forward price For_t, the fair price for this contract if it was to be re-evaluated at a later time $0 \leqq t \leqq T$.

Definition 9.4 (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 9.11.

A. Given our assumption of a constant interest rate, For_t satisfies the equation

$$S_t - e^{-r(T-t)} \text{For}_t = 0. \quad \square \tag{9.37}$$

B. Note that $\text{For}_0 = K$. 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 (9.37) for For_t and obtain

$$\text{For}_t = e^{r(T-t)} S_t. \tag{9.38}$$

D. Note that, for a given time t,

the forward price For_t is NOT the price (or value) $f(t, S_t)$ of a forward contract. $\square$

We recall from Definition 6.3 on p.88 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 pricing process of a European put option.

We relate puts and calls by mean of the following simple identity.

Lemma 9.1. *For any real number α,*

$$\alpha = \alpha^+ - (-\alpha)^+. \tag{9.39}$$

PROOF:

$$\textbf{Case 1}: \ \alpha \geqq 0 \ \Rightarrow \ \alpha^+ = \alpha, \ (-\alpha)^+ = 0 \ \Rightarrow \ \alpha^+ - (-\alpha)^+ = \alpha - 0 = \alpha.$$
$$\textbf{Case 2}: \ \alpha < 0 \ \Rightarrow \ \alpha^+ = 0, \ (-\alpha)^+ = -\alpha \ \Rightarrow \ \alpha^+ - (-\alpha)^+ = 0 - (-\alpha) = \alpha. \ \blacksquare$$

Corollary 9.1.

$$f\big(T, S_T\big) = S_T - K = (S_T - K)^+ - (K - S_T)^+ = c\big(T, S_T\big) - p\big(T, S_T\big).$$

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 9.1. ■

Proposition 9.7 (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**:*

$$f(t, x) = c(t, x) - p(t, x), \quad \textit{for all } x \geqq 0, \ 0 \leqq t \leqq T. \tag{9.40}$$

PROOF: We apply the pricing principle to the formula $p\big(T, S_T\big) = c\big(T, S_T\big) - f\big(T, S_T\big)$ which is immediate from Corollary 9.1 and obtain

$$p(t, x) = c(t, x) - f(t, x), \quad \text{for all } x \geqq 0, \ 0 \leqq t \leqq T. \ \blacksquare$$

Proposition 9.8. *The pricing function $p(t, x)$ of a European put with strike price K satisfies*

$$\begin{aligned} p(t, x) &= x\big(N\big(d_+(T-t, x)\big) - 1\big) - Ke^{-r(T-t)}\big(N\big(d_-(T-t, x)\big) - 1\big) \\ &= Ke^{-r(T-t)}N\big(-d_-(T-t, x)\big) - x\big(N\big(-d_+(T-t, x)\big)\big), \end{aligned} \tag{9.41}$$

PROOF: This follows from put–call parity, the explicit formulas (9.21) and (9.22) for $c(t, x)$ (see p.147), and formula (9.36) on p.150 for $f(t, x)$. ■

9.6 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 9.12.

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., $V_t^{\vec{h}}(\omega) = -c\big(t, S_t(\omega)\big)$. $\square$

9.7 Exercises for Ch.9

None at this time!

10 Multidimensional Stochastic Calculus

We generalize in this chapter the results of Chapter 8 (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.

10.1 Multidimensional Brownian Motion

Definition 10.1 (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 $\sigma\big(W_t^{(i)} : t \geqq 0\big)$ and $\sigma\big(W_t^{(j)} : t \geqq 0\big)$ 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 $t \geqq 0$.

(4) Future increments are independent of the past: If $t \geqq 0 and h > 0$ then the vector $\vec{W}_{t+h} - \vec{W}_t$ is independent of $\mathfrak{F}_t$. $\square$

Remark 10.1. 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 10.2 (Cross variation)**.** $\star$

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 $[0, T]$. We call the random variable

$$C_\Pi[X, Y]_T \;:=\; \sum_{k=0}^{n-1} (X_{t_{k+1}} \;-\; X_{t_k})\,(Y_{t_{k+1}} \;-\; Y_{t_k})$$

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 call the process $[X, Y]_t$ the **cross variation** of X and Y. $\square$

Remark 10.2. Note that if $X = Y$ then the process $[X, X]_t$ is the quadratic variation of X. □

Theorem 10.1. *Let $\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)} \dots, W_t^{(d)}\big)$ 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 $1 \leqq i < j \leqq d$. Then*

$$\Big[W^{(i)}, W^{(j)}\Big]_t \;=\; 0.$$

PROOF: See SCF2 ch.4.6.1. ■

Theorem 10.2. *Let $\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)} \dots, W_t^{(d)}\big)$ 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 $1 \leqq i, j \leqq d$ and $i \neq j$. Then*

$$dW^{(i)}\, dW^{(j)} \;=\; 0.$$

PROOF: This can be shown with help of Theorem 10.1 on p.155. See SCF2 ch.4.6. for details. ■

10.2 The Multidimensional Itô Formula

One can generalize The Itô formula which computes the differential $f(t, X_t)$ to more than one Itô process X_t, each of which is driven by a d–dimensional Brownian motion in the sense of the next definition.

Definition 10.3. ⋆

Let $\vec{W}_t = \big(W_t^{(1)}, W_t^{(2)} \dots, W_t^{(d)}\big)$ 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 X has dynamics

$$\begin{aligned} dX_t &= \Theta_t\, dt \,+\, \sum_{j=1}^{d} \sigma_j(t)\, dW_t^{(j)} \;=\; \Theta_t\, dt \,+\, \sigma_1(t)\, dW_t^{(1)} \,+ \dots\, \sigma_d(t)\, dW_t^{(d)}, \\ X_0 &= x, \end{aligned} \tag{10.1}$$

for suitable adapted and sufficiently integrable processes Θ_t and $\vec{\sigma}(t) = \big(\sigma_1(t) \dots, \sigma_n(t)\big)$. In integrated form (10.1) is equivalent to

$$X_t \;= x \,+\, \int_0^t \Theta_u\, du \,+\, \sum_{j=1}^{d} \int_0^t \sigma_j(u)\, dW_u^{(j)}. \;\square \tag{10.2}$$

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.

Notations 10.1. Let $n \in \mathbb{N}$. If $\vec{\Gamma}_t = \big(\Gamma_t^{(1)}, \dots, \Gamma_t^{(n)}\big)$ and $\vec{A}_t = \big(A_t^{(1)}, \dots, A_t^{(n)}\big)$ are vector valued stochastic processes for which the expressions $\int\limits_0^t \Gamma_u^{(j)} dA_u^{(j)}$ exist then we define

$$\begin{aligned} \vec{\Gamma}_t \bullet d\vec{A}_t \;&:=\; \sum_{j=1}^{n} \Gamma_t^{(j)}\, dA_t^{(j)}, \\ \int_0^t \vec{\Gamma}_u \bullet d\vec{A}_u \;&:=\; \sum_{j=1}^{n} \int_0^t \Gamma_u^{(j)}\, dA_u^{(j)}, \;\square \end{aligned} \tag{10.3}$$

With this notation we can rewrite (10.1) and (10.2) as follows.

$$\begin{aligned} dX_t &= \Theta_t\,dt + \vec{\sigma}(t)\,d\vec{W}_t; \quad X_0 = x, \\ X_t &= x + \int_0^t \Theta_u\,du + \int_0^t \vec{\sigma}(u)\,d\vec{W}_u. \ \square \end{aligned}$$

Remark 10.3. It should be mentioned that Itô's Lemma not only generalizes to d–dimensional Brownian motions for $d > 2$ but also to functions

$$f(t, \vec{x}) = f(t, x_1, x_2, \dots, x_n)$$

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 do that but rather follow SCF2 and limit ourselves to two Itô processes X and Y, which are driven by a twodimensional Brownian motion. $\square$

Notations 10.2. From now on we assume that $\vec{W}_t = \left(W_t^{(1)}, W_t^{(2)}\right)$ is a twodimensional Brownian motion and that X_t and Y_t are the following Itô processes, driven by $\vec{W}_t$.

$$\begin{aligned} dX_t &= \Theta_1(t)\,dt + \sigma_{11}(t)\,dW_t^{(1)} + \sigma_{12}(t)\,dW_t^{(2)}, \\ dY_t &= \Theta_2(t)\,dt + \sigma_{21}(t)\,dW_t^{(1)} + \sigma_{22}(t)\,dW_t^{(2)}. \end{aligned} \tag{10.4}$$

The integrands $\Theta_i(u)$ and $\sigma_{ij}(u)$ are adapted processes. We integrate and get

$$\begin{aligned} X_t &= x_0 + \int_0^t \Theta_1(u)\,du + \int_0^t \sigma_{11}(u)\,dW_u^{(1)} + \int_0^t \sigma_{12}(u)\,dW_u^{(2)}, \\ Y_t &= y_0 + \int_0^t \Theta_2(u)\,du + \int_0^t \sigma_{21}(u)\,dW_u^{(1)} + \int_0^t \sigma_{22}(u)\,dW_u^{(2)}. \end{aligned} \tag{10.5}$$

Theorem 10.3. *The multiplication rules for the multidimensional Itô calculus are*

$$\begin{aligned} dt\,dt &= 0, & dt\,dW_t^{(i)} &= 0, \\ dW_t^{(i)}\,dW_t^{(i)} &= t, & dW_t^{(i)}\,dW_t^{(j)} &= 0 \text{ for } i \neq j. \end{aligned}$$

PROOF: This follows from the onedimensional case (see Remark 7.15 on p.129), together with Theorem 10.1 on p.155. ■

Remark 10.4. 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 (**??**) then

$$\begin{aligned} dX_t\,dX_t\; &= \left[d\big(\Theta_1(t)\,dt\,+\,\sigma_{11}(t)\,dW_t^{(1)}\,+\,\sigma_{12}(t)\,dW_t^{(2)}\big)\right]^2 \\ &= \Theta_1(t)^2 dt\,dt\,+\,\Theta_1(t)dt\,\sigma_{11}(t)\,dW_t^{(1)}\,+\,\Theta_1(t)dt\,\sigma_{12}(t)\,dW_t^{(2)} \\ &\quad +\cdots+\,\sigma_{12}(t)^2\,dW_t^{(2)}\,dW_t^{(2)} \end{aligned}$$

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

$$dX_t\,dX_t\; = \sigma_{11}(t)^2\,dt\,+\,\sigma_{12}(t)^2\,dt.$$

Here is one more example.

$$\begin{aligned} dX_t\,dY_t\; &= \Theta_1(t)\Theta_2(t)dt\,dt\,+\,\Theta_1(t)dt\,\sigma_{21}(t)\,dW_t^{(1)}\,+\,\Theta_1(t)dt\,\sigma_{22}(t)\,dW_t^{(2)} \\ &\quad +\cdots+\,\sigma_{12}(t)\sigma_{22}(t)\,dW_t^{(2)}\,dW_t^{(2)} \end{aligned}$$

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 twodimensional Brownian motion. This is SCF2 Theorem 4.6.2

Theorem 10.4 (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*

$$\begin{aligned} df\big(t,X_t,Y_t\big) \;&=\; f_t\big(t,X_t,Y_t\big)\,dt\,+\,f_x\big(t,X_t,Y_t\big)\,dX_t\,+\,f_y\big(t,X_t,Y_t\big)\,dY_t \\ &\quad +\,\frac{1}{2}\,f_{xx}\big(t,X_t,Y_t\big)\,dX_t dX_t\,+\,f_{xy}\big(t,X_t,Y_t\big)\,dX_t dY_t \\ &\quad +\,\frac{1}{2}\,f_{yy}\big(t,X_t,Y_t\big)\,dY_t dY_t. \end{aligned} \tag{10.6}$$

PROOF: Omitted. ■

Remark 10.5. Here is the Itô formula with integrals rather than differentials.

$$
\begin{aligned}
& f(t,X_t,Y_t) - f(0,X_0,Y_0) \\
&= \int_0^t \left[\sigma_{11}(u)\, f_x(u,X_u,Y_u) + \sigma_{21}(u)\, f_y(u,X_u,Y_u)\right] dW_1(u) \\
&+ \int_0^t \left[\sigma_{12}(u)\, f_x(u,X_u,Y_u) + \sigma_{22}(u)\, f_y(u,X_u,Y_u)\right] dW_2(u) \\
&+ \int_0^t \Big[f_t(u,X_u,Y_u) + \Theta_1(u)\, f_x(u,X_u,Y_u) + \Theta_2(u)\, f_y(u,X_u,Y_u) \\
&\qquad + \frac{1}{2}\left(\sigma_{11}^2(u) + \sigma_{12}^2(u)\right) f_{xx}(u,X_u,Y_u) \\
&\qquad + \left(\sigma_{11}(u)\sigma_{21}(u) + \sigma_{12}(u)\sigma_{22}(u)\right) f_{xy}(u,X_u,Y_u) \\
&\qquad + \frac{1}{2}\left(\sigma_{21}^2(u) + \sigma_{22}^2(u)\right) f_{yy}(u,X_u,Y_u) \Big] du
\end{aligned}
\tag{10.7}
$$

Even though this version of the Itô formula is mathematically more precise than (10.6) it is harder to remember and more cumbersome to use. Here is the other extreme, with all arguments of the tunction $f(t,x,y)$ and its partial derivatives omitted.

$$
\begin{aligned}
df(t,X,Y) &= f_t\,dt + f_x\,dX + f_y\,dY \\
&\quad + \frac{1}{2} f_{xx}\,dX_t\,dX_t + f_{xy}\,dX_t\,dY_t + \frac{1}{2} f_{yy}\,dY_t\,dY_t. \quad \square
\end{aligned}
\tag{10.8}
$$

The following is arguably the most useful application of the multidimensional Itô formula.

