VIENNA GRADUATE SCHOOL OF FINANCE (VGSF)

LECTURE NOTES

Introduction to Probability Theory and Stochastic Processes (STATS)

Helmut Strasser

Department of Statistics and Mathematics
Vienna University of Economics and Business Administration
Helmut.Strasser@wu-wien.ac.at

January 25, 2006

Contents

Preliminaries

0.1 Introduction

The goal of this course is to give an introduction into some mathematical concepts and tools which are indispensable for understanding the modern mathematical theory of finance. Let us give an overview of historic origins of some of the mathematical tools.

The central topic will be those probabilistic concepts and results which play an important role in mathematical finance. Therefore we have to deal with mathematical probability theory. Mathematical probability theory is formulated in a language that comes from measure theory and integration. This language differs considerably from the language of classical analysis, known under the label of calculus. Therefore, our first step will be to get an impression of basic measure theory and integration.

We will not go into the advanced problems of measure theory where this theory becomes exciting. Such topics would be closely related to advanced set theory and topology which also differ basically from set theoretic language and topologically driven slang which is convenient for talking about mathematics but nothing more. Similarly, our usage of measure theory and integration is sort of a convenient language which on this level is of little interest in itself. For us its worth arises with its power to give insight into exciting applications like probability and mathematical finance.

Therefore, our presentation of measure theory and integration will be an overview rather than a specialized training program. We will become more and more familiar with the language and its typical kind of reasoning as we go into those applications for which we are highly motivated. These will be probability theory and stochastic calculus.

In the field of probability theory we are interested in probability models having a dynamic structure, i.e. a time evolution governed by endogeneous correlation properties. Such probability models are called stochastic processes.

Probability theory is a young theory compared with the classical cornerstones of mathematics. It is illuminating to have a look at the evolution of some fundamental ideas of defining a dynamic structure of stochastic processes.

One important line of thought is looking at stationarity. Models which are themselves stationary or are cumulatives of stationary models have determined the econometric literature for decades. For Gaussian models one need not distinguish between strict and weak (covariance) stationarity. As for weak stationarity it turns out that typical processes follow difference or differential equations driven by some noise process. The concept of a noise process is motivated by the idea that it does not transport any information.

From the beginning of serious investigation of stochastic processes (about 1900) another idea was leading in the scientific literature, i.e. the Markov property. This is not the place to go into details of the overwhelming progress in Markov chains and processes achieved in the first half of the 20th century. However, for a long time this theory failed to describe the dynamic behaviour of continuous time Markov processes in terms of equations between single states at different times. Such equations have been the common tools for deterministic dynamics (ordinary difference and differential equations) and for discrete time stationary stochastic sequences. In contrast, continuous time Markov processes were defined in terms of the dynamic behaviour of their distributions rather than of their states, using partial difference and differential equations.

The situation changed dramatically about the middle of the 20th century. There were two ingenious concepts at the beginning of this disruption. The first is the concept of a martingale introduced by Doob. The martingale turned out to be the final mathematical fixation of the idea of noise. The notion of a martingale is located between a process with uncorrelated increments and a process with independent increments, both of which were the competing noise concepts up to that time. The second concept is that of a stochastic integral due to K. Ito. This notion makes it possible to apply differential reasoning to stochastic dynamics.

At the beginning of the stochastic part of this lecture we will present an introduction to the ideas of martingales and stopping times at hand of stochastic sequences (discrete time processes). However, the main subject of the second half of the lecture will be continuous time processes with a strong focus on the Wiener process. However, the notions of martingales, semimartingales and stochastic integrals are introduced in a way which lays the foundation for the study of more general process theory. The choice of examples is governed by be the needs of financial applications (covering the notion of gambling, of course).

0.2 Literature

Let us give some comments to the bibliography.

The popular monograph by Bauer, [1], has been for a long time the standard textbook in Germany on measure theoretic probability. However, probability theory has many different faces. The book by Shiryaev, [21], is much closer to those modern concepts we are heading to. Both texts are mathematically oriented, i.e. they aim at giving complete and general proofs of fundamental facts, preferable in abstract terms. A modern introduction into probability models containing plenty of fascinating phenomena is given by Bremaud, [6] and [7]. The older monograph by Bremaud, [5], is not located at the focus of this lecture but contains as appendix an excellent primer on probability theory.

Our topic in stochastic processes will be the Wiener process and the stochastic analysis of Wiener driven systems. A standard monograph on this subject is Karatzas and Shreve, [15]. The Wiener systems part of the probability primer by Bremaud gives a very compact overview of the main facts. Today, Wiener driven systems are a very special framework for modelling financial markets. In the meanwhile, general stochastic analysis is in a more or less final state, called semimartingale theory. Present and future research applies this theory in order to get a much more flexible modelling of financial markets. Our introduction to semimartingale theory follows the outline by Protter, [20] (see also [19]).

Let us mention some basic literature on mathematical finance.

There is a standard source by Hull, [11]. Although this book heavily tries to present itself as not demanding, nevertheless the contrary is true. The reason is that the combination of financial intuition and the appearently informal utilization of advanced mathematical tools requires on the reader's side a lot of mathematical knowledge in order to catch the intrinsics. Paul Wilmott, [22] and [23], tries to cover all topics in financial mathematics together with the corresponding intuition, and to make the analytical framework a bit more explicit and detailed than Hull does. I consider these books by Hull and Wilmott as a must for any beginner in mathematical finance.

The books by Hull and Wilmott do not pretend to talk about mathematics. Let us mention some references which have a similar goal as this lecture, i.e. to present the mathematical theory of stochastic analysis aiming at applications in finance.

A very popular book which may serve as a bridge from mathematical probability to financial mathematics is by Björk, [4]. Another book, giving an introduction both to the mathematical theory and financial mathematics, is by Hunt and Kennedy, [12].

Standard monographs on mathematical finance which could be considered as cornerstones marking the state of the art at the time of their publication are Karatzas and Shreve, [16], Musiela and Rutkowski, [17], and Bielecki and Rutkowski, [3]. The present lecture should lay some foundations for reading books of that type.

0.3 Nature of these notes

These lecture notes are not intended to be a self-contained text to be used for self study. The notes are rather an outline of the main concepts and facts.

The classroom lecture will be a selection of the notes but in parts present more explanantion and motivation. Diagrams will be drawn on the blackboard and are not copied to the notes. Many facts are formulated in the notes as exercises with or without hints. Some of the exercises will be solved during the lecture, some are home exercises or classroom exercises to be presented by students.

The style of the text is very formal. Together with informal explanations during the lecture the text should train the students for studying more advanced literature. In order to meet the different skills of the audience (applied, theoretic, formal or informal) the exercises are classified with respect to difficulty and required mathematical skills.

The written exams will be open book exams, meaning that the lecture notes without any additional comments may be used during the exam. The problems of the exams will be exercises of the notes or additional exercises posed in the classroom. The solutions should include extensive references of the concepts used in the solution to the notions contained in the notes.

The final collection of those problems where the exams are sampled from will be fixed during the lecture.

The notes are not yet finished. There are some chapters and sections missing (”under construction”) and a lot of review questions have still to be formulated to be used for the exams. Filling of the gaps will be performed in the light of student reactions and feedback during the classes.

0.4 Time table

The following is concerning with the course from January to march 2006.

Week	Unit	Subject
2	1	measure and probability
	2	measure and probability
	3	measurable functions
3	4	exercises
	5	integrals and expectation
	6	integrals and expectation
	7	integrals and expectation
4	8	exercises
	9	conditional expectation
	10	stochastic sequences (gambling)
	11	stochastic sequences (martingales)
5	12	exercises
	13	Wiener process
	14	Wiener process (first passage times)
	15	Wiener process (stopping times)
6		skiing, midterm test
7	16	exercises
	17	the financial market picture (discrete time trading)
	18	stochastic calculus (Stieltjes integrals, differential notation)
	19	stochastic calculus (semimartingales, stochastic integrals)
8	20	exercises
	21	stochastic calculus (Ito calculus)
	22	applications to financial markets (Black Scholes model)

9	23	exercises
	24	linear stochastic differential equations
	25	martingales and stochastic integrals
10	26	martingales (Levy's theorem, martingale representation)
	27	martingales (Girsanov theorem)
	28	exercises

Part I

Measure theory

Chapter 1

Measure and probability

1.1 Fields and contents

We start with the notion of a field. Roughly speaking, a field is a system of subsets where the basic set operations (union, intersection, complementation) can be performed without leaving the system.

1.1 Definition. Let $\Omega \neq \emptyset$ be a set. A *field* on Ω is a system $\mathcal{A}$ of subsets $A \subseteq \Omega$ satisfying the following conditions:
(1) $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$
(2) If $A_1,\ A_2 \in \mathcal{A}$ then $A_1 \cup A_2 \in \mathcal{A}$ and $A_1 \cap A_2 \in \mathcal{A}$
(3) If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$

1.2 Problem. (*easy*)
Discuss minimal sets of conditions such that a system is a field.

The second basic notion is that of a content. A content is an additive set function on a field.

1.3 Definition. A *content* is a set function μ defined on a field $\mathcal{A}$ such that
(1) $\mu(A) \in [0, \infty]$ whenever $A \in \mathcal{A}$
(2) $\mu(\emptyset) = 0$
(3) $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$ whenever $A_1,\ A_2 \in \mathcal{A}$ and $A_1 \cap A_2 = \emptyset$

1.4 Problem. (*easy*)
Let $\mu|\mathcal{A}$ be a content. Then $A_1 \subseteq A_2$ implies $\mu(A_1) \leq \mu(A_2)$.

1.5 Problem. (*intermediate*)
(a) Show that every content satisfies the inclusion-exclusion law:

$$\mu(A_1) + \mu(A_2) = \mu(A_1 \cup A_2) + \mu(A_1 \cap A_2)$$

(b) The preceding problem gives a formula for $\mu(A_1 \cup A_2)$ provided that all sets have finite content. Extend this forumla to the union of three sets.

1.6 Definition. Let $\mathcal{A}$ be a field and let $\mu|\mathcal{A}$ be a content. The content μ is called *σ-additive* if

$$\mu\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \sum_{i\in\mathbb{N}} \mu(A_i)$$

for every pairwise disjoint sequence $(A_i)_{k\in\mathbb{N}} \subseteq \mathcal{A}$ such that $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$.

If a content is σ-additive then the content has several continuity properties which facilitate caculations.

1.7 Lemma. *Let $\mu|\mathcal{A}$ be a content on a field. Consider the following properties:*

(a) $\mu|\mathcal{A}$ is σ-additive.

(b) For every sequence $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}$ such the $A_i \uparrow A \in \mathcal{A}$ we have $\mu(A_i) \uparrow \mu(A)$.

(c) For every sequence $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}$ such the $A_i \downarrow A \in \mathcal{A}$ with $\mu(A_1) < \infty$ we have $\mu(A_i) \downarrow \mu(A)$.

(d) For every sequence $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}$ such the $A_i \downarrow \emptyset$ with $\mu(A_1) < \infty$ we have $\mu(A_i) \downarrow 0$.

Then: (a) $\Leftrightarrow$ (b) $\Rightarrow$ (c) $\Leftrightarrow$ (d)

If $\mu(\Omega) < \infty$ then all assertions are equivalent.

Proof: See Bauer, [1]. □

Further reading: Shiryaev [21], chapter II, paragraph 1.

As we shall see later when we are dealing with measures it is very easy to construct contents. But it is not easy to construct contents with given properties, e.g. with special geometric properties. The next paragraphs deal with the most important examples of contents which later will be extended to those measures that are most common in applications.

Contents on the real line

Let $\Omega = (-\infty, \infty]$ and let $\mathcal{R}$ be the system of subsets arising as unions of finitely many intervals of the form $(a, b]$ where $-\infty \le a < b \le \infty$ (left-open and right-closed intervals).

1.8 Problem. (*intermediate*)
Explain why $\mathcal{R}$ is a field. (Include $\emptyset$ as the union of nothing).

1.9 Problem. (*advanced*)
Show that each element $B \in \mathcal{R}$ can be written as a union of disjoint intervals

$$B = \bigcup_{i=1}^{n} (a_i, b_i] \tag{1}$$

where $-\infty \le a_1 < b_1 \le a_2 < b_2 \le a_3 < \ldots < b_{n-1} \le a_n < b_n \le \infty$.
Hint: Let $\mathcal{H}$ be the system of disjoint unions of intervals. First, show that $\mathcal{H}$ is closed under intersections. Be careful when applying the distributive law. Second, that any finite union of intervals can be written as

$$\bigcup_{i=1}^{n} I_i = I_1 \cup (I_2 \setminus I_1) \cup (I_3 \setminus (I_1 \cup I_2)) \cdots$$

where

$$I_k \setminus (I_1 \cup \cdots \cup I_{k-1})$$

is in $\mathcal{H}$. For the latter apply the first part of the proof.

Let $\alpha : \mathbb{R} \to \mathbb{R}$ be an increasing function. Define $\alpha(-\infty) = \inf \alpha$ and $\alpha(\infty) = \sup \alpha$. It will be shown in the following exercises that

$$\lambda_\alpha((a, b]) := \alpha(b) - \alpha(a) \tag{2}$$

determines a content on $\mathcal{R}$. (Note that in probablity theory this is the usual way to define probability distributions by distribution functions !)