Corollary 10.1 (Itô product rule)**.** *If X_t and Y_t are two Itô processes then*

$$
d(X_t\,Y_t) = X_t\,dY_t + Y_t\,dX_t + dX_t\,dY_t. \tag{10.9}
$$

PROOF: We apply formula (10.8) 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. ■

10.3 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 $W(j)_t$ is a (onedimensional) 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 onedimensional version. This is SCF2 Theorem 4.6.4

Theorem 10.5 (Lévy's characterization of onedimensional Brownian Motion)**.** *Let $M_t, t \geqq 0$, be a martingale relative to a filtration $\mathfrak{F}_t, t \geqq 0$. Assume that $M_0 = 0$, M_t has continuous paths, and $[M, M]_t = t$ for all $t \geqq 0$. Then M_t is a Brownian motion.*

PROOF: [⋆] An outline of the proof can be found in SCF2. We just 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. • Itô integrals $\int 0^t Z_u dM_u$, We have the same multiplication rules

$$dt\,dt \;=\; dt\,dM_t \;=\; dM_t\,dt \;=\; 0,\; dM_t\,dM_t \;=\; t.$$

The last rule results from $[M, M]_t = t$.
• Itô processes $X_t = X_0 + \int_0^t \Delta_u dM_u + \int_0^t \Theta_u du$ driven by a continouus martingale M_t, • The Itô formula for the differential $df(t, X_t)$ where X_t is an Itô process driven by a continouus martingale M_t,

(2) Fix $u \in \mathbb{R}$. The Itô formula is applied to the function

$$f(t, x) \;:=\; \exp\left[ux \,-\, \frac{1}{2}\,u^2 t\right].$$

with the result that

$$E\left[\,e^{uM_t}\,\right] \;=\; e^{\frac{1}{2}u^2 t}.$$

(3) Thus M_t has the same MGF as a Brownian motion W_t, i.e., it is Brownian motion. ■

(4) Finally the independence of the increments $M_{t+h} - M_t$ and $\mathfrak{F}_t$ must be shown for all $t, h \geqq 0$. ■

And this is the multidimensional version of Lévy's theorem (SCF2 Theorem 4.6.5).

Theorem 10.6 (Lévy's characterization of multidimensional Brownian Motion)**.** *Let the process $\vec{M}_t = \big(M_t^{(1)}, \dots, M_t^{(d)}\big)$ have continuous $\mathfrak{F}_t$–martingales $M_t^{(j)}$ as its coordinate processes. Assume further that $\vec{M}_0 = 0$, that the quadratic variations satisfy $[M^{(j)}, M^{(j)}]_t = t$ for all j, and that the cross variations $[M^{(i)}, M^{(j)}]_t$ are zero for $i \neq j$. then $\vec{M}_t$ is a d–dimensional Brownian motion. In particular, the 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 onedimensional case. Make again use of the fact that the Itô formula applies to Itô processes driven by continuous martingales and apply it, for fixed $\vec{u} = (u_1, \dots, u_d)$, to the function

$$f(t, x_1, \dots, x_d) \; := \; \exp\left[\sum_{j=1}^{d} u_j x_j \; - \; \frac{1}{2}\, t \sum_{j=1}^{d} u_j^2 \right]$$

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) but also that 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 10.1. ⋆

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 $-1 \leqq \rho \leqq 1$ 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 the same constants α_2 and σ_2.

(3) *$W_t^{(1)}$ and W_t^* have correlation ρ for all t. Thus those two processes are not independent. Thus $\big(W_t^{(1)}, W_t^*\big)$ is not a twodimensional Brownian motion.*

PROOF:

W_t^* is a continuous martingale as the sum ofcontinuous martingales and $W_0^* = 0$. Further,

$$\begin{aligned} dW_t^*\, dW_t^* &= \rho^2\, dW_t^{(1)}\, dW_t^{(1)} \; + \; 2\rho\sqrt{1-\rho^2}\, dW_t^{(1)}\, dW_t^{(2)} \; + \; (1-\rho^2)\, dW_t^{(2)}\, dW_t^{(2)} \\ &= \rho^2\, dt \; + \; (1-\rho^2)\, dt \; = \; dt. \end{aligned}$$

Thus $[W^*, W^*]_t = t$ and assertion **(1)** follows from Theorem 10.5 (Lévy's characterization of onedimensional Brownian Motion).

Assertion **(2)** is true by definition of W_t^* and since we just proved that this process is a Brownian motion.

To prove assertion **(3)** we compute $\mathrm{Cov}[W_t^1, W_t^*]$. Since $dW_t^{(1)} dW_t^{(2)} = 0$,

$$\begin{aligned} dW_t^{(1)}\, dW_t^* &= dW_t^{(1)} \left(\rho\, dW_t^{(1)} \; + \; \sqrt{1-\rho^2}\, dW_t^{(2)} \right) \\ &= \rho\, dW_t^{(1)}\, dW_t^{(1)} \; + \; \sqrt{1-\rho^2}\, dW_t^{(1)}\, dW_t^{(2)} \; = \; \rho\, dt. \end{aligned}$$

We integrate and take expectation and obtain

$$W_t^{(1)} W_t^* \;=\; \int_0^t W_u^{(1)}\, dW_u^* \;+\; \int_0^t W_u^*\, dW_u^{(1)} \;+\; \rho t.$$

Since the Itô integrals on the right–hand side are martingales,

$$E\left[\int_0^t W_u^{(1)}\, dW_u^*\right] = E\left[\int_0^0 W_u^{(1)}\, dW_u^*\right] \;=\; 0, \quad \text{and} \quad E\left[\int_0^t W_u^*\, dW_u^{(1)}\right] \;=\; E\left[\int_0^0 W_u^*\, dW_u^{(1)}\right] \;=\; 0.$$

From this and $E[W_t^1] \;=\; E[W_t^*] \;=\; 0$ we conclude that

$$\mathrm{Cov}[W_t^{(1)}, W_t^*] \;=\; E[W_t^{(1)} W_t^*] \;=\; \rho t.$$

Assertion **(3)** follows since $\mathrm{Var}[W_t^{(1)}] = \mathrm{Var}[W_t^* = t]$ ■

10.4 Exercises for Ch.??

None at this time!

11 Girsanov's Theorem and the Martingale Representation Theorem

11.1 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 11.1. *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 $\frac{d\widetilde{P}}{dP}$. See Chapter 4.6 (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.,

$$\widetilde{E}(Y) \;=\; \int_\Omega U\,d\widetilde{P}\,.$$

For the following 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.

$$\widetilde{E}[Y] \;= E[YZ_t], \tag{11.1}$$

$$\widetilde{E}\big[Y \,|\, \mathfrak{F}_t\big] \;= \frac{1}{Z_t}\,E[Y Z_{t+h} \,|\, \mathfrak{F}_t] \tag{11.2}$$

PROOF: ⋆

A. We show that $\widetilde{P}$ is a probability measure which is equivalent to P.

$$\widetilde{P}(\Omega) \;=\; \int_\Omega Z\,dP \;=\; E[Z] \;=\; 1.$$

This proves that $\widetilde{P}$ is a probability measure. 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 4.13 on p.70.
Let $Z' := (1/Z)1_{Z>0}$. Then

$$\begin{aligned} 0 \;=\widetilde{P}(A) \;&=\; \int_A 1\,dP \;=\; \int_A ZZ'\,dP \;+\; \int_A 1\cdot 1_{Z=0}\,dP \;=\; \int_A ZZ'\,dP \;+\;0 \\ &=\; \int (1_A Z')Z\,dP \;=\; \int 1_A Z'\,d\widetilde{P} \;=\; \int_A Z'\,d\widetilde{P} \;=\; 0. \end{aligned}$$

The last equality follows from Proposition 4.13, applied to $\mu := \widetilde{P}$ and $f := Z'$. We have shown that all $\widetilde{P}$–null sets are P–null sets, thus $P \sim \widetilde{P}$.

B. Proof of (11.1). 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:

$$\widetilde{E}[Y] = E[YZ] = E\big[E[YZ \mid \mathfrak{F}_t] \mid\big] = E\big[Y\, E[Z \mid \mathfrak{F}_t]\big] = E[YZ_t]. \quad \blacksquare$$

C. Proof of (11.2). To prove that $\frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]$ is the conditional expectation of Y w.r.t. $\mathfrak{F}_t$ and $\widetilde{P}$ (not P!) we must show that

(1) $\frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]$ is $\mathfrak{F}_t$–measurable,

(2) $\frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]$ satisfies the partial averaging property

$$(\mathbf{A}) \qquad \int_A \frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]\, d\widetilde{P} = \int_A Y\, d\widetilde{P} \ \text{ for all } A \in \mathfrak{F}_t.$$

(1) is trivially since $E[\cdots \mid \mathfrak{F}_t]$ enforces $\mathfrak{F}_t$–measurability.

To prove **(2)** we first note that formula (11.1) with $1_A \frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]$ in place of Y yields

$$(\mathbf{B}) \qquad \widetilde{E}\left[1_A \frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]\right] = E\left[1_A \frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t] \cdot Z_t\right] = E\big[1_A E[YZ_{t+h} \mid \mathfrak{F}_t]\big],$$

and when we apply it with $1_A Y$ in place of Y and Z_{t+h} in place of Z_t then we obtain

$$(\mathbf{C}) \qquad \widetilde{E}[1_A Y] = E[1_A Y Z_{t+h}].$$

To prove (**A**) we write

$$\int_A \frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]\, d\widetilde{P} = \widetilde{E}\left[1_A \frac{1}{Z_t} E[YZ_{t+h} \mid \mathfrak{F}_t]\right] \overset{(\mathbf{B})}{=} E\big[1_A E[YZ_{t+h} \mid \mathfrak{F}_t]\big]$$
$$= E\big[E[1_A Y Z_{t+h} \mid \mathfrak{F}_t]\big] = E\big[1_A Y Z_{t+h}\big] \overset{(\mathbf{C})}{=} \widetilde{E}[1_A Y] = \int_A Y\, d\widetilde{P}.$$

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. $\blacksquare$

11.2 Onedimensional Girsanov and Martingale Representation Theorems

The following is SCF2 Theorem 5.2.3.

Theorem 11.1 (Girsanov's Theorem in one dimension). *Let $T > 0$ and let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ 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 $0 \leqq t \leqq T$. Let W_t, be a Brownian motion on this filtered space and let Θ_t, be an adapted process which satisfies the integrability condition*

$$(11.3) \qquad \boxed{\star} \qquad E\left[\int_0^T \Theta_u^2\, Z_u^2\, du\right] < \infty.$$

where the process Z_t is defined in terms of Θ_t by formula (11.4) below.

Let

$$Z_t := \exp\left\{-\int_0^t \Theta_u \, dW_u - \frac{1}{2}\int_0^t \Theta_u^2 \, du\right\}, \tag{11.4}$$

$$\widetilde{P}(A) := \int_A Z_T \, dP \text{ for all } A \in \mathfrak{F}_T \quad i.e., \quad Z_T = \frac{d\widetilde{P}}{dP}, \tag{11.5}$$

$$\widetilde{W}_t = W_t + \int_0^t \Theta_u \, du, \quad i.e., \quad d\widetilde{W}_t = dW_t + \Theta_t \, dt. \tag{11.6}$$

Then $\widetilde{P}$ is a probability equivalent to P and $\widetilde{W}_t$, $0 \leqq t \leqq T$, is a Brownian motion w.r.t. $\widetilde{P}$.

PROOF ⋆ : See the proof of SCF2 Theorem 5.2.3. ■

Remark 11.1. ⋆

Strictly speaking it is not correct to write $Z_T = \frac{d\widetilde{P}}{dP}$ in (11.5) 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 5.1 (Existence Theorem for Conditional Expectations) on p.82. □

Remark 11.2. 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

$$dS_t = \alpha_t S_t \, dt + \sigma_t S_t \, dW_t, \quad 0 \leqq t \leqq T. \tag{11.7}$$

and we have a discount process with an interest rate R_t which can be stochastic (adapted):

$$D_t = \exp\left[-\int_0^t R_s \, ds\right], \tag{11.8}$$

(see Definition 6.5 on p.91), and if we define an adapted process Θ_t to be the so called market price of risk,

$$\Theta_t = \frac{\alpha_t - R_t}{\sigma_t}, \tag{11.9}$$

then the discounted stock price has the dynamics

$$d\,(D_t S_t) = \sigma_t D_t S_t \big[\Theta_t dt + dW_t\big]. \tag{11.10}$$

We apply Girsanov theorem and substitute W_t with the Brownian motion $\widetilde{W}_t$ of formula (11.6) in that last equation. We obtain

$$d\,(D_tS_t) \;=\; \sigma_t D_t S_t\, d\widetilde{W}_t]. \tag{11.11}$$

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 $\widetilde{P}$, hence,

$$D_tS_t \;=\; \widetilde{E}[D_TS_T \mid \mathfrak{F}_t]. \tag{11.12}$$

Now let us switch to self–financing portfolios

$$\vec{H}_t \;=\; \left(H_t^B, H_t^S\right) \;=\; \left(X_t - \Delta_tS_t, \Delta_t\right)$$

Here we have given both the notion of MF454 Chapter 6 (Financial Models - Part 1) 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 (11.12) it will follow that the discounted portfolio value process has dynamics

$$d\big(D_tX_t\big) \;= \Delta_t\sigma_t D_tS_t\, d\widetilde{W}_t. \tag{11.13}$$

Thus D_tX_t also is a $\widetilde{P}$–martingale. We obtain

$$D_tX_t \;=\; \widetilde{E}[D_TX_T \mid \mathfrak{F}_t]. \tag{11.14}$$

Now we get to the really important part. If we have a contingent claim $\mathcal{X}$ with pricing process $\Pi_t(\mathcal{X})$ and $\vec{H}$ 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 $D_TX_T = D_T\mathcal{X}$ and the pricing principle which results from the no arbitrage condition implies that

$$X_t \;=\; \Pi_t(\mathcal{X}), \qquad \text{hence} \quad D_tX_t \;=\; D_t\Pi_t(\mathcal{X}) \;\text{ for } 0 \leqq t \leqq T. \tag{11.15}$$

We have found the long sought after pricing formula for a contingent claim based on a risky asset with generalized GBM as its price process S_t. It follows from (11.14) and (11.15) that

$$\Pi_t(\mathcal{X}) \;=\; \frac{1}{D_t}\widetilde{E}[D_TX_T \mid \mathfrak{F}_t]. \tag{11.16}$$

This formula will be used, e.g., to derive the formula (9.21) of Theorem 9.1 on p.147 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 (11.15) and (11.16) 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. $\square$

We have seen that being a martingale represents a very strong condition concerning what such a process can look like. Lévy's characterization of onedimensional Brownian Motion (Theorem 10.5 on p.159) 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 11.2 (Martingale representation, one dimension)**.**

Let $T > 0$. Assume that
- $W_t, 0 \leqq t \leqq T$ *is a Brownian motion on a probability space* $(\Omega, \mathfrak{F}, P)$,
- $\mathfrak{F}_t^W, 0 \leqq t \leqq T$ *is the filtration generated by this Brownian motion,*
- $M_t, 0 \leqq t \leqq T$, *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* $0 \leqq s \leqq t \leqq T$.

Then there exists an adapted process $\Gamma_u, 0 \leqq u \leqq T$, *such that*

$$M_t = M_0 + \int_0^t \Gamma_u \, dW_u, \quad 0 \leqq t \leqq T. \tag{11.17}$$

PROOF: Will not be given here. It also cannot be found in SCF2. ■

Remark 11.3.

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.

We will not get to that point in this course but note that this is done in SCF2 Chapter 11 (Introduction to Jump Processes) in which stock price is driven by (generalized) Poisson processes in addition to Brownian motion. □

We add the assumption $\mathfrak{F}_t = \mathfrak{F}_t^W$ to Girsanov's Theorem 11.1. This results in the following corollary (SCF2 Corollary 5.3.2).

Corollary 11.1. *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 $\mathfrak{F}_t^W, 0 \leqq t \leqq T$, i.e., the filtration* ***generated by*** *W_t (!) which satisfies the integrability condition*

$$\boxed{\star} \qquad E\left[\int_0^T \Theta_u^2 \, Z_u^2 \, du\right] < \infty. \tag{11.18}$$

- *Let* $Z_t := \exp\left\{ -\int\limits_0^t \Theta_u \, dW_u - \frac{1}{2}\int\limits_0^t \Theta_u^2 \, du \right\}$,
- $\widetilde{P}(A) := \int_A Z_T \, dP$ *for all* $A \in \mathfrak{F}_T$, *i.e.,* $Z_T = \dfrac{d\widetilde{P}}{dP}$,
- $\widetilde{W}_t = W_t + \int\limits_0^t \Theta_u \, du$, *i.e.,* $d\widetilde{W}_t = dW_t + \Theta_t \, dt$.
- *Let* $\widetilde{M}_t, 0 \leqq t \leqq T$, *be an* $\mathfrak{F}_t^W$*–martingale under* $\widetilde{P}$ (***not P!***)

Then there exists an $\mathfrak{F}_t^W$*–adapted process* $\widetilde{\Gamma}_u, 0 \leqq u \leqq T$, *such that*

$$\widetilde{M}_t = \widetilde{M}_0 + \int_0^t \widetilde{\Gamma}_u \, d\widetilde{W}_u, \ 0 \leqq t \leqq T. \tag{11.19}$$

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 11.2 on p.164 showed the significance of Girsanov's Theorem and alluded to that of the martingale representation theorem when modeling contingent claims with one underlying risky asset. 11.1. We need multidimensional versions of those theorems to model claims with several underlying risky assets.

11.3 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 10.1 on p.155.

The following is SCF2 Theorem 5.4.1.

Theorem 11.3 (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* $0 \leqq t \leqq T$. *Let* $\vec{W}_t$ *be a multidimensional Brownian motion*

$$\vec{W}_t = (W_t^{(1)}, \dots, W_t^{(1)})$$

(thus the coordinate processes $W_i(t)$ *are independent). w.r.t. the filtration* $\mathfrak{F}_t, 0 \leqq t \leqq T$. *Let*

$$\vec{\Theta}_t = (\Theta_t^{(1)}, \dots, \Theta_t^{(1)})$$

be a d–dimensional adapted process which satisfies the integrability condition

$$\boxed{\star} \qquad E\left[\int\limits_0^T \|\vec{\Theta}_u\|_2^2 \, Z_u^2 \, du \right] < \infty. \tag{11.20}$$

Here $\|\vec{x}\|_2$ *is the standard Euclidean norm in* $\mathbb{R}^d$*. See Example* 7.4 *on p.*124*.*

Let

(11.21) $$Z_t := \exp\left\{-\int\limits_0^t \vec{\Theta}_u \bullet d\vec{W}_u - \frac{1}{2}\int\limits_0^t \|\vec{\Theta}_u\|^2\,du\right\},$$

(11.22) $$\widetilde{P} : A \mapsto \int_A Z_T\,dP, \quad \text{i.e.,} \quad Z_T = \frac{d\widetilde{P}}{dP},$$

(11.23) $$\widetilde{\vec{W}}_t = \vec{W}_t + \int\limits_0^t \vec{\Theta}_u\,du, \quad \text{i.e.,} \quad d\widetilde{\vec{W}}_t = d\vec{W}_t + \vec{\Theta}_t\,dt.$$

Then $\widetilde{P}$ *is a probability equivalent to* P *and* $\widetilde{\vec{W}}_t$*,* $0 \leqq t \leqq T$*, is a Brownian motion w.r.t.* $\widetilde{P}$*.*

Note that the vector equations in 11.23 *are to be understood componentwise:*

$$\widetilde{W}_t^{(j)} = W_t^{(j)} + \int\limits_0^t \Theta_u^{(j)}\,du, \quad \text{i.e.,} \quad d\widetilde{W}_t^{(j)} = dW_t^{(j)} + \Theta_t^{(j)}\,dt \;\; \text{for } j = 1, \dots, d.$$

PROOF [⋆] : Will not be given here. ■

Remark 11.4. 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 **independent** w.r.t. the new probability $\widetilde{P}$. This is not at all an obvious consequence of 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 11.2 (Martingale representation, one dimension) on p.166. This is SCF2 Theorem 5.4.2.

Theorem 11.4 (Martingale representation theorem, multiple dimensions)**.** *Let* T *be a fixed positive time, and assume that*

- $\vec{W}_t, 0 \leqq t \leqq T$ *is a d–dimensional Brownian motion on a probability space* $(\Omega, \mathfrak{F}, P)$*,*
- $\mathfrak{F}_t^{\vec{W}}, 0 \leqq t \leqq T$ *is the filtration generated by this Brownian motion,*
- $M_t, 0 \leqq t \leqq T$*, is a (one–dimensional) P–martingale with respect to this filtration.*

Then there is an adapted d–dimensional process $\vec{\Gamma}_u = \big(\Gamma_1(u), \dots, \Gamma_d(u)\big), 0 \leqq u \leqq T$*, such that*

(11.24) $$M_t = M_0 + \int_0^t \vec{\Gamma}_u \bullet d\vec{W}_u, \quad 0 \leqq t \leqq T.$$

We now assume in addition to the assumptions stated so far the notation and assumptions of Girsanov's Theorem in multiple dimensions (Theorem 11.3*). Then the following also is true.*

Let $\widetilde{M}_t, 0 \leqq t \leqq T$, be a (one–dimensional) $\widetilde{P}$–martingale with respect to $\mathfrak{F}_t^{\vec{W}}, 0 \leqq t \leqq T$, 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 $\vec{\widetilde{\Gamma}}_u = \left(\widetilde{\Gamma}_u^{(1)}, \dots, \widetilde{\Gamma}_u^{(d)}\right), 0 \leqq u \leqq T$, such that

$$\widetilde{M}_t = \widetilde{M}_0 + \int_0^t \vec{\widetilde{\Gamma}}_u \bullet d\vec{\widetilde{W}}_u, 0 \leqq t \leqq T. \tag{11.25}$$

PROOF: Will not be given here. ■

11.4 Exercises for Ch.11

11.4.1 Exercises for xxx2

None yet

12 Black–Scholes Model Part II: Risk–neutral Valuation

In this chapter we elaborate on Remark 11.2 which gave an outline of how Girsanov's Theorem (Theorem 11.1 would be crucial in pricing a contingent claim.

12.1 The Onedimensional Generalized Black–Scholes Model

In Chapter 9 (Black–Scholes Model Part I: The PDE), Definition 9.1 on p.143 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 risky asset price S_t are assumed to be constant. We weaken those assumptions for most of this entire chapter 12.

Definition 12.1 (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 $0 \leqq t \leqq T$ Let $W_t, 0 \leqq t \leqq T$, be a Brownian motion w.r.t $\mathfrak{F}_t$.

We no more requrie that the instantaneous mean rate of return α, the volatility σ of the risky asset S_t, and the interest rate r that governs investments in the bank account 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

(12.1) $$E\left[\int_0^T \Theta_u^2 \, Z_u^2 \, du\right] < \infty,$$

We speak of a **generalized Black–Scholes market model** if

(12.2) $$dD_t = -R_t D_t \, dt; \; D_0 = 1;$$

(12.3) $$dS_t = \alpha_t S_t \, dt + \sigma_t S_t \, dW_t; \;\; S_0 \in]0, \infty[\,; \alpha_t, \sigma_t \in]0, \infty[\,;$$

(12.4) The market is efficient: No arbitrage portfolios.

- 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). □

Remark 12.1. First some remarks about the process D_t.

(1) From (12.2) we obtain

(12.5) $$D_t = \exp\left[-\int_0^t R_u du\right].$$

This follows easily from differentiating the right hand side with respect to t.

(2) We could have worked instead with the interest rate process

$$dB_t \;=\; R_t B_t dt;\; B_0 = 1, \quad \text{i.e.}, B_t \;=\; \exp\left[\int_0^t R_u du\right] \;=\; \frac{1}{D_t}$$

but using D_t instead will make it easier to relate the contents of this chapter to the SCF2 text.

Also be aware of the following.

(3) Formula (12.3) 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

(12.6) $$S_t \;=\; \exp\left[\int_0^t \sigma_u dW_u \;+\; \int_0^t \left(\alpha_u - \frac{1}{2}\sigma_u^2\right) ds\right],$$

See Remark 8.1 on p.138, the subsequent Remark 8.5, and (8.15) on p.138. .

(4) It was not necessary to explicitly require the adaptedness of the processes S_t and D_t. Formula (12.2) (equivalently, formula (12.5)) implies that, as far as measurability is concerned, D_t only depends on the adapted process R_s for $s \leqq t$, and thus only on information in $\mathfrak{F}_t$, i.e., D_t is adapted. We conclude similarly that formula (12.3) (equivalently, formula (12.6)) implies that measurability of S_t only depends on the adapted process W_s. Thus S_t is adapted.

(5) Recall from Assumption 6.1 on p.94 that we always assume that, besides being free of arbitrage, the market has complete liquidity, no transaction costs and no bid–ask spread. □

Remark 12.2. The degree of uncertainty, i.e., the risk of investing in the bank account is qualitatively much smaller than that of investing in the stock for the following reasons.

Only the randomness of the process R_t within a small interval $[t, t+h]$ affects that of the change $D_{t+h} - D_t$. This results in quadratic variation $[D, D]_t = 0$ since $dtdt = 0$, 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 moption W_t which 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 (11.9) of Remark 11.2 on p.164 already introduced the market price of risk. Here is the formal definition.

Definition 12.2. For the generalized Black–Scholes market economy of Definition 12.1 on p.170,

the **market price of risk**. is the process

$$\Theta_t \;=\; \frac{\alpha_t - R_t}{\sigma_t}, \tag{12.7}$$

Note that Θ_t is adapted as the difference and quotient of adapted processes. □

Remark 12.3. The assumption (12.1) on p.170,

$$E\left[\int_0^T \Theta_u^2\, Z_u^2\, du\right] \;<\; \infty, \tag{12.8}$$

will allow us to apply Girsanov's Theorem to the market price of risk process. □

12.2 Risk–Neutral Measure in a Generalized Black–Scholes Market

Assumption 12.1.

We assume for the entire remainder of this Chapter 12 (Black–Scholes Model Part II: Risk–neutral Valuation) that we have a generalized Black–Scholes market as defined in Definition 12.1 on p.170. □

Introduction 12.1. We recall definitions (6.13) on p.99 and (6.16) on p.102 of the binomial asset model in which we defined a risk–neutral measure, also called there a martingale measure, as a probability measure Q equivalent to the "true" probability which made discouned stock price D_tS_t a Q–martingale. To see that, observe that the (not continuously) compounded interest earned between times t and $t+k$ ($k \in \mathbb{N}$) in the bank is $(1+R)^k$, thus the discount factor is

$$D_t = \frac{1}{(1+R)^k}.$$

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 12.3 (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 12.1. *The discounted stock price has the following dynamics and explicit representation.*

(12.9) $$d\,(D_tS_t) \;=\big(\alpha_t \;-\; R_t\big)D_tS_t\,dt \;+\; \sigma_tD_tS_t\,dW_t$$

(12.10) $$D_tS_t \;= S_0\;\exp\left\{\int_0^t \sigma_s\,dW_s \;+\; \int_0^t \big(\alpha_s \;-\; R_s \;-\; \frac{1}{2}\sigma_s^2\big)\,ds\right\},$$

> *In other words, discounting S_t transforms this generalized GBM with instantaneous mean rate of return α_t and volatility σ_t into another generalized GBM with reduced instantaneous mean rate of return $\alpha_t - R_t$ and unchanged volatility σ_t.*

We further can express $d(D_tS_t)$ with help of the market price of risk process Θ_t given in (12.7) as follows.