1.10 Problem. (*advanced*)
(a) Show that any (hypothetical) content λ_α satisfying (2) necessarily satisfies

$$A = \bigcup_{i=1}^{n} I_i, \text{ where } (I_i) \text{ are pw. dj. intervals } \Rightarrow \lambda_\alpha(A) = \sum_{i=1}^{n} \lambda_\alpha(I_i) \tag{3}$$

(b) Show that using (3) as a definition is unambiguous.
(c) Show that (3) defines a content on $\mathcal{R}$ which is finite on bounded sets.

1.11 Definition. The content $\lambda_\alpha|\mathcal{R}$ is called a *Lebesgue-Stieltjes content*.

1.12 Example. Lebesgue content
Let $\alpha(x) = x$. Then $\lambda_\alpha((a, b]) = b - a$ is the length of the interval $(a, b]$. Therefore, in this special case the content λ_α is simply the geometric volume function. It is called the *Lebesgue content* and is denoted by λ.

We have seen that any increasing function $\alpha : \mathbb{R} \to \mathbb{R}$ defines a content $\lambda_\alpha|\mathcal{R}$. In order to extend such contents to greater families of sets we have to check whether $\lambda_\alpha|\mathcal{R}$ is σ-additive.

1.13 Lemma. *The content $\lambda_\alpha|\mathcal{R}$ is σ-additive iff α is right continuous.*
Proof: Assume that $\lambda_\alpha|\mathcal{R}$ is σ-additive. Then from

$$(a, b] = \bigcap_{n \in \mathbb{N}} (a, b + 1/n]$$

it follows

$$\alpha(b) - \alpha(a) = \lambda_\alpha((a,b]) = \lim_{n\to\infty} \lambda_\alpha((a,b+1/n]) = \lim_{n\to\infty} \alpha(b+1/n) - \alpha(a).$$

This means that α is right continuous at b.

This was the easy part. The proof of the converse is a bit more tricky. We show the converse for bounded α only ($\lambda_\alpha(\Omega) < \infty$). Let us prove that 1.7(d) is satisfied.

Let $(A_n) \subseteq \mathcal{R}$ such that $A_n \downarrow \emptyset$. Choose $\epsilon > 0$. For every A_n we may find a compact set K_n and a set $B_n \in \mathcal{R}$ such that $B_n \subseteq K_n \subseteq A_n$ and $\lambda_\alpha(A_n \setminus B_n) < \epsilon$. (At this point right-continuity goes in !) Since $A_n \downarrow \emptyset$ it follows that $K_n \downarrow \emptyset$. Since the sets K_n are compact there is some N such that $K_N = \emptyset$, hence $B_N = \emptyset$. (This is the so-called finite intersection-property of compect sets). It follows that $\lambda_\alpha(A_N) < \epsilon$. Since ϵ is arbitrarily small we have $\lambda_\alpha(\bigcap_{n=1}^\infty A_n) = 0$. This proves 1.7(d) and the assertion for finite contents. □

Contents on $\mathbb{R}^d$

The following is a summary of facts. Proofs are similar but sometimes a bit more complicated than in the one-dimensional case.

Denote $\overline{\mathbb{R}} := (-\infty, \infty]$ and let $\mathcal{Q}^d$ be the collection of all subsets $Q \subseteq \overline{\mathbb{R}}^d$ of the form

$$Q = \prod_{i=1}^d (a_i, b_i], \quad -\infty \le a_i < b_i \le \infty,$$

(so-called left-open right-closed parallelotops). Denote by $\mathcal{R}^d$ the set of all finite unions of sets in $\mathcal{Q}^d$. The sets in $\mathcal{R}^d$ are called figures.

1.14 Theorem. *(a) The set $\mathcal{R}^d$ is a field.*
(b) Each set $Q \in \mathcal{R}^d$ is a union of pairwise disjoint sets of $\mathcal{Q}^d$.

For the proof see Bauer [1].

In order to define a content on $\mathcal{R}^d$ we first have to define the content on $\mathcal{Q}^d$ and then try to extend it to $\mathcal{R}^d$. This was exactly the procedure that we performed for $d = 1$. For $d > 1$ it is natural to consider the geometric volume

$$\lambda^d\Big(\prod_{i=1}^d (a_i, b_i]\Big) := \prod_{i=1}^d (b_i - a_i) \qquad (4)$$

This can actually be extended to $\mathcal{R}^d$ resulting in a content called the *Lebesgue content*.

1.15 Theorem. *There is a uniquely determined content λ^d on $\mathcal{R}^d$ such that (4) is satisfied. The content λ^d is σ-additive.*

For a proof see Bauer [1].

Finite fields

Since many probabilistic applications are concerned with finite fields it is illuminating to discuss the structure of finite fields in more detail. Let us collect the main facts in terms of exercises.

1.16 Problem. (*easy*)
Let $\mathcal{C} = (C_1, C_2, \ldots, C_m)$ be a finite partition of Ω. Show that

$$\mathcal{R} := \Big\{ \bigcup_{i \in \alpha} C_i : \alpha \subseteq (1, \ldots, m) \Big\}$$

is a field on Ω and that it is the smallest field containing $\mathcal{C}$.

In the situation of 1.16 we say that the partition $\mathcal{C}$ *generates* the field $\mathcal{R}$.

1.17 Problem. (*intermediate*)
Show that every finite field is generated by a partition.
Hint: A set $A \in \mathcal{R}$ is called an atom if

$$A \neq \emptyset, \quad \text{and} \quad \emptyset \neq B \subseteq A,\ B \in \mathcal{R} \ \Rightarrow\ B = A$$

Show that the collection $\mathcal{C}$ of all atoms of $\mathcal{R}$ is a partition generating $\mathcal{R}$. (Show that for $x \in \Omega$ the set $A_x := \bigcap \{A \in \mathcal{R} : x \in A\}$ is the unique atom containing x.)

1.18 Problem. (*easy*)
Let $\mathcal{R}$ be a finite field and let $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$ be the generating partition. Show that for every choice of numbers $a_i \geq 0$ there exists exactly one content $\mu|\mathcal{R}$ sucht that $\mu(C_i) = a_i$.

The preceding assertions are the basis of the elementary theory of probability. The so-called Laplacian definition of a probability content results in the uniform content, i.e. $\mu(C_i) = 1/m$.

Further reading: Shiryaev [21], chapter I.

1.19 Review questions. Explain the structure and generation of finite fields. How to define contents on finite fields ?

1.2 Sigma-fields and measures

1.20 Definition. A field $\mathcal{F}$ on Ω is a *σ-field* if

$$(F_i)_{i \in \mathbb{N}} \subseteq \mathcal{F} \ \Rightarrow\ \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$$

A pair $(\Omega, \mathcal{F})$ where $\mathcal{F}$ is a σ-field on Ω is called a *measurable space*.

1.21 Problem. (*intermediate*)
(a) A field $\mathcal{F}$ is a σ-field iff

$$(F_i)_{i\in\mathbb{N}} \subseteq \mathcal{F} \;\Rightarrow\; \bigcap_{i\in\mathbb{N}} A_i \in \mathcal{F}$$

(b) A field $\mathcal{F}$ is a σ-field iff the union of every increasing (decreasing) sequence of sets in $\mathcal{F}$ is in $\mathcal{F}$, too.

(c) A field $\mathcal{F}$ is a σ-field iff the union of every pairwise disjoint sequence of sets in $\mathcal{F}$ is in $\mathcal{F}$, too.

1.22 Definition. A σ-additive content which is defined on a σ-field is called a *measure*.

A measure (resp. content) is called *finite* if $\mu(\Omega) < \infty$. A measure P (resp. content) is called a *probability measure* (resp. content) if $P(\Omega) = 1$. If $\mu|\mathcal{F}$ is a measure then $(\Omega, \mathcal{F}, \mu)$ is a *measure space*. If $P|\mathcal{F}$ is a probability measure then $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

As for the existence of measures things are easy with finite fields. Actually any finite field is a σ-field and any content on a finite field is σ-additive (in a trivial sense) and is therefore a measure. Therefore the concept of a measure differs from the concept of a content only on infinite σ-fields.

In the following we perform some warming up by discussing some very simple examples of measures.

1.23 Problem. (*easy*)
Let $(\Omega, \mathcal{F})$ be any measurable space. Let $x \in \Omega$ some point and keep it fixed. For every $A \in \mathcal{F}$ define

$$\delta_x(A) = \begin{cases} 1 & \text{whenever } x \in A \\ 0 & \text{whenever } x \notin A \end{cases}$$

Show that $\delta_x : A \mapsto \delta_x(A)$ is a measure (the *one-point measure* at the point x).
(Note that the case $\mathcal{F} = 2^\Omega$ is covered by this definition.)

1.24 Problem. (*easy*)
(a) Show that every finite linear combination of measures with nonnegative coefficients is a measure.

(b) Show that every countable linear combination of measures with nonnegative coefficients is a measure.

Any linear combination of point ist called a *discrete measure*.

1.25 Problem. (*easy*)

Describe the values of finite linear combinations of one-point measures:
Let $a_1, a_2, \ldots, a_m$ be any pairwise different points in Ω and keep them fixed. Let $p_1, p_2, \ldots, p_m$ be any nonnegative numbers. Then

$$\mu := \sum_{i=1}^{n} p_i \delta_{a_i} \;\Rightarrow\; \mu(A) = \sum_{i: a_i \in A} p_i, \; A \subseteq \Omega.$$

1.26 Problem. (*easy*)

(a) Write the binomial distribution as a linear combination of point measures.

(b) Write the geometric distribution as a linear combination of point measures.

1.27 Problem. (*easy*)

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be any finite sequence of elements in Ω (e.g. an empirical sample). Let $\{a_1, a_2, \ldots, a_n\}$ be the set of different components of $\boldsymbol{x}$ and denote by f_j the relative frequency of a_j in $\boldsymbol{x}$.

Show that the *empirical measure* satisfies

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} = \sum_{j=1}^{m} f_j \delta_{a_j}$$

(This is the ”frequency table” of an empirical distribution.)

When we are dealing with point measures or linear combinations of point measures we need not worry about σ-fields since such measures are well-defined on $\mathcal{F} = 2^{\Omega}$. In general, however, it is not possible to define measures with given properties on $\mathcal{F} = 2^{\Omega}$. We have to be more modest and to be satisfied if we find measures that are defined on σ-fields containing at least reasonable sets indispensable for applications.

Usually it is not very difficult to find a field which contains sufficiently many reasonable sets, e.g. the field of figures in $\mathbb{R}^d$ containing all rectangles. But how to proceed from fields to σ-fields ?

Let $\mathcal{A}$ be a field which is not a σ-field. We would like to enlarge $\mathcal{A}$ in such a way that the result is a σ-field. The following questions arise:

(1) Are there any σ-fields $\mathcal{F}$ containg $\mathcal{A}$?

(2) If yes, is there a smallest σ-field containing $\mathcal{A}$?

The answer to both questions is yes.

1.28 Problem. (*advanced*)

(a) The system 2^{Ω} (system of all subsets of Ω is a σ-field.

(b) The intersection of any family of σ-fields is a σ-field.

(c) Let $\mathcal{C}$ be any system of subsets on Ω and denote by $\sigma(\mathcal{C})$ the intersection of all σ-fields containing $\mathcal{C}$:

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{C}\subseteq\mathcal{F}} \mathcal{F}$$

Then $\sigma(\mathcal{C})$ is the smallest σ-field containing $\mathcal{C}$:

$$\mathcal{C} \subseteq \mathcal{F},\ \mathcal{F} \text{ is a } \sigma\text{-field} \ \Rightarrow\ \mathcal{C} \subseteq \sigma(\mathcal{C}) \subseteq \mathcal{F}$$

1.29 Definition. For any system $\mathcal{C}$ of sets in Ω the smallest σ-field $\mathcal{F}$ that contains $\mathcal{C}$ is called the *σ-field generated by $\mathcal{C}$* and is denoted by $\mathcal{F} = \sigma(\mathcal{C})$. The system $\mathcal{C}$ is called a *generator of $\mathcal{F}$*.

It turns out that the situation is simple as long as Ω is a countable set.

1.30 Problem. (*easy*)
Show that the σ-field on $\mathbb{N}$ which is generated by the one-point sets of $\mathbb{N}$ is $\mathcal{F} = 2^{\mathbb{N}}$.

The preceding exercise shows: If we want to have all one-point sets in the σ-filed then for countable Ω every subset of Ω has to be in the σ field.

1.31 Problem. (*intermediate*)
(a) Let $\Omega = \mathbb{N}$ and $\mathcal{F} = 2^{\mathbb{N}}$. Define $\mu(A) := |A|$, $A \subseteq \mathbb{N}$ (the *counting measure*). Show that $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$ is a measure space.
(b) Show that for every sequence of numbers $a_n \geq 0$ there is exactly one measure $\mu|\mathcal{F}$ such that $\mu(\{n\}) := a_n$.
(c) Discuss how to define probability measures on $(\mathbb{N}, 2^{\mathbb{N}})$.

The following exercise shows that for $\Omega = \mathbb{R}$ the system of one-point sets is not sufficient to generate a reasonable σ-field.

1.32 Problem. (*intermediate for mathematicians*)
What is the σ-field on $\mathbb{R}$ that is generated by the one-point sets of $\mathbb{R}$?
Answer: The system of sets which are either countable or the complement of a countable set.

The preceding exercise shows that one-point sets do not generate a σ-field on $\mathbb{R}$ which contains intervals ! Therefore we have to include intervals in our generating system. The starting point is the algebra $\mathcal{R}$ of figures.