(12.11) $$d\,(D_tS_t) \;=\sigma_tD_tS_t\big(\Theta_t\,dt \;+\; dW_t\big).$$

PROOF: Formula (12.10) is obtained by multiplying the right hand sides of (12.5) and (12.6).
We replace in the formulas (8.18) on p.138 and (8.21) on p.139 which α_t with $\alpha_t - R_t$ and this proves that the dynamics of the generalized GBM (12.9) are, in fact, given by (12.10).
Formula (**??**) follows immediately from (12.10) because $\alpha_t - R_t \;=\; \Theta_t\sigma_t$. ■

As a consequence of Girsanov's Theorem we can prove the existence of a risk–neutral measure.

Theorem 12.1. *Let the process $Z_t(0 \leqq t \leqq T)$ 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 12.2 on p.172. 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, $d\widetilde{W}_t \;=\; dW_t \;+\; \Theta_t\,dt$ and $\widetilde{W}_0 = 0$,*
  is an $\mathfrak{F}_t$–Brownian motion w.r.t the new probability measure $\widetilde{P}$.
- *Discounted stock price D_tS_t is a $\widetilde{P}$–martingale.*

PROOF: We may apply Theorem 11.1 (Girsanov's Theorem in one dimension) on p.163 to the market price of risk process Θ_t since (12.1) implies that the integrability condition (11.3) of that theorem is satisfied. The only item that is not an immediate consequence of Theorem 11.1 is the assertion that D_tS_t is a $\widetilde{P}$–martingale.
We see this by substituting $d\widetilde{W}_t \;=\; dW_t + \Theta_t dt$ into formula (12.11). We obtain

(12.12) $$\begin{aligned} d\,(D_tS_t) \;&= \sigma_tD_tS_t\big(d\widetilde{W}_t\big),\\ \text{i.e., } D_tS_t \;&= S_0 \;+\; \int_0^t \sigma_uD_uS_u\,d\widetilde{W}_u. \end{aligned}$$

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 12.1 (Existence of a risk–neutral measure)**.**

- *The probability measure $\widetilde{P}$ of Theorem 12.1 is a risk–neutral measure for the generalized Black–Scholes market in he sense of Definition 12.3 on p.172.*
- *The dynamics of discounted stock price when using $\widetilde{W}_t$ instead of W_t are*

$$d\left(D_t S_t\right) \;=\; \sigma_t D_t S_t\big(d\widetilde{W}_t\big). \tag{12.13}$$

PROOF: Formula (12.13) was established in the proof of Theorem 12.1. The remainder is an obvious consequence of that theorem. ■

Remark 12.4. Note the following.

- (12.13) 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 $0 \leqq t \leqq T$.
- 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 9.2 on p.144.) 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 $\alpha_t > R_t$. □

12.3 Dynamics of Discounted Stock Price and Portfolio Value

We saw in Chapter 9.2 (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 (12.14) on p.174,

$$dX_t \;=\; \Delta_t dS_t \,+\, r\big(X_t - \Delta_t S_t\big)\,dt.$$

In the generalized Black–Scholes market we obtain dX_t by replacing the constant interest rate r with the varying interest rate $R_t(\omega)$.

Proposition 12.2. *The budget equation for a self–financing portfolio strategy in a Black–Scholes market is*

$$dX_t \;= \Delta_t\, dS_t \,+\, R_t\big(X_t - \Delta_t\, S_t\big)\,dt \tag{12.14}$$

Further we have the following equation for the portfolio value dynamics.

$$dX_t \;= R_t X_t\,dt \,+\, \Delta_t \sigma_t S_t\left[\,\Theta_t\,dt \,+\, dW_t\,\right]. \tag{12.15}$$

PROOF: Equation (12.14) is obvious. It just states that the number Δ_t of shares held in the risky asset increases by the change dS_t inasset price, and the value of the bank account holdings $X_t - \Delta_t S_t$ changes during dt according to the interest rate, R_t.

We repeat here the proof of (12.15) as can be found in SCF2, Chapter 5.2.3 (Value of Portfolio Process Under the Risk–Neutral Measure).

$$\begin{aligned} dX_t &= \Delta_t\, dS_t + R_t\big(X_t - \Delta_t\, S_t\big)\, dt \\ &= \Delta_t\left(\alpha_t S_t\, dt + \sigma_t S_t\, dW_t\right) + R_t\big(X_t - \Delta_t\, S_t\big)\, dt \\ &= R_t X_t\, dt + \Delta_t\big(\alpha_t - R_t\big) S_t\, dt + \Delta_t \sigma_t S_t\, dW_t \\ &= R_t X_t\, dt + \Delta_t \sigma_t S_t \left[\, \Theta_t\, dt + dW_t \,\right]. \end{aligned}$$

The last equation follows from $\alpha_t - R_t = \sigma_t \Theta_t$. ■

Proposition 12.3. *The discounted portfolio value $D_t X_t$ has dynamics*

$$d\big(D_t X_t\big) = \Delta_t \sigma_t D_t S_t\, d\widetilde{W}_t. \tag{12.16}$$

PROOF: Again we follow SCF2. To obtain $d(D_t X_t)$ we use the Itô product rule

$$d(D_t X_t) = D_t\, dX_t + X_t dD_t + dD_t\, dX_t. \tag{\textbf{A}}$$

First we note that

$$dD_t dX_t = (-R_t D_t dt)\left((R_t X_t + \Delta_t \sigma_t S_t \Theta_t)\, dt + \Delta_t \sigma_t S_t\, dW_t\right) = 0,$$

because $dtdt = 0$ and $dtdW_t = 0$. That plus $dD_t = -R_t D_t dt$ applied to (**A**) yields

$$d(D_t X_t) = D_t\, dX_t - X_t(R_t D_t dt) + 0.$$

Next we apply (12.15) and obtain

$$\begin{aligned} d(D_t X_t) &= D_t\left(R_t X_t\, dt + \Delta_t \sigma_t S_t \left[\, \Theta_t\, dt + dW_t \,\right]\right) - X_t(R_t D_t dt) \\ &= D_t R_t X_t\, dt + D_t \Delta_t \sigma_t S_t \left[\, \Theta_t\, dt + dW_t \,\right] \\ &\quad - X_t R_t D_t dt \\ &= D_t \Delta_t \sigma_t S_t \left[\, \Theta_t\, dt + dW_t \,\right]. \end{aligned}$$

This proves (12.15). ■

It follows from Proposition 12.3 that $D_t X_t$ is a martingale under $\widetilde{P}$, thus

$$D_t X_t = \widetilde{E}[D_T X_T \,|\, \mathfrak{F}_t] = \widetilde{E}[D_T V_T \,|\, \mathfrak{F}_t] \ \text{ for all } 0 \leqq t \leqq T. \tag{12.17}$$

Now assume that X_t is the value of the hedging portfolio for a contingent claim $\mathcal{X}$ We follow SCF2 notation and write

$$V_T \text{ instead of } \mathcal{X}, \quad \text{and} \quad V_t \text{ instead of } \Pi_t(\mathcal{X}),$$

to denote contract value and pricing function of a contingent claim.

It follows from the pricing principle that $V_t = X_t$ and thus $D_t V_t = D_t X_t$ must be satisfied to avoid arbitrage. We obtain from Proposition 12.3 the following

Corollary 12.2. *Assume that X_t is the value process of a hedging portfolio for a contingent claim with price process V_t ($0 \leqq t \leqq T$). Then*

$$\begin{aligned} D_t V_t &= \widetilde{E}[D_T V_T \,|\, \mathfrak{F}_t],\ 0 \leqq t \leqq T. \\ V_t &= \widetilde{E}\left[e^{-\int_t^T R_u du} V_T \,\middle|\, \mathfrak{F}_t\right],\ 0 \leqq t \leqq T. \end{aligned}$$

PROOF: The equation for $D_t V_t$ results from this process being a $\widetilde{P}$–martingale. The formula for V_t 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 12.4 (Risk–neutral valuation formula)**.** We call either one of the Corollary 12.2 formulas,

(12.18) $$D_t V_t = \widetilde{E}[D_T V_T \,|\, \mathfrak{F}_t],\ 0 \leqq t \leqq T.$$

(12.19) $$V_t = \widetilde{E}\left[e^{-\int_t^T R_u du} V_T \,\middle|\, \mathfrak{F}_t\right],\ 0 \leqq t \leqq T.$$

the **risk–neutral pricing formula**, also the **risk–neutral valuation formula** for a contingent claim with contract function V_T. □

12.4 Risk–Neutral Pricing of a European Call

Assumption 12.2. 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 security payoff is $V_T = (S_T - K)^+$. □

We now derive the Black–Scholes formula for the price of a European call. [22] Since the contract function for a European call is

$$V_T = \Phi(S_T) = (S_T - K)^+,$$

[22]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 would not be Markov since the entire past enters the dynamics $dS_t = \alpha_t S_t dt + \sigma S_t dt$.

the right–hand side of the risk–neutral valuation formula (**??**) on p.**??** reads

(12.20) $$\widetilde{E}\left[e^{-r(T-t)}(S_T-K)^+\,|\mathfrak{F}_t\right].$$

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

(12.21) $$c\big(t,S_t\big) \;=\widetilde{E}\left[e^{-r(T-t)}(S_T-K)^+\,|\,\mathfrak{F}_t\right].$$

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 (12.21) 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 $0\leqq t\leqq T$ and define

(12.22) $$\tau\;:=T\,-\,t;\qquad\qquad Y\;:=\;-\,\frac{\widetilde{W}_T\,-\,\widetilde{W}_t}{\sqrt{\tau}}.$$

(12.23) $$h(t;x,y)\;:=e^{-r\tau}\left(x\cdot\exp\left\{-\sigma\sqrt{\tau}y\,+\,\left(r\,-\,\frac{\sigma^2}{2}\right)\tau\right\}\,-\,K\right)^+.$$

Note that Y is standard normal w.r.t. $\widetilde{P}$ since $\widetilde{W}_t, t\geqq 0$, is a $\widetilde{P}$–Brownian motion.
We next provide three lemmas which have the following purpose.

- Lemma 12.1 shows that we can work with $h(t;S_t,Y)$ instead of $e^{-r\tau}(S_T-K)^+$.
- Lemma 12.2 gives the definition of $c(t,x)$ in terms of $h(t;x,y)$.
- Lemma 12.3 allows us to actually compute $c(t,x)$. The result will be formula (9.21) of Theorem **??** on p.147 where it was stated without proof.

Lemma 12.1. *With the above definitions we can rewrite the risk–neutral valuation formula* (12.20) *for a European call as follows.*

(12.24) $$\widetilde{E}\left[e^{-r\tau}(S_T-K)^+\,|\mathfrak{F}_t\right]\;=\widetilde{E}\left[h(t;S_t,Y)\,|\,\mathfrak{F}_t\right]$$

PROOF: According to (**??**) on p.**??**,

$$S_t\;=S_0\,\exp\left\{\int_0^t\sigma_s\,d\widetilde{W}_s\,+\,\int_0^t\left(R_s ds\,-\,\frac{1}{2}\,\sigma_s^2\right)ds\right\}\;=\;S_0\,\exp\left\{\sigma\,\widetilde{W}_t\,+\,\left(r\,-\,\frac{1}{2}\,\sigma^2\right)\,t\right\}.$$

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

$$\begin{aligned} S_T &= S_t \cdot \exp\left\{\sigma\,(\widetilde{W}_T \,-\, \widetilde{W}_t) \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\,(T-t)\right\} \\ &= S_t \cdot \exp\left\{-\sigma\tau\,\frac{-(\widetilde{W}_T \,-\, \widetilde{W}_t)}{\tau} \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\,(T-t)\right\} \\ &\overset{(12.22)}{=} S_t \cdot \exp\left\{-\sigma\tau\,Y \,+\, \left(r \,-\, \frac{1}{2}\,\sigma^2\right)\,(T-t)\right\}. \end{aligned}$$

It follows from that equation for S_T that

$$\begin{aligned} h(t; S_t, Y) &= e^{-r\tau}\left(S_t \cdot \exp\left\{-\sigma\sqrt{\tau}Y \,+\, \left(r \,-\, \frac{\sigma^2}{2}\right)\tau\right\} \,-\, K\right)^+ \\ &= e^{-r\tau}(S_T \,-\, K)^+. \end{aligned}$$

We apply conditional expectations $\widetilde{E}[\cdots \mid \mathfrak{F}_t]$ to both sides and assertion (12.24) follows. ■

We remember our goal: find a function $(t,x) \mapsto c(t,x)$ such that (12.21) holds:

(12.25) $$c\big(t, S_t\big) \;= \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+ \,\middle|\, \mathfrak{F}_t\right].$$

Lemma 12.1 allows us to reformulate this problem as follows: Let $h(t;x,y)$ be the function given in formula (12.23). We want to find a function $(t,x) \mapsto c(t,x)$ such that

(12.26) $$c(t, S_t) \;= \widetilde{E}\left[h(t; S_t, Y) \,\middle|\, \mathfrak{F}_t\,\right].$$

The next lemma shows how to define this function $c(t,x)$.

Lemma 12.2. *Let*

(12.27) $$c(t,x) \;:= \widetilde{E}[h(t;x,Y)].$$

where $h(t;x,y)$ is the function defined in (**??**)*. Then $c(t,S_t)$ satisfies* (12.26) *and hence also the risk–neutral pricing formula* (12.21)*, i.e.,*

(12.28) $$c(t, S_t) \;=\; \widetilde{E}\left[e^{-r\tau}(S_T - K)^+ \,\middle|\, \mathfrak{F}_t\,\right].$$

PROOF: We fix $0 \leqq t \leqq T$. 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 $0 \leqq t \leqq T$ from the Independence Lemma (Lemma 5.3 on p.86) [23] that

$$c(t, S_t) \;= \widetilde{E}\left[h(t; S_t, Y) \,\middle|\, \mathfrak{F}_t\,\right].$$

[23] 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 (12.26). We apply Lemma 12.1 and (12.28) 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

$$V_t \;=\; c\big(t, S_t\big) \;= \widetilde{E}\left[e^{-r(T-t)}(S_T - K)^+ \,\big|\, \mathfrak{F}_t\right]. \tag{12.29}$$

It follows from the definition of $h(t;x,y)$ given in (12.23) 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 Z. 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 (**??**) on p.**??**:

$$d_\pm(\tau,x) \;= \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} \,+\, \left(r \pm \frac{\sigma^2}{2}\right)\tau\right], \tag{12.30}$$

Lemma 12.3. *The pricing function $c(t,x)$ for a European call option is given by the formula*

$$c(t,x) \;= x\, N\big(d_+(\tau,x)\big) \,-\, e^{-r\tau} K\, N\big(d_-(\tau,x)\big). \tag{12.31}$$

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

$$\begin{aligned} c(t;x) \;&\overset{(12.27)}{=}\; \widetilde{E}[h(t;x,Y)] \;=\; \int_{-\infty}^{\infty} h(t;x,y)\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}\, dy \\ &\overset{(12.23)}{=}\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-r\tau}\left(x\cdot\exp\left\{-\sigma\sqrt{\tau}y \,+\, \left(r \,-\, \frac{\sigma^2}{2}\right)\tau\right\} \,-\, K\right)^+ \; e^{-\frac{y^2}{2}}\, dy \end{aligned}$$

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

$$\begin{aligned} &\log x \,+\, \left\{-\sigma\sqrt{\tau}y \,+\, \left(r \,-\, \frac{\sigma^2}{2}\right)\tau\right\} \;>\; \log K \\ \Leftrightarrow\; &\log x \,-\, \log K \,+\, \left(r \,-\, \frac{\sigma^2}{2}\right)\tau \;>\; \sigma\sqrt{\tau}y \\ \Leftrightarrow\; &\sigma\sqrt{\tau}y \;<\; \frac{\log x}{\log K} \,+\, \left(r \,-\, \frac{\sigma^2}{2}\right)\tau \\ \Leftrightarrow\; &y \;<\; \frac{1}{\sigma\sqrt{\tau}}\left[\frac{\log x}{\log K} \,+\, \left(r \,-\, \frac{\sigma^2}{2}\right)\tau\right] \;=\; d_-(\tau,x). \end{aligned} \tag{12.32}$$

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 can 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 $\int_{-\infty}^{d_-(\tau,x)} e^{-\frac{1}{2}y^2}\, dy$ 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 (12.31). The last equation holds because, according to (12.30),

$$\begin{aligned} d_+(\tau,x) &= d_-(\tau,x) + \sigma\sqrt{\tau} \\ &= \frac{1}{\sigma\sqrt{\tau}} \left[\log\frac{x}{K} + \left(r + \frac{1}{2}\sigma^2 \right) \tau \right]. \quad \blacksquare \end{aligned} \tag{12.33}$$

We have thus given the proof of Theorem 9.1 on p.147 since the classical Black–Scholes market condidions under which it was stated satisfy the assumptions 12.2 on p.176. The difference is that the function $c(t,x)$ was given there as the solution to the (deterministic) Black–Scholes PDE (9.16)

$$c_t(t,x) + rx\, c_x\big(t,x\big) + \frac{1}{2}\sigma^2 x^2 c_{xx}(t,x) = r\, c\big(t,x\big), \; x \geqq 0,$$

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 12.2. *We defined in Remark 9.7 on p. 148, for $\tau = T - t$, i.e., $t = T - \tau$,*

$$BSM(\tau, x; K, r, \sigma) := c(t, x), \quad \text{where } c(t,x) = x\,N\big(d_+(\tau, x)\big) - e^{-r\tau} K\,N\big(d_-(\tau, x)\big). \tag{12.34}$$

If we redefine BSM $(\tau, x; K, r, \sigma)$ to be

$$BSM(\tau, x; K, r, \sigma) = \widetilde{E}\left[e^{-r\tau}\left(x \exp\left\{-\sigma\sqrt{\tau}Y + \left(r - \frac{1}{2}\sigma^2\right)\tau\right\} - K\right)^+\right], \tag{12.35}$$

where Y is a standard normal random variable under $\widetilde{P}$, then the following holds true:

$$BSM(\tau, x; K, r, \sigma) = x\,N\big(d_+(\tau, x)\big) - e^{-r\tau} K\,N\big(d_-(\tau, x)\big). \tag{12.36}$$

PROOF: Follows from the preceding Lemmas and the fact that the right–hand side of (12.36) matches the definition of $c(t, x)$ given in (12.34). ■

12.5 Completeness of the Onedimensional Generalized Black–Scholes Model

We have seen in Corollary 12.2 on p.176 that any contingent claim that can be replicated can be priced by means of the risk–neutral valuation formula.

$$V_t = \widetilde{E}\left[e^{-\int_t^T R_u du} V_T \,\middle|\, \mathfrak{F}_t\right], \quad 0 \leqq t \leqq T. \tag{12.37}$$

The question that has not been aswered is the following. What claims can be hedged? We will explore that in this cnapter.

We assume that we operate in a generalized Black–Scholes market as was defined in Definition 12.1 on p.170, in particular, that the market price of risk process Θ_t is such that the integrability condition (12.1) given in that definition is satisfied and thus Girsanov's Theorem can be applied.

Assumption 12.3. 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,

$$\mathfrak{F}_t = \mathfrak{F}_t^W = \sigma\{W_u : u \leqq t\} \text{ for all } 0 \leqq t \leqq T,$$
$$\mathfrak{F} = \mathfrak{F}_T^W.$$

We have the following result. See SCF2, ch.5.3.2 (Hedging with One Stock).

Theorem 12.3 (Completeness of the onedimensional Generalized Black–Scholes market)**.** *Given the additional assumptions 12.3, we have the following.*

The onedimensional Generalized Black–Scholes market is complete, i.e., every contingent claim can be hedged. Further, the quantity Δ_t of the replicating portfolio is given for any $0 \leqq t \leqq T$ by either of

$$\Delta_t \sigma_t D_t S_t = \widetilde{\Gamma}_t, \tag{12.38}$$

$$\Delta_t = \frac{\widetilde{\Gamma}_t}{\sigma_t D_t S_t}. \tag{12.39}$$

Here the process $\widetilde{\Gamma}_t$ is implicitly defined by the equation

$$D_t V_t = V_0 + \int_0^t \widetilde{\Gamma}_u \, d\widetilde{W}_u, \; 0 \leqq t \leqq T, \tag{12.40}$$

$$\text{i.e., } \; d(D_t V_t) = \widetilde{\Gamma}_t \, d\widetilde{W}_t \; 0 \leqq t \leqq T. \tag{12.41}$$

PROOF: We create the hedge $\vec{H}_t$ by first looking at the pricing function V_t of the claim $\mathcal{X} = V_T$ that the value process X_t of $\vec{H}_t$ must replicate for each t We then deduce from that the quantity Δ_t of the underlying risky asset (and thus the bank account holdings $X_t - S_t\Delta_t$) for $\vec{H}_t$.

Since $\vec{H}$ replicates $\mathcal{X}$, the pricing principle mandates $X_t = V_t$ for all t. From risk–neutral validation (12.37) we obtain

$$V_t = \widetilde{E}\left[e^{-\int_t^T R_u du} V_T \,\middle|\, \mathfrak{F}_t \right], \quad 0 \leqq t \leqq T. \tag{12.42}$$

Since $V_t = X_t$, $D_t V_t = D_t X_t$. This plus the other risk–neutral validation formula which expresses the fact that the discounted portfolio value $D_t X_t$ is a $\widetilde{P}$–martingale yields

$$D_t V_t = \widetilde{E}\left[D_T V_T \,\middle|\, \mathfrak{F}_t \right], \quad 0 \leqq t \leqq T. \tag{12.43}$$

It now follows from Corollary 11.1 (p.166) to the martingale representation theorem in one dimension that there exists an $\mathfrak{F}_t^W$–adapted process $\widetilde{\Gamma}_u, 0 \leqq u \leqq T$, such that (12.40) holds. Here we made use of the fact that

$$D_0 = e^{-\int_0^0 R_u du} = e^0 = 1, \qquad \text{hence,} \quad D_0 V_0 = V_0.$$

We compare (12.41) to formula (12.16) on p.175 for the differential of $D_t V_t$,

$$d(D_t X_t) = \Delta_t \sigma_t D_t S_t \, d\widetilde{W}_t.$$

Since $\sigma_t D_t S_t > 0$ as the product of three strictly positive quantities, we obtain the desired quantity Δ_t for the number of shares of a hedge $\vec{H}$ for our claim if it is chosen by either of (12.38) or (12.39). ■

Remark 12.5. Note that the formulas for Δ_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. □

12.6 Multidimensional Financial Market Models

Assumption 12.4. 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)

$$\vec{\mathscr{A}} \;=\; \big(\mathscr{A}^{(0)}, \mathscr{A}^{(1)}, \dots, \mathscr{A}^{(n)}\big),$$

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

$$dS_t^{(i)} \;=\; \alpha_t^{(i)} S_t^{(i)}\, dt \;+\; S_t^{(i)} \sum_{j=1}^{d} \sigma_{ij}(t)\, dW_t^{(j)}, \quad i = 1, \dots, m, \tag{12.44}$$

and that we have the usual discount process which is based on an adapted interest rate process R_t.

$$dD_t \;=\; -R_t D_t\, dt, D_0 \;=\; 1, \quad \text{i.e.,} \quad D_t \;=\; \exp\left(-\int_0^t R_u\, du\right). \tag{12.45}$$

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, 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

$$\sigma_t^{(i)} \;:=\; \sqrt{\sum_{j=1}^{d} \sigma_{ij}(t)}, \quad i = 1, \dots, m. \tag{12.46}$$

$$B_t^{(i)} \;:=\; \sum_{j=1}^{d} \int_0^d \frac{\sigma_{ij}(u)}{\sigma_u^{(i)}}\, dW_u^{(j)}, \quad i = 1, \dots, m. \tag{12.47}$$

$$\rho_{ik}(t) \;:=\; \frac{1}{\sigma_t^{(i)} \sigma_t^{(k)}} \sum_{j=1}^{d} \sigma_{ij}(t) \sigma_{kj}(t). \quad i, k = 1, \dots, m. \tag{12.48}$$

We also assume that $\sigma_t^{(i)} > 0$ for all t. ☐

We have the following result.

Proposition 12.4. ⋆ *Each process $B_t(i)$ is a Brownian motion. The multiplication table is*

$$dB_t^{(i)}\, dB_t^{(i)} \;=\; \sum_{j=1}^{d} \frac{\sigma_{ij}^2(t)}{(\sigma_t^{(i)})^2}\, dt \;=\; dt, \quad i = 1, \dots, m, \tag{12.49}$$

$$dB_t^{(i)}\, dB_t^{(k)} \;=\rho_{ik}(t)\, dt \quad i, k = 1, \dots, m,\ i \neq k, \tag{12.50}$$

and the covariances are

$$Cov\big[B_t^{(i)}B_t^{(k)}\big] \;=E\int_0^t \rho_{ik}(u)\,du. \tag{12.51}$$

PROOF: See Chapter 5.4.2 (Multidimensional Market Model) in SCF2. ■

Corollary 12.3. [⋆] *Consider the special case that the volatility matrix is constant in t and ω, allowing us to write*

$$\sigma_{ij} \;:=\; \sigma_{ij}(t,\omega), \qquad \sigma^{(i)} \;:=\; \sigma_t^{(i)}(\omega).$$

A. *Then the right hand side of* (12.48) *also is constant in t and ω and we can write*

$$\rho_{ik} \;:=\; \rho_{ik}(t) \;=\; \frac{1}{\sigma^{(i)}\sigma^{(k)}}\sum_{j=1}^{d}\sigma_{ij}\sigma_{kj} \;\text{ for } i,k=1,\dots,m.$$

B. *Further* $Cov\big[B_t^{(i)},B_t^{(k)}\big] \;=\; \rho_{ik}\,t$ *and the correlation between* $B_t^{(i)}$ *and* $B_t^{(j)}$ *is* ρ_{ik}.

PROOF: [⋆] Trivial. as far as constancy of ρ_{ik} and the equation $Cov\big[B_t^{(i)},B_t^{(k)}\big] \;=\; \rho_{ik}\cdot t$ are concerned The assertion about the correlation follows from

$$\mathrm{Var}\big[B_t^{(i)}\big] \;=\; t \quad \text{for all } i=1,\dots,m. \;■$$

Now some terminology.

Definition 12.5. [⋆] When the volatility matrix is **not** constant in t and ω then we call $\rho_{ik}(t) = \rho_{ik}(t,\omega)$ the **instantaneous correlation** between $B_t^{(i)}$ and $B_t^{(k)}$.