1.33 Definition. The σ-field on $\mathbb{R}$ ($\mathbb{R}^d$) which is generated by the algebra $\mathcal{R}$ ($\mathcal{R}^d$) is called the *Borel σ-field* and is denoted by $\mathcal{B} = \mathcal{B}(\mathbb{R})$ ($\mathcal{B}(\mathbb{R}^d)$).

The sets in the Borel σ-field are called *Borel sets*.

1.34 Problem. (*easy*)
(a) Show that $\mathcal{B}$ contains all intervals (including one-point sets).
(b) Is $\mathbb{Q}$ a Borel set ?

1.35 Problem. (*easy for mathematicians*) Show that $\mathcal{B}(\mathbb{R}^d)$ contains all open sets and all closed sets.

1.36 Review questions. What is a field and what is a σ-field ? What is the difference between a content and a measure.

1.37 Review questions. Explain the ideas of generating a σ-field by a system of sets. Explain how the measurable spaces $(\mathbb{N}, 2^{\mathbb{N}})$, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ are generated.

1.3 The extension theorem

The fundamental problem of measure theory is the extension problem. The extension problem deals with the question whether a given content on a field $\mathcal{A}$ can be extended to a measure on the σ-field $\sigma(\mathcal{A})$.

It is clear that for the existence of an extension the content must be σ-additive. This is a necessary condition. For finite contents it is even sufficient.

1.38 Theorem. *Every finite σ-additive content $\mu|\mathcal{A}$ defined on a field has a uniquely determined measure extension to $\mathcal{F} = \sigma(\mathcal{A})$.*
Proof: (Outline. Further reading: Bauer [1])
Let us indicate some ideas of the proof. W.l.g. we assume that $\mu(\Omega) = 1$.

The first step of the proof is to try an extension of the content to all subsets $M \subseteq \Omega$. This is done by

$$\mu^*(M) = \inf\Big\{ \sum_{i\in\mathbb{N}} \mu(A_i) : (A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}, M \subseteq \bigcup_{i\in\mathbb{N}} A_i \Big\}$$

Unfortunately, this definition does not result in a measure, even not in a content. The set function μ^* is a so-called outer measure. It is clear that $\mu^*|\mathcal{A} = \mu|\mathcal{A}$.

Next, define

$$\mathcal{M} := \{M \subseteq \Omega : \mu^*(M) + \mu^*(M^c) = 1\}$$

It is clear that $\mathcal{A} \subseteq \mathcal{M}$. It turns out that $\mathcal{M}$ is a σ-field and therefore $\sigma(\mathcal{A}) \subseteq \mathcal{M}$. Moreover, it is shown that the restriction $\mu^*|\mathcal{M}$ is a measure. Therefore it is a measure extension of $\mu|\mathcal{A}$ to some σ-field containing $\mathcal{A}$, thus at least a measure extension to $\sigma(\mathcal{A})$.

The uniqueness of the extension is shown in the following way. Let $\mu_1|\sigma(\mathcal{A})$ and $\mu_2|\sigma(\mathcal{A})$ be two extensions of $\mu|\mathcal{A}$. Let

$$\mathcal{M}_1 := \{M \in \sigma(\mathcal{A}) : \mu_1(M) = \mu_2(M)\}$$

By assumption we have $\mathcal{A} \subseteq \mathcal{M}_1$. It can be shown that $\mathcal{M}_1$ is a σ-field. Then it follows that $\sigma(\mathcal{A}) \subseteq \mathcal{M}_1$. □

1.39 Remarks.

The following remarks can be understood (”proved”) with the information provided by the outline of the proof of the measure extension theorem. Assume that $\mu(\Omega) < \infty$.

(1) Although the sets in $\sigma(\mathcal{A})$ or $\mathcal{M}$ can be rather complicated they don't differ very much from sets in $\mathcal{A}$: For every $M \in \mathcal{M}$ and every (arbitrarily small) $\epsilon > 0$ there is a set $A \in \mathcal{A}$ such that $\mu(M \setminus A) < \epsilon$ and $\mu(A \setminus M) < \epsilon$.

(2) The proof of the measure extension theorem results in an extension to $\mathcal{M}$ which actually is a larger σ-field than $\sigma(\mathcal{A})$. However, from (1) it follows that for every $M \in \mathcal{M}$ there is some $A \in \sigma(\mathcal{A})$ such that $\mu(M \setminus A) = 0$ and $\mu(A \setminus M) = 0$.

1.40 Problem. (*advanced for mathematicians*)
(1) Prove assertion (1) of Remark 1.39.
(2) Prove assertion (2) of Remark 1.39.

What about the extension of non-finite contents ?

1.41 Definition. A content $\mu|\mathcal{A}$ is called *σ-finite* if there is a sequence $(A_i)_{i\in\mathbb{N}} \subseteq \mathcal{A}$ such that $\bigcup_{i\in\mathbb{N}} A_i = \Omega$ and $\mu(A_i) < \infty$ for every $i \in \mathbb{N}$.

1.42 Theorem. *Every σ-finite σ-additive content $\mu|\mathcal{A}$ defined on a field has a uniquely determined measure extension to $\mathcal{F} = \sigma(\mathcal{A})$.*

The proof is similar, but a bit more complicated than in the finite case.

We may apply the measure extension theorem since every $\lambda_\alpha|\mathcal{R}$ is obviously σ-finite.

1.43 Corollary. *For every increasing and right continuous function $\alpha : \mathbb{R} \to \mathbb{R}$ there is a uniquely determined measure $\lambda_\alpha|\mathcal{B}$ such that $\lambda_\alpha((a, b]) = \alpha(b) - \alpha(a)$.*

For $\alpha(x) = x$ the measure $\lambda_\alpha = \lambda$ is called the Lebesgue measure.

It should be noted that there are subsets of $\mathbb{R}$ (resp. $\mathbb{R}^d$) that are not Borel sets. However, the construction of such sets can be very complicated.

1.44 Review questions. State the measure extension theorem. Show how to apply this theorem for defining Borel measures on $\mathbb{R}$.

Chapter 2

Measurable functions and random variables

2.1 The idea of measurability

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $(Y, \mathcal{B})$ be a measurable space. Moreover, let $f : \Omega \to Y$ be a function. We are going to consider the problem of mapping the measure μ to the set Y be means of the function f.

The following example serves as a first motivation. More details concerning measure theoretic probability concepts are given in section 2.4.

2.1 Example. Distribution of a random variable

The concept of the distribution of a random variable an important special case of mapping a measure from one set to another.

Let X be a random variable. This is a function from a probability space $(\Omega, \mathcal{A}, P)$ to $\mathbb{R}$. Since the probability space is a rather abstract object it is convenient to put the essentials of the random variable X into analytically tractable terms. Usually we are only interested in the probabilities $P(X \in B)$ the collection of which is the distribution P^X, i.e. $P^X(B) = P(X \in B)$, $B \in \mathcal{B}$. The distribution is a set function on $(\mathbb{R}, \mathcal{B})$ and it is defined by mapping the probability measure P to $(\mathbb{R}, \mathcal{B})$ via the function X.

However, for defining $P^X(B)$ it is essential that the expression $P(X \in B)$ makes sense. This the case iff the inverse image $(X \in B) = X^{-1}(B)$ is in $\mathcal{A}$. Therefore a random variable cannot be an arbitrary function $X : \Omega \to \mathbb{R}$ but must satisfy $(X \in B) \in \mathcal{A}$ for all $B \in \mathcal{B}$. This property is called measurability.

2.2 Definition. A function $f : (\Omega, \mathcal{A}) \to (Y, \mathcal{B})$ is called $(\mathcal{A}, \mathcal{B})$-measurable if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.

If $f : (\Omega, \mathcal{A}, \mu) \to (Y, \mathcal{B})$ is $(\mathcal{A}, \mathcal{B})$-measurable then we may define

$$\mu^f(B) := \mu(f \in B) = \mu(f^{-1}(B)), \quad B \in \mathcal{B}.$$

This is the *image of μ under f* or the *distribution of f under μ*.

2.3 Problem. (*easy*)
Show that μ^f is indeed a measure on $\mathcal{B}$.

Let us agree upon some terminology.

(1) When we consider real-valued functions then we always use the Borel-σ-field in the range of f.
E.g.: If $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ then we simply say that f is $\mathcal{F}$-measurable if we mean that it is $(\mathcal{F}, \mathcal{B})$-measurable.

(2) When we consider functions $f : \mathbb{R} \to \mathbb{R}$ then $(\mathcal{B}, \mathcal{B})$-measurability is called *Borel measurability*. The term "Borel" is thus concerned with the σ-field in the domain of f.

To get an idea what measurability means let us consider some simple examples.

2.4 Problem. (*easy*)
Let $\Omega, \mathcal{F}, \mu)$ be a measure space and let $f = 1_A$ where $A \subseteq \Omega$.
(a) Show that f is $\mathcal{F}$-measurable iff $A \in \mathcal{F}$.
(b) Find μ^f.

It follows that very complicated functions are Borel-measurable, e.g. $f = 1_{\mathbb{Q}}$.

Recall that a *simple function* is a real-valued function which has only finitely many values. Any simple function f can be written as

$$f = \sum_{i=1}^{n} a_i 1_{F_i}$$

if $F_i = (f = a_i)$ where $\{a_1, a_2, \ldots, a_n\}$ denotes the set of different function values of f. This is the *canonical representation* of f.

Any linear combination of indicator functions is simple but need not be canonical. It is canonical iff both the sets supporting the indicators are pairwise disjoint and the coefficients are pairwise different. There is exactly one canonical representation.

2.5 Problem. (*easy*)
Let $\Omega, \mathcal{F}, \mu)$ be a measure space and let $f : \Omega \to \mathbb{R}$ be a simple function.
(a) Show that f is $\mathcal{F}$-measurable iff all sets of the canonical representation are in $\mathcal{F}$.
(b) Find μ^f.

2.2 The basic abstract assertions

There are two fundamental principles for dealing with measurability. The first proiciple says that measurability is a property which is preserved under composition of functions.

2.6 Theorem. *Let $f : (\Omega, \mathcal{A}) \to (Y, \mathcal{B})$ be $(\mathcal{A}, \mathcal{B})$-measurable, and let $g : (Y, \mathcal{B}) \to (Z, \mathcal{C})$ be $(\mathcal{B}, \mathcal{C})$-measurable. Then $g \circ f$ is $(\mathcal{A}, \mathcal{C})$-measurable.*

2.7 Problem. (*easy*) Prove 2.6.

The second principle is concerned with checking measurability. For checking measurability it is sufficient to consider the sets in a generating system of the range σ-field.

2.8 Theorem. *Let $f : (\Omega, \mathcal{A}) \to (Y, \mathcal{B})$ and let $\mathcal{C}$ be a generating system of $\mathcal{B}$, i.e. $\mathcal{B} = \sigma(\mathcal{C})$. Then f is $(\mathcal{A}, \mathcal{B})$-measurable iff $f^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$.*

Proof: Let $\mathcal{D} := \{D \subseteq Y : f^{-1}(D) \in \mathcal{A}\}$. It can be shown that $\mathcal{D}$ is a σ-field. If $f^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$ then $\mathcal{C} \subseteq \mathcal{D}$. This implies $\sigma(\mathcal{C}) \subseteq \mathcal{D}$. □

2.9 Problem. (*intermediate*) Fill in the details of the proof of 2.8.

2.10 Review questions. Explain the abstract concept of a measurable function. State the basic abstract properties of measurable functions.

2.3 The structure of real-valued measurable functions

Let $(\Omega, \mathcal{F})$ be a measurable space. Let $\mathcal{L}(\mathcal{F})$ be the set of all $\mathcal{F}$-measurable real-valued functions. We start with the most common and most simple criterion for checking measurability of a real-valued function.

2.11 Problem. (*intermediate*)

Show that a function $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable iff $(f \le \alpha) \in \mathcal{F}$ for every $\alpha \in \mathbb{R}$.

Hint: Apply 2.8.

This provides us with a lot of examples of Borel-measurable functions.

2.12 Problem. (*easy*)

(a) Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is Borel-measurable.

(b) Show that every continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{B}^n$-measurable.

Hint: Note that $(f \le \alpha)$ is a closed set.

(c) Let $f : (\Omega, \mathcal{F}) \to \mathbb{R}$ be $\mathcal{F}$-measurable. Show that f^+, f^-, $|f|$, and every polynomial $a_0 + a_1 f + \cdots + a_n f^n$ are $\mathcal{F}$-measurable.

The next exercise is a first step towards the measurability of expressions involving several measurable functions.

2.13 Problem. (*intermediate*) Let $(f_1, f_2, \ldots, f_n$ be measurable functions. Then

$$f = (f_1, f_2, \ldots, f_n) : \Omega \to \mathbb{R}^n$$

is $(\mathcal{F}, \mathcal{B}^n)$-measurable.

2.14 Corollary. *Let $f_1, f_2, \ldots, f_n$ be measurable functions. Then for every continuous function $\phi : \mathbb{R}^n \to \mathbb{R}$ the composition $\phi(f_1, f_2, \ldots, f_n)$ is measurable.*

Proof: Apply 2.6. □

2.15 Corollary. *Let f_1, f_2 be measurable functions. Then $f_1 + f_2$, $f_1 \cdot f_2$, $f_1 \cap f_2$, $f_1 \cup f_2$ are measurable functions.*

2.16 Problem. Prove 2.15.

As a result we see that $\mathcal{L}(\mathcal{F})$ is a space of functions where we may perform any algebraic operations without leaving the space. Thus it is a very convenient space for formal manipulations. Moreover, we may even perform all of those operations involving a countable set (e.g. a sequence) of measurable functions !

2.17 Theorem. *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions. Then $\sup_n f_n$, $\inf_n f_n$ are measurable functions. Let $A := (\exists \lim_n f_n)$. Then $A \in \mathcal{F}$ and $\lim_n f_n \cdot 1_A$ is measurable.*

Proof: Since

$$(\sup_n f_n \le \alpha) = \bigcap_n (f_n \le \alpha)$$

it follows from 2.11 that $\sup_n f_n$ and $\inf_n f_n = -\sup_n(-f_n)$ are measurable. We have

$$A := (\exists \lim_n f_n) = \left(\sup_k \inf_{n\ge k} f_n = \inf_k \sup_{n\ge k} f_n \right)$$

This implies $A \in \mathcal{F}$. The last statement follows from

$$\lim_n f_n = \sup_k \inf_{n\ge k} f_n \quad \text{on } A.$$

□

Note that the preceding corollaries are only very special examples for the power of theorem 2.6. Roughly speaking, any function which can be written as an expression involving countable many operations with countable many measurable functions is measurable. It is rather difficult to construct non-measurable functions.

Next we turn to the question how typical measurable functions look like. Let us denote the set of all $\mathcal{F}$-measurable simple functions by $\mathcal{S}(\mathcal{F})$. Clearly, all limits of simple measurable functions are measurable. The remarkable fact being fundamental for almost everything in integration theory is the converse of this statement.

2.18 Theorem. *(a) Every measurable function f is the limit of some sequence of simple measurable functions.*

(b) If f is bounded then the approximating sequence can be chosen to be uniformly convergent.

(c) If $f \geq 0$ then the approximating sequence can be chosen to be increasing.

Proof: The fundamental statement is (c).

Let $f \geq 0$. For every $n \in \mathbb{N}$ define

$$f_n := \begin{cases} (k-1)/2^n & \text{whenever } (k-1)/2^n \leq f < k/2^n, \ k = 1, 2, \ldots, n2^n \\ n & \text{whenever } f \geq n \end{cases}$$

Then $f_n \uparrow f$. If f is bounded then (f_n) converges uniformly to f. Parts (a) and b follow from $f = f^+ - f^-$. □

2.19 Problem. (*easy*)
Draw a diagram illustrating the construction of the proof of 2.18.

2.20 Review questions. Describe the structure of the set of real-valued measurable functions. Explain the role of simple functions.

2.4 Probability models

The term random variable is simply the probabilistic name of a measurable function.

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

2.21 Definition. Any $\mathcal{F}$-measurable real-valued function $X : \Omega \to \mathbb{R}$ is called a *random variable*.

Let X be a random variable. Then the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) := P(X \leq x), \quad x \in \mathbb{R}$$

is the *distribution function* of X. The distribution of X is P^X, i.e. the image of P under X defined by

$$P^X(B) := P(X^{-1}(B)) = P(X \in B), \quad B \in \mathcal{B}.$$

Thus, the distribution function F_X determines the values of the distribution P^X on intervals by

$$P^X((a, b]) = F(b) - F(a).$$

2.22 Problem. (*easy*)
(a) Show that any distribution function is right-continuous.
(b) Show that the distribution $P^X = \lambda_F$.

A major problem of probability theory is the converse problem: Given a function F, does there exist a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X such that $P^X = \lambda_F$?

2.23 Problem. (*easy*)
Let $F : \mathbb{R} \to [0,1]$ be increasing. Show that λ_F is a σ-additive probability content iff F is right-continuous and satisfies $F(-\infty) = 0$ and $F(\infty) = 1$.

2.24 Definition. A *distribution function* is a function $F : \mathbb{R} \to [0,1]$ which is increasing, right-continuous and satisfies $F(-\infty) = 0$ and $F(\infty) = 1$.

2.25 Problem. (*intermediate*)
Let F be a distribution function. Show that there is a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X such that $P(X \le x) = F(x)$.
Hint: Let $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, $P = \lambda_F$ and $X(\omega) = \omega$.

2.26 Example. Joint distribution functions

Let $X := (X_1, X_2, \dots, X_d)$ be a random vector. Then the distribution

$$P^X(Q) := P(X \in Q), \quad Q \in \mathcal{R}^d,$$

is a σ-additive content on $\mathcal{R}^d$.

The joint distribution function of X is defined to be

$$F(x_1, x_2, \dots, x_d) := P(X_1 \le x_1, X_2 \le x_2, \dots, X_d \le x_d)$$

For defining joint distributions one usually goes the other way round and starts with a function $F : \mathbb{R}^d \to \mathbb{R}$ to define

$$P(X_1 \le x_1, X_2 \le x_2, \dots, X_d \le x_d) := F(x_1, x_2, \dots, x_d)$$

But this only makes sense if this definition can be extended to an σ-additive content $P^X(Q) := P(X \in Q), \quad Q \in \mathcal{R}^d$. There are conditions on F to guarantee the possibility of such an extension.

Further reading: Shiryaev [21], chapter II, paragraph 3, sections 1-3.

2.27 Review questions. Explain how the measure extension theorem is applied to construct probability spaces and random variables with given distributions.

Chapter 3

Integral and expectation

3.1 The integral of simple functions

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We start with defining the μ-integral of a measurable simple function.

3.1 Definition. Let $f = \sum_{i=1}^n a_i 1_{F_i}$ be a nonnegative simple $\mathcal{F}$-measurable function with its canonical representation. Then

$$\int f \, d\mu := \sum_{i=1}^n a_i \mu(F_i)$$

is called the μ-*integral* of f.

We had to restrict the preceding definition to nonnegative functions since we admit the case $\mu(F) = \infty$. If we were dealing with a finite measure μ the definition would work for all $\mathcal{F}$-measurable simple functions.

3.2 Example.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X = \sum_{i=1}^n a_i 1_{F_i}$ be a simple random variable. Then we have $E(X) = \int X \, dP$.

3.3 Theorem. *The μ-integral on $\mathcal{S}(\mathcal{F})^+$ has the following properties:*
(1) $\int 1_F d\mu = \mu(F)$,
(2) $\int (sf + tg) \, d\mu = s \int f \, d\mu + t \int g \, d\mu$ *if* $s, t \in \mathbb{R}^+$ *and* $f, g \in \mathcal{S}(\mathcal{F})^+$
(3) $\int f \, d\mu \leq \int g \, d\mu$ *if* $f \leq g$ *and* $f,\, g \in \mathcal{S}(\mathcal{F})^+$
Proof: The only nontrivial part is to prove that $\int (f+g) d\mu = \int f d\mu + \int g d\mu$. □

3.4 Problem. (*intermediate*)
Show that $\int (f+g) d\mu = \int f d\mu + \int g d\mu$ for $f,\, g \in \mathcal{S}(\mathcal{F})^+$.
Hint: Try to find the canonical representation of $f + g$ in terms of the canonical representations of f and g.

It follows that the defining formula of the μ-integral can be applied to any (non-negative) linear combination of indicators, not only to canonical representations !

3.5 Theorem. *(Transformation theorem)*
Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $g \in \mathcal{L}(\mathcal{F})$. Then for every $f \in \mathcal{S}_+(\mathcal{B})$

$$\int f \circ g \, d\mu = \int f \, d\mu^g$$

3.6 Problem. (*easy*)
Prove 3.5.

3.7 Problem. (*easy*)
Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable with distribution function F. Explain the formula

$$E(f \circ X) = \int f \, d\lambda_F$$

3.2 The extension to nonnegative functions

We know that every nonnegative measurable function f is the limit of an increasing sequence (f_n) of measurable simple functions: $f_n \uparrow f$. It is a natural idea to think of the integral of f as something like

$$\int f \, d\mu := \lim_n \int f_n \, d\mu \tag{5}$$

This is actually the way we will succeed. But there are some points to worry about. The reader whose does not like to worry should grasp Beppo Levi's theorem and proceed to the next section.

First of all, we should ask whether the limit on the right hand side exists. The integrals $\int f_n d\mu$ form an increasing sequence in $[0, \infty]$. This sequence either has a finite limit or it increases to ∞. Both cases are covered by our definition.

The second and far more subtle question is whether the definition is compatible with the definition of the integral on $\mathcal{S}(\mathcal{F})$. This is the only nontrivial part of the extension process of the integral and it is the point where σ-additivity of μ is required.

3.8 Theorem. *Let $f \in \mathcal{S}(\mathcal{F})^+$ and $(f_n) \subseteq \mathcal{S}(\mathcal{F})^+$. Then*

$$f_n \uparrow f \;\Rightarrow\; \lim_n \int f_n \, d\mu = \int f \, d\mu$$

Proof: Note that ”$\leq$” is clear. For an arbitrary $\epsilon > 0$ let $B_n := (f \leq f_n \cdot (1+\epsilon))$. It is clear that

$$\int 1_{B_n} f\, d\mu \leq \int 1_{B_n} f_n \cdot (1+\epsilon)\, d\mu \leq \int f_n\, d\mu \cdot (1+\epsilon)$$

From $B_n \uparrow \Omega$ it follows that $A \cap B_n \uparrow A$ and $\mu(A \cap B_n) \uparrow \mu(A)$ by σ-additivity. We get

$$\int f\, d\mu = \sum_{j=1}^{n} \alpha_j \mu(A_j) = \lim_n \sum_{j=1}^{n} \alpha_j\, \mu(A_j \cap B_n) = \lim_n \int 1_{B_n} f\, d\mu$$

which implies

$$\int f\, d\mu \leq \lim_n \int f_n\, d\mu \cdot (1+\epsilon)$$

Since ϵ is arbitrarily small the assertion follows. □

The third question is whether the value of the limit is independent of the approximating sequence. This is straightforward using 3.8.

3.9 Theorem. *Let* (f_n) *and* $g_n)$ *be increasing sequences of nonnegative measurable simple functions. Then*

$$\lim_n f_n = \lim_n g_n \;\Rightarrow\; \lim_n \int f_n d\mu = \lim_n \int g_n d\mu.$$

Proof: It is sufficient to prove the assertion with ”$\leq$” replacing ”$=$”. Since $\lim_k f_n \cap g_k = f_n \cap \lim_k g_k = f_n$ we obtain by 3.8

$$\int f_n\, d\mu = \lim_k \int f_n \cap g_k\, d\mu \leq \lim_k \int g_k\, d\mu$$

□

Thus, we have a valid definition (5) of the integral on $\mathcal{L}(\mathcal{F})^+$. It is now straightforward that 3.3 carries over to $\mathcal{L}(\mathcal{F})^+$.

3.10 Theorem. *The* μ*-integral on* $\mathcal{L}(\mathcal{F})^+$ *has the following properties:*
(1) $\int 1_F d\mu = \mu(F)$,
(2) $\int (sf + tg)\, d\mu = s \int f\, d\mu + t \int g\, d\mu$ *if* $s, t \in \mathbb{R}^+$ *and* $f, g \in \mathcal{L}(\mathcal{F})^+$
(3) $\int f\, d\mu \leq \int g\, d\mu$ *if* $f \leq g$ *and* $f,\, g \in \mathcal{L}(\mathcal{F})^+$

The extension process is complete if we succed to extend 3.8 to $\mathcal{L}(\mathcal{F})^+$.

3.11 Theorem. *(Theorem of Beppo Levi)*
Let $f \in \mathcal{L}(\mathcal{F})^+$ *and* $(f_n) \subseteq \mathcal{L}(\mathcal{F})^+$*. Then*

$$f_n \uparrow f \;\Rightarrow\; \lim_n \int f_n\, d\mu = \int f\, d\mu$$

Proof: We have to show "$\geq$".

For every $n \in \mathbb{N}$ let $(f_{nk})_{k\in\mathbb{N}}$ be an increasing sequence in $\mathcal{S}(\mathcal{F})^+$ such that $\lim_k f_{nk} = f_n$. Define

$$g_k := f_{1k} \cup f_{2k} \cup \ldots \cup f_{kk}$$

Then

$$f_{nk} \leq g_k \leq f_k \leq f \text{ whenever } n \leq k.$$

It follows that $g_k \uparrow f$ and

$$\int f\,d\mu = \lim_k \int g_k\,d\mu \leq \lim_k \int f_k\,d\mu$$

□

3.12 Problem. (*intermediate for mathematicians*) Prove Fatou's lemma: For every sequence (f_n) of nonnegative measurable functions

$$\liminf_n \int f_n\,d\mu \geq \int \liminf_n f_n\,d\mu$$

Hint: Recall that $\liminf_n x_n = \lim_k \inf_{n\geq k} x_n$. Consider $g_k := \inf_{n\geq k} f_n$ and apply Levi's theorem to (g_k).

3.13 Problem. (*intermediate for mathematicians*)

For every sequence (f_n) of nonnegative measurable functions we have

$$\int \Big(\sum_{n=1}^{\infty} f_n\Big)\,d\mu = \sum_{n=1}^{\infty} \int f_n\,d\mu$$

3.14 Problem. (*intermediate*)

Let $f \in \mathcal{L}(\mathcal{F})^+$. Prove Markoff's inequality

$$\mu(f > a) \leq \frac{1}{a}\int f\,d\mu, \quad a > 0.$$

3.15 Problem. (*intermediate*)

Let $f \in \mathcal{L}(\mathcal{F})^+$. Show that $\int f\,d\mu = 0$ implies $\mu(f \neq 0) = 0$.
Hint: Show that $\mu(f > 1/n) = 0$ for every $n \in \mathbb{N}$.