Proposition 12.5. [⋆] *Given the dynamics* (12.44) *for $\vec{S}_t$ and* (12.45) *for D_t, the discounted stock price vector $D_t\vec{S}_t$ has dynamics*

$$d\left(D_tS_t^{(i)}\right) \;= D_tS_t^{(i)}\left[\,\left(\alpha_t^{(i)}-R_t\right)dt \,+\, \sum_{j=1}^{d}\sigma_{ij}(t)\,dW_t^{(j)}\,\right]. \tag{12.52}$$

PROOF: See Chapter 5.4.2 (Multidimensional Market Model) in SCF2. ■

We must generalize the definition of risk–neutral measure given in Definition 12.3 on p.172 for a financial market with a single risky asset price driven by a single Brownian motion to the multidimensional model.

Definition 12.6 (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 12.4 on p.183 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 $D_tS_t^{(i)}$ is a $\widetilde{P}$–martingale w.r.t. the filtration $\mathfrak{F}_t$ for **ALL** $i=1,\dots,m$. □

The following is SCF2 Lemma 5.4.5.

Proposition 12.6. ⋆ *Let $\vec{H}_t$ be a self–financing portfolio with price process X_t. If a risk–neutral measure $\widetilde{P}$ exists in the model then the discounted portfolio value $D_t X_t$ is an $\mathfrak{F}_t$–martingale under $\widetilde{P}$.*

PROOF: See SCF2, before the statement of Lemma 5.4.5. ■

Remark 12.6. We state here for the reader's convenience the definition 6.10 of an arbitrage portfolio. on p.93 in SCF2 notation.
A portfolio $\vec{H}_t$ is an arbitrage portfolio if its value process X_t satisfies

$$X_0 = 0, \tag{12.53}$$
$$P\{X_T \geqq 0\} = 1, \tag{12.54}$$
$$P\{T > 0\} > 0. \;\square \tag{12.55}$$

Here is how we define the vector valued version of a market price of risk process.

Definition 12.7.

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 solves the system of equations, called the **market price of risk equations**,

$$\alpha_i(t) - R_t = \sum_{j=1}^{d} \sigma_{ij}(t)\Theta_j(t), \quad i = 1, \dots, m, \tag{12.56}$$

and which satisfies the Girsanov integrability condition (formula (11.20) on p.167). □

Remark 12.7. 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 thus are not able to guarantee that there are no possibilities for arbitrage. For this see the first fundamental theorem of asset pricing below (Theorem 12.5 on p.186). □

Theorem 12.4.

If a solution to the market price of risk equations

$$\alpha_i(t) - R_t = \sum_{j=1}^{d} \sigma_{ij}(t)\Theta_j(t), \quad i = 1, \dots, m,$$

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 11.3 (Girsanov's Theorem in multiple dimensions) on p.167. We recall that the process $\vec{\widetilde{W}}_t = (\widetilde{W}_t^1, \dots, \widetilde{W}_t^d)$ with dynamics

$$d\widetilde{W}_t^{(j)} = dW_t^{(j)} + \Theta_t^{(j)}\,dt, \qquad \widetilde{W}_0^{(j)} = 0, \tag{12.57}$$

is a d–dimensional $\mathfrak{F}_t$–Brownian motion under the probability $\widetilde{P}$. We plug the market price of risk equations into formula (12.52) on p.184: and obtain

$$\begin{aligned}
d\left(D_t S_t^{(i)}\right) &= D_t S_t^{(i)} \left[\sum_{j=1}^{d} \sigma_{ij}(t)\Theta_j(t)\,dt + \sum_{j=1}^{d} \sigma_{ij}(t)\,dW_t^{(j)} \right] \\
&= D_t S_t^{(i)} \sum_{j=1}^{d} \sigma_{ij}(t)\left[\Theta_j(t)\,dt + dW_t^{(j)}\right] \\
&\overset{(12.57)}{=} D_t S_t^{(i)} \sum_{j=1}^{d} \sigma_{ij}(t) d\widetilde{W}_t^{(j)}].
\end{aligned}$$

Since each $\widetilde{W}_t^{(j)}$ is a $\widetilde{P}$–martingale so is each discounted stock price $D_t S_t^{(i)}$. ∎

Next comes SCF2 Theorem 5.4.7.

Theorem 12.5.

First fundamental theorem of asset pricing:
If the market model given in Assumption 12.4 on p.183 has a risk–neutral probability measure, then it does not admit arbitrage.

PROOF: ⋆ Let $\widetilde{P}$ be a risk–neutral measure and assume we have aelf–financing portfolio $\vec{H}$ with initial value $X_0 = 0$. Since $D_t VT$ is a $\widetilde{P}$–martingale and thus has constant expectation across all times $0 \leqq t \leqq T$ and $D_0 = e^{-\int_0^0 R_u du} = e^0 = 1$ we have

$$\widetilde{E}[D_T X_T] = \widetilde{E}[D_0 X_0] = X_0 = 0. \tag{12.58}$$

Assume further that $\vec{H}$ satisfies condition (12.54),

$$P\{X_T \geqq 0\} = 1. \quad \text{Then} \quad P\{X_T < 0\} = 0, \quad \text{thus} \quad \widetilde{P}\{X_T < 0\} = 0. \tag{12.59}$$

If we can show that it is impossible for $\vec{H}$ to satisfy (12.55): $P\{X_T > 0\} > 0$, then we are done since this means that no self–financing portfolio can satisfy all three conditions (12.53) (12.54), (12.55) of an arbitrage portfolio. So

(A) let us assume to the contrary that $P\{X_T > 0\} > 0$.

Since $P \sim \widetilde{P}$ and thus both probabilities assign zero to the same events, we also have $\widetilde{P}\{X_T > 0\} > 0$. Moreover $\{X_T > 0\} = \{D_T X_T > 0\}$ because $D_T(\omega)$ is strictly positive for all ω as an exponential. Let $A_j := \{D_T X_T \geqq \frac{1}{j}\}$ and $A := \{D_T X_T > 0\}$. If we write $2a$ for $\widetilde{P}(A)$ then $a > 0$. Since

$$A = \bigcup_{j \in \mathbb{N}} A_j \qquad \text{and thus, by (4.27b) on p.47,} \qquad \widetilde{P}(A_j) \uparrow 2a,$$

there is some index j_0 such that $\widetilde{P}(A_{j_0}) \geqq a$. We have

$$0 \overset{(12.58)}{=} \widetilde{E}[D_T X_T] = \int_\Omega D_T X_T \, d\widetilde{P} = \int_A D_T X_T \, d\widetilde{P} + \int_{\{D_T X_T = 0\}} D_T X_T \, d\widetilde{P} + \int_{\{D_T X_T < 0\}} D_T X_T \, d\widetilde{P}.$$

The second integral of the right hand expression is zero because the integrand vanishes on $\{D_T X_T = 0\}$. The third integral of the right hand expression is zero any integral over a set of measure zero is zero. This follows from Proposition 4.13 on p.70. Hence,

$$\int_A D_T X_T \, d\widetilde{P} = 0.$$

Since $A_{j_0} \subset A$ and $D_T X_T > 0$ on A,

$$0 = \int_A D_T X_T \, d\widetilde{P} \geqq \int_{A_{j_0}} D_T X_T \, d\widetilde{P} \geqq \int_{A_{j_0}} \frac{1}{j_0} \, d\widetilde{P} = \frac{1}{j_0} \widetilde{P}(A_{j_0}) \geqq \frac{a}{j_0} > 0.$$

Thus assumption (**A**) has lead us to the contradiction $0 > 0$. This proves that $P\{X_T > 0\} > 0$ and 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 12.8. 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

$$\widetilde{P}(...) = 0 \Leftrightarrow P(...) = 0,$$
$$\widetilde{P}(...) > 0 \Leftrightarrow P(...) > 0,$$
$$\widetilde{P}(...) = 1 \Leftrightarrow P(...) = 1.$$

Theorem 12.3 (Completeness of the onedimensional Generalized Black–Scholes market) in Subchapter 12.5 (Completeness of the Onedimensional Generalized Black–Scholes Model) gave conditions under which the onedimensional market is complete, i.e., every contingent claim that is reasonably integrable can be hedged. See Definition 6.12 (Hedging/Replicating Portfolio) on p.94. We now want to examine under which conditions the multidimensional market is complete.

Assumption 12.5. We add to Assumption 12.4 the following conditions.

(1) The market price of risk equations of Definition 12.7 on p.185,

$$\alpha_i(t) \,-\, R_t \;=\; \sum_{j=1}^{d} \sigma_{ij}(t)\Theta_j(t), \quad i = 1, \dots, m,$$

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., our filtration is generated by the d–dimensional Brownian motion $\vec{W}_t$. $\square$

Remark 12.9. The first of the above conditions implies that the conditions of Theorem 12.4 on p.185 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

$$M_t \;=\; M_0 \,+\, \sum_{j=1}^{d} \int_0^t \widetilde{\Gamma_u}\, d\widetilde{W}_u.$$

Here the process $\widetilde{W}_t$ is the $\widetilde{P}$–d–dimensional Brownian motion

$$\widetilde{W}_t \;=\; W_t \,+\, \int_0^t \Theta_u\, du.; \;\square$$

Theorem 12.6 (Completeness of the multidimensional market)**.**

This item has been removed!

The next theorem is SCF2 Theorem 5.4.9.

Theorem 12.7.

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! ■

12.7 Exercises for Ch.12

Exercise 12.1. Prove the formula (12.9) of Proposition 12.1 on p.173:

$$d\, D_t S_t \;= \big(\alpha_t - R_t\big) D_t S_t \, dt \,+\, \sigma_t D_t S_t \, dW_t$$

directly from the dynamics given in Definition 12.1 on p.170,

$$
\begin{aligned}
dD_t &= -R_t D_t\,dt,\\
dS_t &= \alpha_t S_t\,dt + \sigma_t S_t\,dW_t,
\end{aligned}
$$

by applying the Itô product rule to $dD_t S_t$. $\square$

Exercise 12.2. Prove the "⇒" direction of Theorem 12.7 (Second fundamental theorem of asset pricing) on p.188 of this document: If the multidimensional market is complete then the risk–neutral probability measure is unique. $\square$

Exercise 12.3.

This item has been removed!

13 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 provide insight into modeling aspects and they also are a good approximation for a mutual fund which holds many different kinds of stocks which all pay their dividends at different dates.

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) $A_t S_t dt$ is paid per share during $[t, t+dt]$ and we must subtract $A_t S_t dt$ 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 onedimensional case: A single (onedimensional) Brownian motion which drives a single underlying risky asset (stock).

We try to use SCF2 notation whenever feasible.

It will be shown in Proposition 13.2 on p.192 that the probability measure $\widetilde{P}$ which is constructed in Girsanov's Theorem by means of the market price of risk process Θ_t is no longer a risk–neutral measure as was defined in Definition 12.3 on p.172 since it does not make discounted stock price $D_t S_t$ a martingale.

It turns out though that discounted portfolio value $D_t X_t$ 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 $\widetilde{P}$.

13.1 Continuously Paying Dividends

Assumption 13.1. Unless stated otherwise we assume that we have a generalized Black–Scholes market as defined in Definition 12.1 (Generalized Black–Scholes market model) on p.170, with the following **modification**.

We assume that the risky asset 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 non–negative We noted in the introduction to this chapter that this will affect the stock price dynamics and we replace formula (12.3) with the following.

$$dS_t = \alpha_t S_t\,dt + \sigma_t S_t\,dW_t - A_t S_t\,dt; \quad S_0 \in]0,\infty[;\ \alpha_t, \sigma_t \in]0,\infty[; \tag{13.1}$$

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 $\widetilde{P}$. □

We thus have

$$dD_t = -R_t D_t\,dt;\ D_0 = 1, \tag{13.2}$$

$$\Theta_t = \frac{\alpha_t - R_t}{\sigma_t}, \tag{13.3}$$

$$d\widetilde{W}_t = dW_t + \Theta_t\,dt;\ \widetilde{W}_0 = 0. \ \square \tag{13.4}$$

Proposition 13.1. *The value and discounted value of a self–financing portfolio have the following dynamics.*

$$dX_t = R_t X_t\,dt + \Delta_t S_t \sigma_t\big(\Theta_t\,dt + dW_t\big) = R_t X_t\,dt + \Delta_t S_t \sigma_t\,d\widetilde{W}_t, \tag{13.5}$$

$$d\big(D_t X_t\big) = \Delta_t D_t S_t \sigma_t\,d\widetilde{W}_t. \tag{13.6}$$

In particular, the discounted portfolio process $D_t X_t$ is a $\widetilde{P}$–martingale.

For the proof see SCF2 ch.5.5.1. ■

Remark 13.1. A. Discounted portfolio value being a $\widetilde{P}$–martingale is all it takes to use risk–neutral valuation for contingent claims. Let $\vec{H}$ with portfolio value X_t be a hedge for a contingent claim $\mathcal{X} = V_T$ with pricing process $V_t = \Pi_t(\mathcal{X})$. Then $X_T = V_T$, thus $D_T V_T = D_T X_T$ and, according to the pricing principle, $V_t = X_t$ for all $0 \leqq t \leqq T$. Moreover, since $D_t X_t$ is an $\mathfrak{F}_t$–martingale under $\widetilde{P}$,

$$D_t V_t = D_t X_t = \widetilde{E}\big[D_T V_T \mid \mathfrak{F}_t\big] \quad \text{for } 0 \leqq t \leqq T,$$

$$\text{thus} \quad V_t = \widetilde{E}\big[D_t^{-1} D_T V_T \mid \mathfrak{F}_t\big] = \widetilde{E}\big[e^{-\int_t^T R_u du} V_T \mid \mathfrak{F}_t\big] \quad \text{for } 0 \leqq t \leqq T.$$

B. Note that formula (13.5) for dX_t matches formula 12.15 on p,174, and Note that formula (13.6) for $d(D_t X_t)$ matches formula 12.16 on p,175.