An assertion A about a measurable function f is said to hold μ-*almost everywhere* (μ-a.e.) if $\mu(A^c) = 0$. Using this terminology the assertion of the preceding exercise can be phrased as:

$$\int f\,d\mu = 0,\ f \geq 0 \ \Rightarrow\ f = 0\ \mu\text{-a.e.}$$

If we are talking about probability measures and random variables ”almost everwhere” is sometimes replaced by ”almost sure”.

3.16 Problem. (*easy*)
Let $f \in \mathcal{L}(\mathcal{F})^+$. Show that $\int f \, d\mu < \infty$ implies $\mu(f > a) < \infty$ for every $a > 0$.

3.3 Integrable functions

Now the integral is defined for every nonnegative measurable function. The value of the integral may be ∞. In order to define the integral for measurable functions which may take both positive and negative values we have to exclude infinite integrals.

3.17 Definition. A measurable function f is μ-*integrable* if $\int f^+ \, d\mu < \infty$ and $\int f^- \, d\mu < \infty$. If f is μ-integrable then

$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu$$

The set of all μ-integrable functions is denoted by $\mathcal{L}^1(\mu) = \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$.

Proving the basic properties of the integral of integrable functions is an easy matter. We collect these fact in a couple of problems.

3.18 Problem. (*easy*)
Show that $f \in \mathcal{L}(\mathcal{F})$ is μ-integrable iff $\int |f| \, d\mu < \infty$.

3.19 Theorem. *The set $L^1(\mu)$ is a linear space and the μ-integral is a linear functional on $L^1(\mu)$.*

3.20 Problem. (*intermediate*)
Prove 3.19.

3.21 Theorem. *The μ-integral is an isotonic functional on $L^1(\mu)$.*

3.22 Problem. (*easy*)
Prove 3.21.

3.23 Problem. (*easy*)
Let $f \in L^1(\mu)$. Show that $|\int f \, d\mu| \le \int |f| \, d\mu$.

3.24 Problem. (*easy*)
Let f be a measurable function and assume that there is an integrable function g such that $|f| \le g$. Then f is integrable.

3.25 Problem. (*easy*)
(a) Discuss the question whether bounded measurable functions are integrable.

(b) Characterize those measurable simple functions which are integrable.

For notational convenience we denote

$$\int_A f\,d\mu := \int 1_A f\,d\mu, \quad A \in \mathcal{F}.$$

3.26 Problem. (*easy*)
(a) Let f be a measurable function such that $f = 0$ μ-a.e.. Then f is integrable and $\int f\,d\mu = 0$.
(b) Let f be an integrable function. Then

$$f = 0\,\mu\text{-a.e.} \;\Leftrightarrow\; \int_A f\,d\mu = 0 \text{ for all } A \in \mathcal{F}$$

3.27 Problem. (*easy*)
(a) Let f and g be measurable functions such that $f = g$ μ-a.e.. Then f is integrable iff g is integrable.
(b) Let f and g be integrable functions. Then

$$f = g\,\mu\text{-a.e.} \;\Leftrightarrow\; \int_A f\,d\mu = \int_A g\,d\mu \text{ for all } A \in \mathcal{F}$$

Many assertions in measure theory concerning measurable functions are stable under linear combinations and under convergence. Assertions of such a type need only be proved for indicators. The procedure of proving (understanding) an assertion for indicators and extending it to nonnegative and to integrable functions is called *measure theoretic induction*.

3.28 Problem. (*easy*)
Extend the transformation theorem by measure theoretic induction.

3.29 Problem. (*easy*)
Show that integrals are linear with respect to the integrating measure.

3.4 Convergence

One of the reasons for the great success of abstract integration theory are the convergence theorems for integrals. The problem is the following. Assume that (f_n) is a sequence of integrable functions converging to some function f. When can we conclude that f is integrable and

$$\lim_n \int f_n\,d\mu = \int f\,d\mu \quad ?$$

The most popular result concerning this issue is Lebesgue's theorem on dominated convergence.

3.30 Theorem. *Dominated convergence theorem*
Let (f_n) be a sequence of measurable function which is dominated by an integrable function g, i.e. $|f_n| \le g$, $n \in \mathbb{N}$. Then

$$f_n \to f \ \mu\text{-a.e.} \ \Rightarrow \ f \in L^1(\mu) \quad \textit{and} \quad \lim_n \int f_n \, d\mu = \int f \, d\mu$$

The dominated convergence can be used and applied like a black box without being aware of its proof. However, the proof is very easy and follows straightforward from Levi's theorem 3.11 and Fatou's lemma 3.12.

Proof: Integrability of f is obvious. Moreover, the sequences $g - f_n$ and $g + f_n$ consist of nonnegative measurable functions. Therefore we may apply Fatou's lemma:

$$\int (g - f) \, d\mu \le \liminf \int (g - f_n) \, d\mu = \int g \, d\mu - \limsup_n \int f_n \, d\mu$$

and

$$\int (g + f) \, d\mu \le \liminf \int (g + f_n) \, d\mu = \int g \, d\mu + \liminf_n \int f_n \, d\mu$$

This implies

$$\int f \, d\mu \le \liminf_n \int f_n \, d\mu \le \limsup_n \int f_n \, d\mu \le \int f \, d\mu$$

□

3.31 Problem. (*easy*)
Show that under the assumptions of the dominated convergence theorem we even have

$$\lim_n \int |f_n - f| \, d\mu = 0$$

(This type of convergence is called *mean convergence*.)

3.32 Problem. (*easy*)
Discuss the question whether a uniformly bounded sequence of measurable functions fulfills is dominated in the sense of the dominated convergence theorem.

Chapter 4

Selected topics

4.1 Spaces of integrable functions

We know that the space $\mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ is a vector space. We would like to define a norm on $\mathcal{L}^1$.

A natural idea is to define

$$||f||_1 := \int |f|\, d\mu, \quad f \in \mathcal{L}^1.$$

It is easy to see that this definition has the following properties:
(1) $||f||_1 \geq 0, \quad f = 0 \;\Rightarrow\; ||f||_1 = 0$,
(2) $||f + g||_1 \leq ||f||_1 + ||g||_1,\; f, g \in \mathcal{L}^1$,
(3) $||\lambda f||_1 \leq |\lambda|\, ||f||_1,\; \lambda \in \mathbb{R},\; f \in \mathcal{L}^1$.
However, we have

$$||f||_1 = 0 \;\Rightarrow\; f = 0 \;\mu\text{-a.e.}$$

A function with zero norm need not be identically zero ! Therefore, $||.||_1$ is not a norm on $\mathcal{L}^1$ but only a pseudo-norm.

In order to get a normed space one has to change the space $\mathcal{L}^1$ in such a way that all functions $f = g$ μ-a.e. are considered as equal. Then $f = 0$ μ-a.e. can be considered as the null element of the vector space. The space of integrable functions modified in this way is denoted by $L^1 = L^1(\Omega, \mathcal{F}, P)$.

4.1 Discussion.
For those readers who want to have hard facts instead of soft wellness we provide some details.

For any $f \in \mathcal{L}(\mathcal{F})$ let

$$\tilde{f} = \{g \in \mathcal{L}(\mathcal{F}) : f = g \;\mu\text{-a.e.}\}$$

denote the equivalence class of f. Then integrability is a class-property and the space

$$L^1 := \{\tilde{f} : f \in \mathcal{L}^1\}$$

is a vector space. The value of the integral depends only on the class and therefore it defines a linear function on L^1 having the usual properties. In particular, $||\tilde{f}||_1 := ||f||_1$ defines a norm on L^1.

It is common practise to work with L^1 instead of $\mathcal{L}^1$ but to write f instead of $\tilde{f}$. This is a typical example of what mathematicians call *abuse of language*.

4.2 Theorem. *The space $L^1(\Omega, \mathcal{F}, P)$ is a Banach space.*
Proof: Let (f_n) be a Cauchy sequence in L^1, i.e.

$$\forall \epsilon > 0\ \exists N(\epsilon) \text{ such that } \int |f_n - f_m|\, d\mu < \epsilon \text{ whenever } n, m \geq N(\epsilon).$$

Let $n_i := N(1/2^i)$. Then

$$\int |f_{n_{i+1}} - f_{n_i}|\, d\mu < \frac{1}{2^i}$$

It follows that for all $k \in \mathbb{N}$

$$\int \Big(|f_{n_1}| + |f_{n_2} - f_{n_1}| + \cdots + |f_{n_{k+1}} - f_{n_k}| \Big)\, d\mu \leq C < \infty$$

Hence the corresponding infinite series converges which implies that

$$|f_{n_1}| + \sum_{i=1}^{\infty} |f_{n_{i+1}} - f_{n_i}| < \infty \quad \mu\text{-a.e.}$$

Since absolute convergence of series in $\mathbb{R}$ implies convergence (here completenes of $\mathbb{R}$ goes in) the partial sums

$$f_{n_1} + (f_{n_2} - f_{n_1}) + \cdots + (f_{n_k} - f_{n_{k-1}}) = f_{n_k}$$

converge to some limit f μ-a.e. Mean convergence of (f_n) follows from Fatou's lemma by

$$\begin{aligned}\int |f_n - f|\, d\mu &= \int \lim_{k\to\infty} |f_n - f_{n_k}|\, d\mu \\ &\leq \liminf_{k\to\infty} \int |f_n - f_{n_k}|\, d\mu < \epsilon \text{ whenever } n \geq N(\epsilon).\end{aligned}$$

□

Let

$$\mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{F}, \mu) = \{f \in \mathcal{L}(\mathcal{F}) : \int f^2 d\mu < \infty\}$$

This is another important space of integrable functions.

4.3 Problem. (*easy*)
(a) Show that $\mathcal{L}^2$ is a vector space.

(b) Show that $\int f^2 \boldsymbol{d}\mu < \infty$ is a property of the μ-equivalence class of $f \in \mathcal{L}(\mathcal{F})$.

By $L^1 = L^1(\Omega, \mathcal{F}\mu)$ we again denote the corresponding space of equivalence classes. On this space there is an inner product

$$< f, g >:= \int fg \, d\mu, \quad f, g \in L^2.$$

The corresponding norm is

$$||f||_2 =< f, f >= \Big(\int f^2 d\mu \Big)^{1/2}$$

4.4 Theorem. *The space $L^2(\Omega, \mathcal{F}, \mu)$ is a Hilbert space.*

4.2 Measures with densities

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f \in \mathcal{L}^+(\mathcal{F})$.

4.5 Problem. (*intermediate*)
Show that $\nu : A \mapsto \int_A f \, d\mu$, $A \in \mathcal{F}$ is a measure.

We would like to say that f is the density of ν with respect to μ but for doing so we have to be sure that f is uniquely determined by ν. This is not true, in general.

4.6 Lemma. *Let $f, g \in \mathcal{L}^+(\mathcal{F})$. Then*

$$\int_A f \, d\mu = \int_A g \, d\mu \; \forall A \in \mathcal{F} \;\Rightarrow\; \mu((f \neq g) \cap A) = 0 \; \forall \mu(A) < \infty.$$

In other words: $f = g$ μ-a.e. on every set of finite μ-measure.

Proof: Let $\mu(M) < \infty$ and define $M_n := M \cap (f \leq n) \cap g \leq n)$. Since $f1_{M_n}$ and $g1_{M_n}$ are μ-integrable it follows that $f1_{M_n} = g1_{M_n}$ μ-a.e. For $n \to \infty$ we have $M_n \uparrow M$ which implies $f1_M = g1_M$ μ-a.e. □

Now, the basic uniquenes theorem follows immediately.

4.7 Theorem. *If μ is finite or σ-finite then*

$$\int_A f \, d\mu = \int_A g \, d\mu \; \forall A \in \mathcal{F} \;\Rightarrow\; f = g \; \mu\text{-a.e.}$$

4.8 Definition. Let μ be σ-finite and define

$$\nu : A \mapsto \int_A f \, d\mu, \quad A \in \mathcal{F}.$$

Then $\nu = f\mu$ and $f =: \dfrac{d\nu}{d\mu}$ is called the density or the Radon-Nikodym derivative of ν with respect to μ.

Which measures have densities w.r.t. other measures ?

4.9 Problem. (*easy*)
Let $\nu = f\mu$. Show that $\mu(A) = 0$ implies $\nu(A) = 0$, $A \in \mathcal{F}$.

4.10 Definition. Let $\mu|\mathcal{F}$ and $\nu|\mathcal{F}$ be measures. The measure ν is said to be absolutely continuous w.r.t the measure $\mu|\mathcal{F}$ ($\nu \ll \mu$) if

$$\mu(A) = 0 \;\Rightarrow\; \nu(A) = 0, \quad A \in \mathcal{F}.$$

We saw that absolute continuity is necessary for having a density. It is even sufficient.

4.11 Theorem. *(Radon-Nikodym theorem)*
Assume that μ is σ-finite. If $\nu \ll \mu$ then $\nu = f\mu$ for some $f \in \mathcal{L}^+(\mathcal{F})$.
Proof: See Bauer, [2]. □

We will meet measures with densities frequently when we explore stochastic analysis. Therefore a this point we present two most important special cases.

The first example deals with Lebesgue-Stieltjes measures. Let $\alpha : \mathbb{R} \to \mathbb{R}$ be an increasing function which is supposed to be continuous on $\mathbb{R}$ and differentiable exept of finitely many points. We will show that $\lambda_\alpha = \alpha'\lambda$.

4.12 Problem. (*intermediate*)
Let $\alpha : \mathbb{R} \to \mathbb{R}$ be an increasing function which is supposed to be continuous on $\mathbb{R}$ and to not differentiable at at most finitely many points. Show that $\lambda_\alpha = \alpha'\lambda$.

A density w.r.t the Lebesgue measure is called a Lebesgue density.

4.13 Problem. (*intermediate*)
Let P and Q be probability measures of a finite field $\mathcal{F}$.
(1) State $Q \ll P$ in terms of the generating partition of $\mathcal{F}$.
(2) If $Q \ll P$ find dQ/dP.

Finally, we have to ask how μ-integrals can be transformed into ν-integrals.

4.14 Problem. (*intermediate*)
Let $\nu = f\mu$. Discuss the validity of

$$\int f \, d\nu = \int f \frac{d\nu}{d\mu} d\mu$$

Hint: Prove it for $f \in \mathcal{S}^+(\mathcal{F})$ and extend it by measure theoretic induction.

4.15 Problem.
Let $(\Omega, \mathcal{F}, P)$ be a measure space and X a random variable with differentiable distribution function F. Explain the formulas

$$P(X \in B) = \int_B F'(t)\,dt \quad \text{and} \quad E(g \circ X) = \int g(t)F'(t)\,dt$$

4.3 Iterated integration

UNDER CONSTRUCTION

Part II

Probability theory

Chapter 5

Measure theoretic language of probability

5.1 Basics

Let $(\Omega, \mathcal{F}, P)$ be a probability space. The probability space serves as a model of a *random experiment*. The σ-field $\mathcal{F}$ is the field of *observable* events. Observability of A means that after performing the random experiment it can be decided whether A is realized or not. In this sense the σ-field can be identified with the *information* which is obtained after having performed the random experiment.

The probability measure P gives to each event $A \in \mathcal{F}$ a *probability* $P(A)$. The intuitive nature of the probability will become clear later.

A function $X : \Omega \to \mathbb{R}$ is a *random variable* if assertions about the function value (e.g. $(X \in B)$, $B \in \mathcal{B}$) are observable, i.e. are in $\mathcal{F}$. Therefore, a random variable is simply an $\mathcal{F}$-measurable function.

Let X be a nonnegative or integrable random variable. Then the integral of X is called expectation of X and is denoted by

$$E(X) = \int X \, dP$$

5.2 The information set of random variables

The information set of a random variable X is the family of events which can be expressed in terms of X, i.e. the family of events $X \in B$, $B \in \mathcal{B}$. This is a sub-σ-field of $\mathcal{F}$ and is denoted by $\sigma(X)$. It is called the σ-field generated by X.

5.1 Problem. (*easy*)
Show that $\sigma(X)$ is a σ-field.

5.2 Problem. (*intermediate*)
(a) Let X be an indicator random variable. Find $\sigma(X)$.

(b) Let X be a simple random variable. Find $\sigma(X)$.

Let X and Y be random variables such that $Y = f \circ X$ where f is some Borel-measurable function. Since $(Y \in B) = (X \in f^{-1}(B))$ it follows that $\sigma(Y) \subseteq \sigma(X)$. In other words: If f is causal dependent of X then the information of Y is contained in the information set of X. This is intuitively very plausible: Any assertion about Y can be stated as an assertion about X.

It is a remarkable fact that even the converse is true.

5.3 Theorem. *(Causality theorem)*
Let X and Y be random variables such that $\sigma(Y) \subseteq \sigma(X)$. Then there exists a measurable function f such that $Y = f \circ X$.

Proof: By measure theoretic induction it is sufficient to prove the assertion for $Y = 1_A$, $A \in \mathcal{F}$.

Recall that $\sigma(Y) = \{\emptyset, \Omega, A, A^c\}$. From $\sigma(Y) \subseteq \sigma(X)$ it follows that $A \in \sigma(X)$, i.e. $A = (X \in B)$ for some $B \in \mathcal{B}$. This means $1_A = 1_B \circ X$. □

5.3 Independence

The notion of independence marks the point where probability theory goes beyond abstract measure theory.

Recall that two events $A, B \in \mathcal{F}$ are independent if the product formula $P(A \cap B) = P(A)P(B)$ is true. This is easily extended to families of events.

5.4 Definition. Let $\mathcal{C}$ and $\mathcal{D}$ be subfamilies of $\mathcal{F}$. The families $\mathcal{C}$ and $\mathcal{D}$ are said to be independent (with respect to P) if $P(A \cap B) = P(A)P(B)$ for every choice $A \in \mathcal{C}$ and $B \in \mathcal{D}$.

It is natural to call random variables X and Y independent if the corresponding information sets are independent.

5.5 Definition. Two random variables X and Y are independent if $\sigma(X)$ and $\sigma(Y)$ are independent.

How to check independence of random variables ? Is it sufficient to check the independence of generators of the inforamtion sets ? This is not true, in general, but with a minor modification it is.

5.6 Theorem. *Let X and Y be random variables and let $\mathcal{C}$ and $\mathcal{D}$ be generators of the corresponding information sets. If $\mathcal{C}$ and $\mathcal{D}$ are independent and closed under intersection then X and Y are independent.*

5.7 Problem. (*intermediate*)
Let $F(x, y)$ be the joint distribution function of (X, Y). Show that X and Y are independent iff $F(x, y) = h(x)k(y)$ for some functions h and k.

For independent random variables there is a product formula for expectations.

5.8 Theorem. *(1) Let $X \geq 0$ and $Y \geq 0$ be independent random variables. Then*

$$E(YX) = E(X)E(Y)$$

(2) Let $X \in L^1$ and $Y \in L^1$ be indpendent random variables. Then $XY \in L^1$ and

$$E(YX) = E(X)E(Y)$$

Proof: Apply measure theoretic induction. □

Chapter 6

Conditional expectation

6.1 The concept

Let us explore the relation between a random variable and a σ-field. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{A} \subseteq \mathcal{F}$ be sub-σ-field.

If a random variable X is $\mathcal{A}$-measurable then the information available in $\mathcal{A}$ tells us everything about X. If the random variable X is not $\mathcal{A}$-measurable we could be interested in finding an optimal $\mathcal{A}$-measurable approximation of X in a sense to be specified. This program leads to the concept of conditional expectation.

A successful way consists in decomposing X into a sum $Y + R$ where Y is $\mathcal{A}$-measurable and R is uncorrelated to $\mathcal{A}$. If we require that $E(X) = E(Y)$ then $E(R) = 0$. The condition on R of being uncorrelated to $\mathcal{A}$ means

$$\int_A R\,dP = 0 \quad \text{for all } A \in \mathcal{A}.$$

In other words the approximating variable Y should satisfy

$$\int_A X\,dP = \int_A Y\,dP \quad \text{for all } A \in \mathcal{A}$$

For these integrals to be defined we need nonnegative or integrable random variables.

6.1 Definition. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{A} \subseteq \mathcal{F}$ be sub-σ-field. Let X be a nonnegative or integrable random variable. The *conditional expectation* $E(X|\mathcal{A})$ of X given $\mathcal{A}$ is an $\mathcal{A}$-measurable random variable Y satisfying

$$\int_A X\,dP = \int_A Y\,dP \quad \text{for all } A \in \mathcal{A}$$

6.2 Theorem. *The conditional expectation $E(X|\mathcal{A})$ exists if X is integrable or $X \geq 0$.*

Proof: This is a consequence of the Radon-Nikodym theorem. If $X \geq 0$ then $\mu(A) := \int_A X\,dP$ defines a σ-finite measure on $\mathcal{A}$ such that $\mu \ll P$. Define $E(X|\mathcal{A}) := \dfrac{d\mu}{dP}$. If X is integrable apply the preceding to X^+ and X^-. □

6.3 Problem.
(a) The conditional expectation $E(X|\mathcal{A})$ is uniquely determined P-a.e.
(b) If $X \geq 0$ then $E(X|\mathcal{A}) \geq 0$ P-a.e.
(c) If X is integrable then $E(X|\mathcal{A})$ is integrable, too.

6.4 Problem. $E(E(X|\mathcal{A})) = E(X)$

6.5 Problem. Find the conditional expectation given a finite field.

6.6 Problem. If X is $\mathcal{A}$-measurable then $E(X|\mathcal{A}) = X$.

6.7 Problem. If X is independent of $\mathcal{A}$ then $E(X|\mathcal{A}) = E(X)$.

6.2 Properties

Since the definition of the conditional expectation is linear in X and Y the following two assertions are almost obvious:

(1) Assume that X and Y are nonnegative random variables. Then

$$E(\alpha X + \beta Y|\mathcal{A}) = \alpha E(X|\mathcal{A}) + \beta E(Y|\mathcal{A}) \text{ whenever } \alpha, \beta \geq 0.$$

(2) Assume that X and Y are integrable random variables. Then

$$E(\alpha X + \beta Y|\mathcal{A}) = \alpha E(X|\mathcal{A}) + \beta E(Y|\mathcal{A}) \text{ whenever } \alpha, \beta \in \mathbb{R}.$$

6.8 Theorem. *Iterated conditioning*
UNDER CONSTRUCTION

6.9 Theorem. *Redundant conditioning*
UNDER CONSTRUCTION

Inequalities

6.10 Problem. Show that $X \leq Y$ implies $E(X|\mathcal{A}) \leq E(Y|\mathcal{A})$. Distinguish between the nonnegative and the integrable case.

6.11 Problem. Show that $|E(X|\mathcal{A})| \leq E(|X||\mathcal{A})$ if X is integrable.

6.12 Theorem. *Jensen's inequality*
UNDER CONSTRUCTION

Further topics:

- L^2-hereditary property
- CS-inequality

Projection properties

UNDER CONSTRUCTION

Convergence

6.13 Problem. Show that $X_n \stackrel{L^1}{\to} X$ implies $E(X_n|\mathcal{A}) \stackrel{L^1}{\to} E(X|\mathcal{A})$.

6.14 Problem. Prove a Fatou's lemma for conditional expectations.

6.15 Problem. Prove a Lebesgue's dominated convergence theorem for conditional expectations.

6.3 Calculation

- given a random variable, causality theorem
- given a dominating product measure
- given several random variables
- the Gaussian case

Chapter 7

Stochastic sequences

7.1 The ruin problem

One player

Let us start with a very simple gambling system.

A gambler bets a stake of one unit at subsequent games. The games are independent and p denotes the probability of winning. In case of winning the gambler's return is the double stake, otherwise the stake is lost.

A stochastic model of such a gambling system consists of a probability space $(\Omega, \mathcal{F}, P)$ and a sequence of random variables $(X_i)_{i\geq 1}$. The random variables are independent with values $+1$ and -1 representing the gambler's gain or loss at time $i \geq 1$. Thus, we have $P(X = 1) = p$. The sequence of partial sums, i.e. the accumulated gains,

$$S_n = X_1 + X_2 + \cdots + X_n$$

is called a *random walk* on $\mathbb{Z}$ starting at zero. If $p = 1/2$ then it is a symmetric random walk.

A major foundation problem of probability theory is the question whether there exists a stochastic model for a random walk. We do not discuss such questions but refer to the literature.

Assume the the gambler starts at $i = 0$ with capital $V_0 = a$. Then her wealth after n games is

$$V_n = a + X_1 + X_2 + \cdots + X_n = a + S_n$$

The sequence $(V_n)_{n\geq 0}$ of partial sums is called a *random walk* starting at a.

We assume that the gambler plans to continue gambling until her wealth is $c > a$ or 0. Let

$$T_x := \min\{n : V_n = x\}$$

Then $q_0(a) := P(T_0 < T_c | V_0 = a)$ is called the probability of ruin. Similarly, $q_c(a) := P(T_c < T_0 | V_0 = a)$ is the probability of winning.

How to evaluate the probability of ruin ? It will turn out that the probability can be obtained by studying the dynamic behaviour of the gambling situation. Thus, this is a basic example of a situation which is typical for stochastic analysis: Probabilities are not only obtained by combinatorial methods but also and often much more easily by a dynamic argument resulting in a difference or differential equation.

The starting point is the following assertion.

7.1 Lemma. *The ruin probabilities satisfy the difference equation*

$$q_c(a) = p\, q_c(a+1) + (1-p)\, q_c(a-1) \text{ whenever } 0 < a < c$$

with boundary conditions $q_c(0) = 0$ *and* $q_c(c) = 1$.

It is illuminating to understand the assertion with the help of an heuristic argument: If the random walk starts at $V_0 = a$, $0 < a < c$, then we have $V_1 = a + 1$ with probability p and $V_1 = a - 1$ with probability $1 - p$. This gives

$$P(T_c < T_0|V_0 = a) = pP(T_c < T_0|V_1 = a + 1) + (1 - p)P(T_c < T_0|V_1 = a - 1)$$

However, the random walk starting at time $i = 1$ has the same ruin probabilites as the random walk starting at $i = 0$. This proves the assertion. In this argument we utilized the intuitively obvious fact that the starting time of the random walk does not affect its ruin probabilites.

In order to calculate the ruin probabilities we have to solve the difference equation.