C. A closer inspection of the proof of Theorem 12.3 (Completeness of the onedimensional Generalized Black–Scholes market) on p.181 shows that it only depends on risk–neutral valuation and the items mentioned in points **A** and **B** of this remark. We thus obtain the next theorem in the case of a stock with a continuously paying dividend. □

Theorem 13.1. *Given the assumptions 12.3 on p.181 in addition to the assumptions 13.1 made at the beginning of this chapter we have the following.*

The onedimensional Generalized Black–Scholes market with continuous dividend payments is complete, i.e., every contingent claim can be hedged. Further, the quantity Δ_t of the replicating portfolio is given for any $0 \leqq t \leqq T$ by either of

$$\Delta_t \sigma_t D_t S_t = \widetilde{\Gamma}_t, \tag{13.7}$$

$$\Delta_t = \frac{\widetilde{\Gamma}_t}{\sigma_t D_t S_t}. \tag{13.8}$$

Here the process $\widetilde{\Gamma}_t$ is implicitly defined by the equation

$$D_t V_t = V_0 + \int_0^t \widetilde{\Gamma}_u \, d\widetilde{W}_u, \ 0 \leqq t \leqq T, \tag{13.9}$$

$$\text{i.e.,} \quad d(D_t V_t) = \widetilde{\Gamma}_t \, d\widetilde{W}_t \ 0 \leqq t \leqq T. \tag{13.10}$$

PROOF: ⋆ We can copy the proof of Theorem 12.3 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 that the discounted value of a self–financing portfolio is a $\widetilde{P}$–martingale. The next proposition shows that this is no more true for discounted stock price.

Proposition 13.2. ⋆ *If $A_t \neq 0$ then the process $D_t S_t$ is not a $\widetilde{P}$–martingale. Instead the process $e^{\int_0^t A_u du} D_t S_t$ is a $\widetilde{P}$–martingale. That process has explicit representation*

$$e^{\int_0^t A_s ds} D_t S_t = \exp\left\{ \int_0^t \sigma_s \, d\widetilde{W}_s - \frac{1}{2} \int_0^t \sigma_s^2 \, ds \right\}.$$

PROOF: See SCF2. ■

13.2 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 risky asset 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.

We need to be able to model continuous time processes that possess a jump at some time t^*.

Definition 13.1. 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 usually us subscripts X_t rather than parenthesized time arguments for stochastic processes $X_t(\omega)$. Then we write X_{t^*-}. □

To take this into account we must modify the assumptions 13.1 of Chapter 13.1 (Continuously Paying Dividends) accordingly.

Assumption 13.2.

(1) Unless stated otherwise we assume that we have a generalized Black–Scholes market as defined in Definition 12.1 (Generalized Black–Scholes market model) on p.170, with the following **modification**.

(2) We assume that the risky asset pays its dividend only at the discrete points in time $0 < t_1 < t_2 < \cdots < t_n < T$. 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 $0 \leqq a_j \leqq 1$ 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.

(3) 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:

$$dS_t = \alpha_t S_t\,dt + \sigma_t S_t\,dW_t, \quad \text{where } \alpha_t, \sigma_t \in]0, \infty[\,; \tag{13.11}$$

$$S_{t_j} = S_{t_j-} - a_{t_j} S_{t_j-}\,. \tag{13.12}$$

(4) 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 $\widetilde{P}$.

We thus have

$$dD_t = -R_t D_t\,dt;\; D_0 = 1, \tag{13.13}$$

$$\Theta_t = \frac{\alpha_t - R_t}{\sigma_t}, \tag{13.14}$$

$$d\widetilde{W}_t = dW_t + \Theta_t\,dt;\; \widetilde{W}_0 = 0.\ \square \tag{13.15}$$

Remark 13.2.

(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

$$S(t_j) = S(t_j-) - a_j S(t_j-) = (1 - a_j) S(t_j-). \tag{13.16}$$

(2) If $a_j = 0$, no dividend is paid and $S_{t_j} = S_{t_j-}$.

(3) If $a_j = 1$, the full value of the asset is paid and $S_t = 0$ for all $t \geqq t_j$. $\square$

Proposition 13.3. *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 13.1 on p.191*

$$dX_t = R_t X_t\,dt + \Delta_t S_t \sigma_t \big(\Theta_t\,dt + dW_t\big) = R_t X_t\,dt + \Delta_t S_t \sigma_t\,d\widetilde{W}_t, \tag{13.17}$$

$$d\big(D_t X_t\big) = \Delta_t D_t S_t \sigma_t\,d\widetilde{W}_t. \tag{13.18}$$

In particular, the discounted portfolio process D_tX_t is a $\widetilde{P}$–martingale and risk–neutral validation still applies:

$$\begin{aligned} D_tV_t &= D_tX_t = \widetilde{E}\big[D_TV_T \mid \mathfrak{F}_t\big] \quad \textit{for } 0 \leqq t \leqq T, \\ \textit{thus} \quad V_t &= \widetilde{E}\big[D_t^{-1}D_TV_T \mid \mathfrak{F}_t\big] = \widetilde{E}\big[e^{-\int_t^T R_u du}V_T \mid \mathfrak{F}_t\big] \quad \textit{for } 0 \leqq t \leqq T. \end{aligned}$$

PROOF: ⋆ For the proof see SCF2 ch.5.5.2. ■

13.3 Constant Dividend Rates

First the continuous time case.

Assumption 13.3. 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 9 (Black–Scholes Model Part I: The PDE). □

In the case of no divdidends we had seen in Subchapter 9.4 (The Black–Scholes PDE for a European Call) that the pricing function of a European call is

$$c(t,x) = xN\big(d_+(T-t,x)\big) - Ke^{-r(T-t)}N\big(d_-(T-t,x)\big), \quad 0 \leqq t < T,\ x > 0, \tag{13.19}$$

where

$$d_\pm(\tau,x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right], \tag{13.20}$$

Here is the main result in the case of continuous and constant dividend payments with rate a.

Proposition 13.4. *Under the assumptions 13.3 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:*

$$c(t,x) = xe^{-a\tau}N\big(d_+(T-t,x)\big) - Ke^{-r(T-t)}N\big(d_-(T-t,x)\big), \tag{13.21}$$

for $0 \leqq t < T$, and $x > 0$ Here we define

$$d_\pm(\tau,x) = \frac{1}{\sigma\sqrt{\tau}}\left[\log\frac{x}{K} + \left(r - a \pm \frac{\sigma^2}{2}\right)\tau\right]. \tag{13.22}$$

As usual N is the cumulative standard normal distribution

$$N(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{z^2}{2}}\,dz = \frac{1}{\sqrt{2\pi}}\int_{-y}^{\infty} e^{-\frac{z^2}{2}}\,dz. \tag{13.23}$$

For the proof see SCF2 ch.5.5.1. ■

Now we switch to discrete time dividend payments.

Assumption 13.4. We replace the assumptions 13.3 with the following.

We assume that the processes $r := R_t$, $\alpha := \alpha_t$, $\sigma := \sigma_t$ are constant in t and ω, i.e., we have a classical Black–Scholes market as in Chapter 9 (Black–Scholes Model Part I: The PDE).

In addition we now also have finite list of discrete time dividend rates a_j as wa had defined in the assumptions 13.2 of Subchapter 13.2 (Dividends Paid at Discrete Times) except that

We assume that those rates a_j are deterministic.

Proposition 13.5. *Under the assumptions 13.3 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:*

$$c(t,x) = x \prod_{j=0}^{n-1} N\big(d_+(\tau,x)\big) - Ke^{-r\tau} N\big(d_-(\tau,x)\big), \tag{13.24}$$

for $0 \leqq t < T$, and $x > 0$ Here we define $\tau := T - t$ and

$$d_\pm(\tau,x) = \frac{1}{\sigma\sqrt{\tau}} \left[\log\frac{x}{K} + \left(r - a \pm \frac{\sigma^2}{2} \right) \tau \right]. \tag{13.25}$$

As usual N is the cumulative standard normal distribution

$$N(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{z^2}{2}}\, dz = \frac{1}{\sqrt{2\pi}} \int_{-y}^{\infty} e^{-\frac{z^2}{2}}\, dz. \tag{13.26}$$

For the proof see SCF2 ch.5.5.1. ■

Remark 13.3. 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 □

13.4 Forward Contracts and Zero Coupon Bonds

We now assume that a dividend is **NOT paid** for the risky asset, thus discounted stock price $D_t S_t$ is a martingale under the Girsanov measure $\widetilde{P}$ and $\widetilde{P}$ is a genuine risk–neutral measure.

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 13.2.

- A **zero–coupon bond** is a contingent claim with contract value $V_T = 1$ at time T. We call T the **maturity date** of the zero–coupon bond.
- We denote the price of such a zero–coupon bond at time $0 \leqq t \leqq T$ by $B(t,T)$. $\square$

Proposition 13.6. *If $\widetilde{P}$ is a risk–neutral probability then*

$$B(t,T) := \frac{1}{D_t}\widetilde{E}[D_T \mid \mathfrak{F}_t], \text{ for } 0 \leqq t \leqq T \leqq \bar{T}. \tag{13.27}$$

PROOF: This is risk–neutral validation applied to a contingent claim with constant value 1 at time T. $\blacksquare$

The following is SCF2, Theorem 5.6.2.

Theorem 13.2. $\boxed{\star}$

Let $T > 0$. Assume that there is unlimited liquidity in the market for zero–coupon bonds with maturity date $0 \leqq T' \leqq T$. Let $\mathcal{X}$ be a forward contract with expiration date T for an underlying asset with price S_t. Then the following holds, regardless of the strike price of that contract.

The forward price For_t at time t (see Definition 9.4 on p.151) is

$$For_S(t,T) = \frac{S_t}{B(t,T)}, \quad 0 \leqq t \leqq T \leqq \bar{T}. \tag{13.28}$$

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. Let K denote the strike price of that contract. Then its value at time T is $V_T = S_T - K$, thus

$$\begin{aligned} V_t &= \frac{1}{D_t}\widetilde{E}[D_T\big(S_T - K\big) \,|\, \mathfrak{F}_t] \\ &= \frac{1}{D_t}\widetilde{E}[D_T S_T \,|\, \mathfrak{F}_t] - \frac{K}{D_t}\widetilde{E}[D_T \,|\, \mathfrak{F}_t]. \end{aligned} \tag{A}$$

Note that $D_t S_t$ is a martingale under risk–neutral probability $\widetilde{P}$ and so is $D_t V'_t$ if V'_t is the pricing function of a claim with contract value $V'_T = 1$, i.e., of a zero–coupon bond with maturity T. Note that $D_T = D_T \cdot 1 = D_T V'_T$ and that $V'_t = B(t,T)$ by the very definition of $B(t,T)$. It follows from **(A)** that

$$V_t = \frac{1}{D_t} D_t S_t - \frac{K}{D_t} D_t\, B(t,T) = S_t - K\, B(t,T).$$

The forward price $\text{For}_S(t,T)$ 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 V_1 of that contract at time t. Thus

$$0 = S_t - \text{For}_S(t,T)\, B(t,T),$$

and we have obtained (13.28). $\blacksquare$

13.5 Exercises for Ch.13

Exercise 13.1. Theorem 13.2 on p.196 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. □

14 Stochastic Methods for Partial Differential Equations

Many if not most stocks pay a dividend per share at discrete times, say,

14.1 Stochastic Differential Equations

Definition 14.1 (Stochastic differential equation)**.** Let $W_t, t \geqq 0$, be a Brownian motion on a filtered probability space $(\Omega, \mathfrak{F}, \mathfrak{F}_t, P)$ and let

$$\beta, \gamma : [0,T] \to \mathbb{R}, \qquad (t,x) \mapsto \beta(t,x), \gamma : [0,T] \to \mathbb{R}, \qquad (t,x) \mapsto \gamma(t,x),$$

be two (measurable) deterministic functions.

We call a stochastic differential plus family of initial conditions,

$$dX_t = \beta\big(t, X_t\big)\, dt + \gamma\big(t, X_t\big)\, dW_t, \tag{14.1}$$

$$X_{t_0} = x_0, \quad \text{for all } 0 \leqq t_0 \leqq t \leqq T \text{ and } x_0 \in \mathbb{R}, \tag{14.2}$$

a **stochastic differential equation** with **drift coefficient** β**, diffusion coefficient** γ, , and **initial conditions**, , (14.2). This can be referred to more compactly as an **SDE with drift** β**, diffusion** γ**, and initial conditions** (14.2).

We say that the SDE with stochastic differential (14.1) and initial conditions (14.2) **has a solution**, if for EACH (u, a) such that $0 \leqq u \leqq T$ and $a \in \mathbb{R}$ there is a stochastic process $X^{u,a} = X_t^{u,a}(\omega)$ with dynamics and initial condition given by

$$dX_t^{u,a} = \beta\big(t, X_t^{u,a}\big)\, dt + \gamma\big(t, X_t^{u,a}\big)\, dW_t, \tag{14.3}$$

$$X_u^{u,a} = a, \quad . \tag{14.4}$$

Since each statement $X_u^{u,a} = a$ uniquely determines a pair $(t,a) \in [0,T] \times \mathbb{R}$ and vice versa, it is convenient to refer to the **initial condition** (t,a). ☐

We will see as part of the next Facts collection that, all processes $X^{u,a}$ are the same and that we can discard the superscripts.