7.2 Discussion. Solving the ruin equation The difference equation

$$x_a = px_{a+1} + (1-p)x_{a-1} \text{ whenever } a = 1, \dots, c-1$$

has the general solution

$$x_a = \begin{cases} A + B\left(\dfrac{1-p}{p}\right)^a & \text{if } p \neq 1/2 \\ A + Ba & \text{if } p = 1/2 \end{cases}$$

The constants A and B are determined by the boundary conditions. This gives

$$q_c(a) = \begin{cases} \dfrac{\left(\dfrac{1-p}{p}\right)^a - 1}{\left(\dfrac{1-p}{p}\right)^c - 1} & \text{if } p \neq 1/2 \\[2ex] \dfrac{a}{c} & \text{if } p = 1/2 \end{cases}$$

In order to calculate $q_0(a)$ we note that $q_0(a) = \bar{q}_c(c-a)$ where $\bar{q}$ denotes the ruin probabilities of a random walk with interchanged transitions probabilities. This implies

$$q_0(a) = \begin{cases} \dfrac{\left(\dfrac{p}{1-p}\right)^{c-a} - 1}{\left(\dfrac{p}{1-p}\right)^{c} - 1} & \text{if} \quad p \neq 1/2 \\ \\ \dfrac{c-a}{c} & \text{if} \quad p = 1/2 \end{cases}$$

Easy calculations show that

$$q_c(a) + q_0(a) = 1$$

which means that gambling ends with probability 1.

7.3 Problem. (*intermediate*)
(a) Fill in the details of solving the difference equation of the ruin problem.
(b) Show that the random walk hits the boundaries almost surely (with probability one).

Two players

Now we assume that two players with initial capitals a and b are playing against each other. The stake of each player is 1 at each game. The game ends when one player is ruined.

This is obviously equivalent to the situation of the preceding section leading to

$$\begin{aligned} P(\text{player 1 wins}) &= q_{a+b}(a) \\ P(\text{player 2 wins}) &= q_0(a). \end{aligned}$$

We know that the game ends with probability one.

Let us turn to the situation where player 1 has unlimited initial capital. Then the game can only end by the ruin of player 2, i.e. if

$$\sup_n S_n \geq b$$

where S_n denotes the accumulated gain of player 1.

7.4 Theorem. *Let (S_n) be a random walk on $\mathbb{Z}$. Then*

$$P(\sup_n S_n \geq b) = \begin{cases} 1 & \textit{whenever } p \geq 1/2 \\ \left(\dfrac{p}{1-p}\right)^b & \textit{whenever } p < 1/2 \end{cases}$$

7.5 Problem. (*advanced*)
Prove 7.4.
Hint: Show that $P(\sup_n S_n \ge b) = \lim_{a\to\infty} q_{a+b}(a)$.

Note that $P(\sup_n S_n \ge 1)$ is the probability that a gambler with unlimited initial capital gains 1 at some time. If $p \ge 1/2$ this happens with probability 1 if we wait sufficiently long. Later we will see that in a fair game ($p = 1/2$) the expected waiting time is infinite.

7.2 Stopping times

Optional stopping

Let us consider the question whether gambling chances can be improved by a gambling system.

Let us start with a particularly simple gambling system, called optional stopping system. The idea is as follows: The gambler waits up to a random time σ and then starts gambling. (The game at period $\sigma + 1$ is the first game to play.) Gambling is continued until a further random time $\tau \ge \sigma$ and then stops. (The game at period τ is the last game to play.) Random times are random variables $\sigma, \tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$.

Now it is an important point that the choice of the starting time σ and the stopping time τ depend only on the information available up to those times since the gambler does not know the future.

Filtrations and stopping times

Let $X_1, X_2, \dots, X_k, \dots$ be a sequence of random variables representing the outcomes of a game at times $k = 1, 2, \dots$.

7.6 Definition. The σ-field $\mathcal{F}_k := \sigma(X_1, X_2, \dots, X_k)$. which is generated by the events $(X_1 \in B_1, X_2 \in B_2, \dots, X_k \in B_k)$, $B_i \in \mathcal{B}$, is called the *past* of the sequence (X_i) at time k.

7.7 Problem. (*advanced*)
State and prove a causality theorem for $\sigma(X_1, X_2, \dots, X_k)$-measurable ransom variables.
Hint: Let $\mathcal{C}$ be the generating system of $\sigma(X_1, X_2, \dots, X_k)$ and let $\mathcal{D}$ be the family of random variables that are functions of $(X_1, X_2, \dots, X_k)$. Show that $\mathcal{D}$ is a σ-field and that $\mathcal{C} \subseteq \mathcal{D}$. This implies that any indicator of a set in $\sigma(X_1, X_2, \dots, X_k)$ is a function of $(X_1, X_2, \dots, X_k)$. Extend this result by measure theoretic induction.

The past at time k is the information set of the beginning $(X_1, X_2, \dots, X_k)$ of the sequence (X_i). The *history* of the game is the family of σ-fields $(\mathcal{F}_k)_{k\ge 0}$ where $\mathcal{F}_0 =$

$\{\emptyset, \Omega\}$. The history is an increasing sequence of σ-fields representing the increasing information in course of time.

7.8 Definition. Any increasing sequence of σ-fields is called a *filtration*.

7.9 Definition. A sequence (X_k) of random variables is *adapted* to a filtration $(\mathcal{F}_k)_{k\geq 0}$ if X_k is $\mathcal{F}_k$-measurable for every $k \geq 0$.

Clearly, every sequence of random variables is adapted to its own history.

Now we are in the position to give a formal definition of a stopping time.

7.10 Definition. Let $(\mathcal{F}_k)$ be a filtration. A random variable $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is a stopping time (relative to the filtration $(\mathcal{F}_k)$) if

$$(\tau = k) \in \mathcal{F}_k \quad \text{for every } k \in \mathbb{N}.$$

7.11 Problem. Let $(\mathcal{F}_k)$ be a filtration and let $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ be a random variable. Show that the following assertions are equivalent:
(a) $(\tau = k) \in \mathcal{F}_k$ for every $k \in \mathbb{N}$
(b) $(\tau \leq k) \in \mathcal{F}_k$ for every $k \in \mathbb{N}$
(c) $(\tau < k) \in \mathcal{F}_{k-1}$ for every $k \in \mathbb{N}$
(d) $(\tau \geq k) \in \mathcal{F}_{k-1}$ for every $k \in \mathbb{N}$
(e) $(\tau > k) \in \mathcal{F}_k$ for every $k \in \mathbb{N}$

7.12 Problem. Let $(X_n)_{n\geq 0}$ be adapted. Show that the *hitting time*

$$\tau = \min\{k \geq 0 : X_k \in B\}$$

is a stopping time for any $B \in \mathcal{B}$. (Note that $\tau = \infty$ if $X_k \notin B$ for all $k \in \mathbb{N}$.)

In view of the causality theorem the realisation of the events $(\tau = k)$ is determined by the values of the random variables $X_1, X_2, \ldots, X_k$, i.e.

$$1_{(\tau=k)} = f_k(X_1, X_2, \ldots, X_k)$$

where f_k are any functions.

Wald's equation

If our gambler applies a stopping system (σ, τ) with finite stopping times then her gain is $S_\tau - S_\sigma$.

7.13 Problem. (*easy*)
Let (X_k) be a sequence adapted to $(\mathcal{F}_k)$ and let τ be a finite stopping time. Then X_τ is a random variable.

Does the stopping system improve the gambler's chances ? We require some preparations.

7.14 Problem. (*easy*)
Let Z be a random variable with values in $\mathbb{N}_0$. Show that $E(Z) = \sum_{k=1}^{\infty} P(Z \geq k)$.

7.15 Theorem. *Wald's equation*
Let (X_k) be an independent sequence of integrable random varibales with a common expectation $E(X_k) = \mu$. If τ is an integrable stopping time then S_τ is integrable and

$$E(S_\tau) = \mu E(\tau)$$

Proof: We will show that the equation is true both for the positive parts and the negative parts of X_k. Let $X_k \geq 0$. Then

$$\begin{aligned} E(S_\tau) &= \sum_{k=1}^{\infty} \int_{\tau=k} S_k \, dP = \sum_{i=1}^{\infty} \int_{\tau \geq i} X_i \, dP \\ &= \sum_{i=1}^{\infty} E(X_i) P(\tau \geq i) = \mu E(\tau) \end{aligned}$$

Note that all terms are ≥ 0 which allows interchanging sums iteration. □

7.16 Problem. (*advanced*)
Let $\tau := \min(k \geq 0 : S_k = -a$ or $S_k = b)$.
(a) Show that Wald's equation is valid.
Hint: Let $\tau_n := \tau \cap n$. Show that τ_n satisfies Wald's equation. Let $n \to \infty$ to obtain integrability of τ.
(b) Calculate $E(\tau)$.

7.17 Problem. (*intermediate*)
Let (S_k) be a symmetric random walk and let $\tau := \min(k \geq 0 : S_k = 1)$. Show that $E(\tau) = \infty$.
Hint: Assume $E(\tau) < \infty$ and derive a contradiction.

The following theorem answers our question for improving chances by stopping strategies. It shows that unfavourable games cannot be turned into fair games and fair games cannot be turned into favourable games. The result is a consequence of Wald's equation and it is the prototype of the fundamental optional stopping theorem of stochastic analysis.

7.18 Theorem. *Optional stopping of random walks*
Let (X_k) be an independent sequence of integrable random variables with a common expectation $E(X_k) = \mu$. Let $\sigma \leq \tau$ be integrable stopping times. Then:

(a) $\mu < 0 \;\Rightarrow\; E(S_\tau - S_\sigma) < 0$.
(b) $\mu = 0 \;\Rightarrow\; E(S_\tau - S_\sigma) = 0$.
(c) $\mu > 0 \;\Rightarrow\; E(S_\tau - S_\sigma) > 0$.

7.3 Gambling systems

Next we generalize our gambling system. We are going to admit that the gambler may vary the stakes. The stopping system is a special case where only 0 and 1 are admitted as stakes.

The stake for game n is denoted by H_n and has to be nonnegative. It is fixed before period n and therefore must be $\mathcal{F}_{n-1}$-measurable since it is determined by the outcomes at times $k = 1, 2, \ldots, n-1$. The sequence of stakes (H_n) is thus not only adapted but even *predictable*.

7.19 Problem.
Determine the stakes H_n corresponding to a stopping system and check predictability.
Hint: Show that $H_n = 1_{(\sigma,\tau]}(n)$.

The gain at game k is $H_k X_k = H_k(S_k - S_{k-1}) = H_k \Delta S_k$. For the wealth of the gambler after n games we obtain

$$V_n = V_0 + \sum_{k=1}^{n} H_k(S_k - S_{k-1}) = V_0 + \sum_{k=1}^{n} H_k \Delta S_k \tag{6}$$

If the stakes are integrable then we have

$$E(V_n) = E(V_{n-1}) + E(H_n)E(X_n).$$

In particular, if $p = 1/2$ we have $E(V_n) = 0$ for all $n \in \mathbb{N}$.

However, if the total gambling time is unbounded (but finite !) then this is no longer true.

7.20 Example. Doubling strategy

Let τ be the waiting time to the first success, i.e.

$$\tau = \min\{k \geq 1 : X_k = 1\}$$

and define

$$H_n := 2^{n-1} 1_{\tau \geq n}$$

Obviously, the stakes are integrable. However, we have

$$P(V_\tau = 1) = 1 \tag{7}$$

for any $p \in (0,1)$. Therefore, a fair game can be transformed into a favourable game by such a strategy. And this is true although the stopping time τ is integrable, actually $E(\tau) = 1/p$!

7.21 Problem.
Prove (7).

The assertion of the optional stopping theorem remains valid for gambling systems if the stopping times are bounded.

7.22 Theorem. *Optional stopping for gambling systems*
Let (X_k) be an independent sequence of integrable random variables with a common expectation $E(X_k) = \mu$. Let (V_n) be the sequence of wealths generated by a gambling system with integrable stakes. If $\sigma \le \tau$ are bounded stopping times then
(a) $\mu < 0 \;\Rightarrow\; E(V_\tau - V_\sigma) < 0$,
(b) $\mu = 0 \;\Rightarrow\; E(V_\tau - V_\sigma) = 0$,
(c) $\mu > 0 \;\Rightarrow\; E(V_\tau - V_\sigma) > 0$.
Proof: Let $N := \max \tau$. Since

$$V_\tau - V_\sigma = \sum_{k=1}^{N} H_k X_k 1_{\sigma < k \le \tau}$$

and since $(\sigma < k \le \tau)$ is independent of X_k it follows that

$$E(V_\tau - V_\sigma) = \mu \sum_{k=1}^{N} E(H_k 1_{\sigma < k \le \tau}).$$

□

7.23 Problem. (*easy*)
Let $p = 1/2$. Show that for the doubling strategy we have $E(V_{\tau \cap n}) = 0$.

7.24 Problem. (*advanced*)
Explain for the doubling strategy, why Lebesgue's theorem on dominated convergence does not imply $E(V_{\tau\cap n}) \to E(V_\tau)$, although $V_{\tau\cap n} \to V_\tau$.
Hint: Show that the sequence $(V_{\tau\cap n})$ is not dominated from below by an integrable random variable.

7.4 Martingales

Let $(X_n)_{n\ge 0}$ be a sequence of integrable random variables adapted to a filtration $(\mathcal{F}_n)_{n\ge 0}$ with $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

In gambler's speech gambling systems are called martingales. This might be the reason for the following mathematical terminology.

7.25 Definition. Let $(\mathcal{F}_k)$ be a filtration and let (Y_k) be an adapted sequence of integrable random variables.

(1) The sequence (Y_k) is called a *martingale* if $E(Y_\sigma) = E(Y_\tau)$ for all bounded stopping times $\sigma \le \tau$.

(2) The sequence (Y_k) is called a *submartingale* if $E(Y_\sigma) \le E(Y_\tau)$ for all bounded stopping times $\sigma \le \tau$.

(3) The sequence (Y_k) is called a *supermartingale* if $E(Y_\sigma) \ge E(Y_\tau)$ for all bounded stopping times $\sigma \le \tau$.