Fact 14.1. ⋆ *Solutions of an SDE have the following properties.*

(1) *The SDE of Definition 14.1 possesses a solution under very general conditions on drift $\beta(t,x)$ and diffusion $\gamma(t,x)$ must satisfy.*

(2) *This solution is described for all initial conditions $X_u = a$, i.e., for all $0 \leqq u \leqq T$ and $a \in \mathbb{R}$, by one and the same process $(t,\omega) \mapsto X_t(\omega)$. In other words, all processes $X_t^{u,a}$ coincide and thus we can and will drop the superscript and write X_t instead of $X_t^{u,a}$.*

(3) *For the following review Remark 5.4 (Factored conditional expectation) on p.85. For any initial condition*

We can associate with every initial condition (u, a) a probability measure $P^{u,a}$ on the codomain $\mathbb{R}$ of the real valued process X_t which acts like the conditional probability

$$P^{u,a}\{X_t \in B\} \;=\; P\{X_t \in B \mid X_u = a\} \tag{14.5}$$

(4) *We can express this in terms of the corresponding expectation $E^{u,a} = \int \dots dP^{u,a}$:*

$$E^{u,a}\{h(X_t)\} \;=\; E\{h(X_t) \mid X_u = a\} \tag{14.6}$$

This formula remains valid if we replace a with $X_u(\omega)$:

$$E^{u,X_u(\omega)}\{h(X_t)\} \;=\; E\{h(X_t) \mid X_u = X_u(\omega)\}. \;=\; E\{h(X_t) \mid X_u\}(\omega).$$

We finally drop the argument ω and obtain

$$E^{u,X_u}\{h(X_t)\} \;=\; E\{h(X_t) \mid X_u\}.$$

Now formula (14.6) asserts that

$$E^{u,X_u}\{h(X_t)\} \;=\; E\{h(X_t) \mid X_u\}. \;\square \tag{14.7}$$

The following is SCF2 Theorem 6.3.1.

Theorem 14.1. *The original expectation $E[\dots]$ of $(\Omega, \mathfrak{T}, \mathfrak{T}_t, P)$ is intimately related to the expectations $E^{u,a}[\dots]$ belonging to the initial conditions (u, a) by means of conditioning:*

$$E^{u,X_u}\{h(X_t)\} \;=\; E\{h(X_t) \mid X_u\} \;=\; E\{h(X_t) \mid \mathfrak{F}_u\}. \tag{14.8}$$

PROOF: ⋆ The first equation is a repetition of (14.7).

The solution X_t is a Markov process, i.e., conditioning on the present information $\sigma(X_u)$ is the same as conditioning on the entire past $\mathfrak{F}_u$. This proves the second equation. ■

The following is SCF2 Theorem 6.4.1.

Theorem 14.2 (Feynman–Kac Theorem)**.**

Let $T > 0$. We examine again the SDE with differential (14.1) and initial conditions (14.2),

$$d\,X_t \;:=\beta\big(t,X_t\big)\,dt \;+\;\gamma\big(t,X_t\big)\,dW_t; \quad X_{t_0} = x_0 \;(0 \leqq t_0 < T,\; x_0 \in \mathbb{R}). \tag{14.9}$$

Let $x \mapsto \Phi(x)$ be Borel–measurable such that $E^{t,x}[\Phi\big(X_T\big)] < \infty$, for all $0 \leqq t \leqq T$ and $x \in \mathbb{R}$. Let $(t,x) \mapsto f(t,x)$ be the function

$$f(t,x) \;:=\; E^{t,x}[\Phi\big(X_T\big)] \tag{14.10}$$

Then $f(t,x)$ is a solution to the following PDE plus terminal condition

$$f_t(t,x) \;+\; \beta(t,x)f_x(t,x) \;+\; \frac{1}{2}\gamma^2(t,x)f_{xx}(t,x) \;=0 \tag{14.11}$$

$$f(T,x) \;=\; \Phi(x) \;\text{ for all } x. \tag{14.12}$$

You can find an outline of the proof in the SCF2 text. ■

The following is SCF2 Theorem 6.4.3.

Theorem 14.3 (Discounted Feynman–Kac)**.**

Let $T > 0$. We examine again the SDE with differential (14.1) and initial conditions (14.2),

$$d\,X_t \;:=\beta\big(t,X_t\big)\,dt \;+\;\gamma\big(t,X_t\big)\,dW_t; \quad X_{t_0} = x_0 \;(0 \leqq t_0 < T,\; x_0 \in \mathbb{R}). \tag{14.13}$$

Let $x \mapsto \Phi(x)$ be Borel–measurable such that $E^{t,x}[\Phi\big(X_T\big)] < \infty$, for all $0 \leqq t \leqq T$ and $x \in \mathbb{R}$. Let $(t,x) \mapsto f(t,x)$ be the function

$$f(t,x) \;:=\; E^{t,x}[e^{-r(T-t)}\Phi\big(X_T\big)] \tag{14.14}$$

Then $f(t,x)$ is a solution to the following PDE plus terminal condition

$$f_t(t,x) \;+\; \beta(t,x)f_x(t,x) \;+\; \frac{1}{2}\gamma^2(t,x)f_{xx}(t,x) - rf(t,x) \;=0, \tag{14.15}$$

$$f(T,x) \;=\; \Phi(x) \;\text{ for all } x. \tag{14.16}$$

You can find an outline of the proof in the SCF2 text. ■

Remark 14.1. 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 of a simple claim at time t if stock price then is $S_t = x$,
- The terminal condition $f(T,x) = \Phi(x)$ will be the contract function of that claim, i.e. $\mathcal{X} = \Phi(S_T)$.

- $f(t,x) = E^{t,x}[e^{-r(T-t)}\Phi(X_T)]$ is guaranteed to be the solution of the PDE $f_t + \beta f_x + \frac{1}{2}\gamma^2 f_{xx} - rf = 0$, but what is it good for if $E[...]$ is not risk neutral measure and $E^{t,S_t}[e^{-r(T-t)}\Phi(X_T)]$ is NOT the arbitrage free price V_t of the option? since the Brownian motion W_t in the dynamics of

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 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

$$f_t(t,x) \;+\; \beta'(t,x)f_x(t,x) \;+\; \frac{1}{2}\gamma'^2(t,x)f_{xx}(t,x) - rf(t,x) \;= 0,$$
$$f(T,x) \;=\; \Phi(x) \;\text{ for all } x.$$

for which the solution $f(t,x)$ does what you wanted: $V_t = f(t,S_t)$.

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. □

14.2 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

$$d\,R_t \;= \beta\big(t,R_t\big)\,dt \;+\; \gamma\big(t,R_t\big)\,d\widetilde{W}_t. \tag{14.17}$$

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 6.7 (Discount process) on p.91 that

$$\text{Int}_t \;= \exp\left\{\int_0^t R_s\,ds\right\}$$

is the money market account price process and

$$D_t \;= \frac{1}{\text{Int}_t} \;= \exp\left\{-\int_0^t R_s\,ds\right\}$$

is the discount process of the bank account.

Clearly the dynamics of those processes are

$$dD_t \;=\; -\,R_t D_t\,dt, \qquad d\text{Int}_t \;= \text{Int}_t\,R_t\,dt.$$

We saw in Chapter 13.4 (Forward Contracts and Zero Coupon Bonds) that a zero–coupon bond with maturity date T is a contingent claim wich constant contract value $V_T = 1$ and that the (arbitrage free) price $B(t,T)$ at time $0 \leqq t \leqq T$ is, under risk–neutral probability $\widetilde{P}$,

$$B(t,T) \;=\; \frac{1}{D_t}\widetilde{E}[D_T \mid \mathfrak{F}_t]. \;=\; \widetilde{E}[e^{-\int_t^T R_s\,ds} \mid \mathfrak{F}_t].$$

Definition 14.2 (Yield)**.** We define the **yield** of zero–coupon bond between times t and T as

$$Y(t,T) \;:=\; -\frac{1}{T-t}\log B(t,T) \tag{14.18}$$

□

Remark 14.2. Formula (14.18) is equivalent to

$$B(t,T) \;=\; e^{-Y(t,T)\,(T-t)}. \tag{14.19}$$

One sees from this formula that $Y(t,T)$ is the constant rate of continuously compounding interest between times t and T that corresponds to the price $B(t,T)$ of a zero–coupon bond maturing at T. □

Proposition 14.1.
Given the dynamics of (14.17) for the interest rate R_t, there is a function $f(t,x$ such that $B(t,T) = f(t,R_t$. This function satisfies the PDE plus terminal condition

$$f_t(t,r) \;+\; \beta(t,r)f_r(t,r) \;+\; \frac{1}{2}\gamma^2(t,r)f_{rr}(t,r) \;=\; rf(t,r), \tag{14.20}$$

$$f(T,r) \;=\; 1 \;\; \textit{for all } r. \tag{14.21}$$

PROOF: See SCF2, Chapter 6.5. ■

14.3 Stochastic Differential Equations and their PDEs in Multiple Dimensions

Theorem 14.4 (Twodimensional Feynman–Kac)**.**
Let $T > 0$. Let $\vec{W}_t = \big(W_1(t), W_2(t)\big)$ be a two-dimensional Brownian motion (i.e., the components $W_1(t), W_2(t)$ are two independent, one–dimensional Brownian motions.

et $\vec{X}_t := (X_1(t), X_2(t))$ be a vector of two Itô processes which satisfy the system of SDEs

$$\begin{aligned} dX_1(s) \;&=\; \beta_1\big(s, X_1(s)\big)\,ds \;+\; \gamma_{11}\big(s, X_1(s), X_2(s)\big)\,dW_1(s) \;+\; \gamma_{12}\big(s, X_1(s), X_2(s)\big)\,dW_2(s),\\ dX_2(s) \;&=\; \beta_2\big(s, X_1(s)\big)\,ds \;+\; \gamma_{21}\big(s, X_1(s), X_2(s)\big)\,dW_1(s) \;+\; \gamma_{22}\big(s, X_1(s), X_2(s)\big)\,dW_2(s). \end{aligned}$$

A. This pair of SDEs has under certain mild conditions on the processes $\beta_i\big(s, X_1(s)\big)$ and $\gamma_{22}\big(s, X_1(s), X_2(s)\big)$ a solution $\vec{X}_t$ starting at $X_1(t) = x_1$ and $X_2(t) = x_2$. Regardless of the initial condition, this solution is a Markov process.

Let a Borel-measurable function $h(y_1, y_2)$ be given. Corresponding to the initial condition t, x_1, x_2, where $0 \leqq t \leqq T$, we define

$$g(t, x_1, x_2) := E^{t,x_1,x_2}\, h\big(X_1(T), X_2(T)\big), \tag{14.22}$$

$$f(t, x_1, x_2) := E^{t,x_1,x_2}\left[e^{-r(T-t)}\, h\big(X_1(T), X_2(T)\big)\right] \tag{14.23}$$

Then

$$\begin{aligned} &g_t + \beta_1 g_{x_1} + \beta_2 g_{x_2} \\ &\quad + \frac{1}{2}(\gamma_{11}^2 + \gamma_{12}^2) g_{x_1x_1} + (\gamma_{11}\gamma_{21} + \gamma_{12}\gamma_{22}) g_{x_1x_2} + \frac{1}{2}(\gamma_{21}^2 + \gamma_{22}^2) g_{x_2x_2} = 0, \end{aligned} \tag{14.24}$$

$$\begin{aligned} &f_t + \beta_1 f_{x_1} + \beta_2 f_{x_2} \\ &\quad + \frac{1}{2}(\gamma_{11}^2 + \gamma_{12}^2) f_{x_1x_1} + (\gamma_{11}\gamma_{21} + \gamma_{12}\gamma_{22}) f_{x_1x_2} + \frac{1}{2}(\gamma_{21}^2 + \gamma_{22}^2) f_{x_2x_2} = rf. \end{aligned} \tag{14.25}$$

Further these PDE solutions $f(t, x_1, x_2)$ and $g(t, x_1, x_2)$ also satisfy the terminal conditions

$$g(T, x_1, x_2) = f(T, x_1, x_2) = h(x_1, x_2) \quad \textit{for all } x_1 \textit{ and } x_2.$$

PROOF: See SCF2 Chapter 6.6 ■

14.4 Exercises for Ch.14

Exercise 14.1. Let T, X_t, $\Phi(x)$, $f(t,x)$ be as defined in Theorem 14.2 (Feynman–Kac Theorem) on p.199. Prove that the process

$$M_t := f(t, X_t) = E^{t,x}[\Phi(X_T)]$$

is a martingale. **Hint**: Use formula (14.8) on p.199. □

15 Other Appendices

15.1 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 $\quad B =$ 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		

15.2 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.

Notations 15.1. **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 $]0, +\infty[$, $\mathbb{R}_{\geqq 0}$ 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 $\mathbb{Z}_{\geqq 0}$ for $\mathbb{N} \biguplus \{0\}$. $\mathbb{Z}_{\geqq 0}$ is the B/G notation. It is very unusual but also very intuitive. ☐

Definition 15.1. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of real numbers. We call that sequence **nondecreasing** or **increasing** if $x_n \leqq x_{n+1}$ for all $n \in \mathbb{N}$.

We call it **strictly increasing** if $x_n < x_{n+1}$ for all $n \in \mathbb{N}$.

We call it **nonincreasing** or **decreasing** if $x_n \geqq x_{n+1}$ for all n.

We call it **strictly decreasing** if $x_n > x_{n+1}$ for all $n \in \mathbb{N}$. ☐

References

[1] Heinz Bauer. Measure and Integration Theory. De Gruyter, Berlin, 1st edition, 2001.

[2] Matthias Beck and Ross Geoghegan. The Art of Proof. Springer, 1st edition, 2010.

[3] Thomas Björk. Arbitrage Theory in Continuous Time. Oxford University Press, 2nd edition, 2004.

[4] Steve E. Shreve. Stochastic Calculus for Finance II: Continuous-Time Models. Springer, 1st edition, 2004.

[5] James Stewart. Single Variable Calculus. Thomson Brooks Cole, 7th edition, 2012.

[6] Dennis Wackerly, William Mendenhall, and Richard L. Scheaffer. Mathematical Statistics with Applications. Thomson Brooks/Cole, 7th edition, 2008.

List of Symbols

Index