We defined martingales by the property of $\tau \mapsto E(X_\tau)$ being constant for bounded stopping times τ. This property can be rewritten in terms of conditional expectations. We start with a fundamental identity.

7.26 Lemma. *If $\sigma \le \tau$ are bounded stopping times then for any $A \in \mathcal{F}$*

$$\int_A (X_\tau - X_\sigma)dP = \sum_{j=1}^n \int_{A\cap(\sigma<j\le\tau)} (E(X_j|\mathcal{F}_{j-1}) - X_{j-1})dP$$

Proof: Let $\tau \le n$. It is obvious that

$$X_\tau = X_0 + \sum_{j\le\tau}(X_j - X_{j-1}) = X_0 + \sum_{j=1}^n 1_{\tau\ge j}(X_j - X_{j-1})$$

This gives

$$\int_A (X_\tau - X_0)dP = \sum_{j=1}^n \int_{A\cap(\tau\ge j)} (X_j - X_{j-1})dP$$

We may replace X_j by $E(X_j|\mathcal{F}_{j-1})$. □

7.27 Problem. Fill in the details of the proof of 7.26.

7.28 Theorem. *The sequence $(X_n)_{n\ge0}$ is a martingale iff*

$$E(X_j|\mathcal{F}_{j-1}) = X_{j-1}, \quad j \ge 1. \tag{8}$$

Proof: The ”if”-part of the assertion is clear from 7.26.
Let $F \in \mathcal{F}_{j-1}$ and define

$$\tau := \begin{cases} j-1 & \text{whenever } \omega \in F, \\ j & \text{whenever } \omega \notin F. \end{cases}$$

Then τ is a stopping time. From $E(X_j) = E(X_\tau)$ the ”only if”-part follows. □

Equation (8) is the common definition of a martingale.

7.29 Problem. Extend 7.28 to submartingales and supermartingales.

We conclude this section by the elementary version of the celebrated Doob-Meyer decomposition.

7.30 Theorem. *Each adapted sequence $(X_n)_{n\geq 0}$ of integrable random variables can be written as*

$$X_n = M_n + A_n, \quad n \geq 0,$$

where (M_n) is a martingale and (A_n) is a predictable sequence, i.e. A_n is $\mathcal{F}_{n-1}$-measurable for every $n \geq 0$.

The decomposition is unique up to constants.

The sequence $(X_n)_{n\geq 0}$ is a submartingale iff (A_n) is increasing, it is a supermartingale iff (A_n) is decreasing, and it is a martingale iff (A_n) is constant.

Proof: Let

$$M_n = \sum_{j=1}^{n}(X_j - E(X_j|\mathcal{F}_{j-1}))$$

and

$$A_n = \sum_{j-1}^{n}(E(X_j|\mathcal{F}_{j-1}) - X_{j-1})$$

This proves existence of the decomposition. Uniqueness follows from the fact that a predictable martingale is constant. The rest is obvious. □

7.31 Problem. Show that a predictable martingale is constant.

Equation (8) extends easily to stopping times after having defined the past of a stopping time.

7.32 Problem. Let τ be a stopping time. Show that

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : F \cap (\tau \leq j) \in \mathcal{F}_j,\, j \geq 0\}$$

is a σ-field (the *past* of the stopping time τ).

7.33 Theorem. *Optional stopping for martingales*

Let $(X_n)_{n\geq 0}$ be a martingale. Then for any pair $\sigma \leq \tau$ of bounded random variables

$$E(X_\tau|\mathcal{F}_\sigma) = X_\sigma$$

Proof: Applying 7.26 to $A \in \mathcal{F}_\sigma$ proves the assertion. □

7.5 Convergence

UNDER CONSTRUCTION

Chapter 8

The Wiener process

8.1 Basic concepts

In this section we introduce the Wiener process or Brownian Motion process.

A *stochastic process* (random process) on a probability space $(\Omega, \mathcal{F}, P)$ is a family $(X_t)_{t\geq 0}$ of random variables. The parameter t is usually interpreted as time. Therefore, the intuitive notion of a stochastic process is that of random system whose state at time t is X_t.

There are some notions related to a stochastic process $(X_t)_{t\geq 0}$ which are important from the very beginning: the *starting value* X_0, the *increments* $X_t - X_s$ for $s < t$, and the *paths* $t \mapsto X_t(\omega)$, $\omega \in \Omega$.

The Wiener process is defined in terms of these notions.

8.1 Definition. A stochastic process $(W_t)_{t\geq 0}$ is called a *Wiener process* if
(1) the starting value is $W_0 = 0$,
(2) the increments $W_t - W_s$ are $N(0, t-s)$-distributed and mutually independent for non-overlapping intervals,
(3) the paths are continuous for P-almost all $\omega \in \Omega$.

8.2 Remark. As it is the case with every probability model one has to ask whether there exist a probability space $(\Omega, \mathcal{F}, P)$ and a family of random variables (W_t) satisfying the properties of Definition 8.1. The mathematical construction of such models is a complicated matter and is one of great achievements of probability theory in the first half of the 20th century. Accepting the existence of the Wiener process as a valid mathematical model we may forget the details of the construction (there are several of them) and start with the axioms stated in 8.1. (Further reading: Karatzas-Shreve [15], section 2.2.)

8.3 Discussion. Wiener process as random walk
Later we will show:

Any process $(X_t)_{t\ge 0}$ with continuous paths and independent increments such that $E(X_t - X_s) = 0$ and $V(X_t - X_s) = t - s$ is necessarily a Wiener process.

This means that there are three structural properties which are essential for the concepts of a Wiener process:
(1) The process has independent increments.
(2) The expectation of the increments is zero.
(3) The variance of the increments is proportional to the length of the time interval.
Let us motivate these properties at hand of specific discrete time models.

Let $X_1, X_2, \ldots$ be independent and such that $P(X_i = 1) = P(X_i = -1) = 1/2$. Then

$$S_n = X_1 + X_2 + \ldots + X_n \text{ whenever } n = 1, 2, \ldots$$

is called a *symmetric random walk*. It is easy to see that the increments $S_n - S_m$ are independent. Moreover, we have

$$E(X_i) = 0 \;\Rightarrow\; E(S_n - S_m) = 0 \quad \text{and} \quad \mathrm{V}(X_i) = 1 \;\Rightarrow\; \mathrm{V}(S_n - S_m) = n - m$$

Thus, the Wiener process can be interpreted as a continuous time version of a symmetric random walk.

8.4 Problem. (*easy*) Let $(W_t)_{t\ge 0}$ be a Wiener process. Show that $X_t := -W_t$, $t \ge 0$, is a Wiener process, too.

8.5 Problem. (*intermediate*) Show that $W_t/t \xrightarrow{P} 0$ as $t \to \infty$.

8.6 Definition. The *past* of a process $(X_t)_{t\ge 0}$ at time t is the σ-field of events $\mathcal{F}_t^X = \sigma(X_s : s \le t)$ generated by variables X_s of the process prior to t, i.e. $s \le t$. The *internal history* of $(X_t)_{t\ge 0}$ is the family $(\mathcal{F}_t^X)_{t\ge 0}$ of pasts of the process.

The intuitive idea behind the concept of past is the following: $\mathcal{F}_t^X$ consists of all events which are observable if one observes the process up to time t. It represents the information about the process available at time t. It it obvious that $t_1 < t_2 \;\Rightarrow\; \mathcal{F}_{t_1}^X \subseteq \mathcal{F}_{t_2}^X$. If X_0 is a constant then $\mathcal{F}_0^X = \{\emptyset, \Omega\}$.

Independence of increments is actually a much stronger property than it sounds.

8.7 Theorem. *The increments $W_t - W_s$ of a Wiener process $(W_t)_{t\ge 0}$ are independent of the past $\mathcal{F}_s^W$.*

Proof: Let $s_1 \le s_2 \le \ldots \le s_n \le s < t$. Then the random variables

$$W_{s_1}, W_{s_2} - W_{s_1}, \ldots, W_{s_n} - W_{s_{n-1}}, W_t - W_s$$

are independent. It follows that even the random variables $W_{s_1}, W_{s_2}, \ldots, W_{s_n}$ are independent of $W_t - W_s$. Since this is valid for any choice of time points $s_i \le s$ the independence assertion carries over to the whole past $\mathcal{F}_t^W$. $\square$

8.2 Quadratic variation

For beginners the most surprising properties are the path properties of a Wiener process.

The paths of a Wiener process are continuous (which is part of our definition). In this respect the paths seem to be not complicated since they have no jumps or other singularities. It will turn out, however, that in spite of their continuity, the paths of a Wiener process are of a very peculiar nature.

8.8 Remark. (This remark is based on section 10.1.)

Recall that for a function $f : [0, \infty) \to \mathbb{R}$ with bounded variation on compacts we have

$$\lim \sum_{i=1}^{n} |f(t_i^n) - f(t_{i-1}^n)| = V_0^t(f) < \infty$$

for each Riemannian sequence of subdivisions $0 = t_0^n < t_1^n < \ldots < t_n^n = t$ and every fixed $t > 0$. Recall that all smooth (continuously differentiable) functions are of bounded variation on compacts. For such functions it follows that their *quadratic variation*, defined as

$$\lim \sum_{i=1}^{n} |f(t_i^n) - f(t_{i-1}^n)|^2$$

by Riemannian sequences, is necessarily zero for every $t > 0$.

8.9 Problem. (*easy for mathematicians*) Show that the quadratic variation of a continuous BV-function is zero on every compact interval.

We will now show that (almost) all paths of a Wiener process have nonvanishing quadratic variation. This implies that the paths cannot be smooth. Actually, it can be shown that they are nowhere differentiable. (Further reading: Karatzas-Shreve [15], section 2.9.)

8.10 Theorem. *Let $(W_t)_{t \geq 0}$ be a Wiener process. For every $t > 0$ and every Riemannian sequence of subdivisions $0 = t_0^n < t_1^n < \ldots < t_n^n = t$*

$$\sum_{i=1}^{n} |W(t_i^n) - W(t_{i-1}^n)|^2 \xrightarrow{P} t, \quad t > 0.$$

8.11 Problem. (*intermediate*) Prove 8.10.

Hint: Let $Q_n := \sum_{i=1}^{n} |W(t_i^n) - W(t_{i-1}^n)|^2$ for a particular Riemannian sequence of subdivisions. Show that $E(Q_n) = t$ and $V(Q_n) \to 0$. Then the assertion follows from Cebysev's inequality.

The assertion of 8.10 can be improved to P-almost sure convergence which implies that the quadratic variation on $[0,t]$ of almost all paths is actually t. It is remarkable that the quadratic variation of the Wiener process is a deterministic function of a very simple nature.

8.3 Martingales

We start with some general definitions.

8.12 Definition. Any increasing family of σ-fields $(\mathcal{F}_t)_{t\ge 0}$ is called a *filtration*. A process $(Y_t)_{t\ge 0}$ is *adapted* to the filtration $(\mathcal{F}_t)_{t\ge 0}$ if Y_t is $\mathcal{F}_t$-measurable for every $t \ge 0$.

The internal history $(\mathcal{F}_t^X)_{t\ge 0}$ of a process $(X_t)_{t\ge 0}$ is a filtration and the process $(X_t)_{t\ge 0}$ is adapted to its internal history. But also $Y_t := \phi(X_t)$ for any measurable function ϕ is adapted to the internal history of $(X_t)_{t\ge 0}$. Adaption simply means that the past of the process $(Y_t)_{t\ge 0}$ at time t is contained in $\mathcal{F}_t$. Having the information contained in $\mathcal{F}_t$ we know everything about the process up to time t.

8.13 Definition. A *martingale* relative to the filtration $(\mathcal{F}_t)_{t\ge 0}$ is an adapted and integrable stochastic process $(X_t)_{t\ge 0}$ such that

$$E(X_t|\mathcal{F}_s) = X_s \text{ whenever } s < t$$

It is a *square integrable martingale* if $E(X_t^2) < \infty$, $t \ge 0$.

8.14 Problem. (*easy*)
Show that the martingale property remains valid if the filtration is replaced by another filtration consisting of smaller σ-fields, provided that the process is still adapted.

Let us consider some important martingales related to the Wiener process. Let $(\mathcal{F}_t)_{t\ge 0}$ be the internal history of the Wiener process $(W_t)_{t\ge 0}$.

8.15 Theorem. *A Wiener process is a square integrable martingale with respect to its internal history.*

Proof: Since $W_t - W_s$ is independent of $\mathcal{F}_s$ it follows that $E(W_t - W_s|\mathcal{F}_s) = E(W_t - W_s) = 0$. Hence $E(W_t|\mathcal{F}_s) = E(W_s|\mathcal{F}_s) = W_s$. □

A nonlinear function of a martingale typically is not a martingale. But the next theorem is a first special case of a very general fact: It is sometimes possible to correct a process by a bounded variation process in such a way that the result is a martingale.

8.16 Theorem. *The process $W_t^2 - t$ is a square integrable martingale with respect to the internal history of the driving Wiener process $(W_t)_{t\geq 0}$.*

Proof: Note that

$$W_t^2 - W_s^2 = (W_t - W_s)^2 + 2W_s(W_t - W_s)$$

This gives

$$E(W_t^2 - W_s^2|\mathcal{F}_s) = E((W_t - W_s)^2|\mathcal{F}_s) + 2E(W_s(W_t - W_s)|\mathcal{F}_s) = t - s$$

□

The assertion of 8.16 can be written in the following way: For $X_t = W_t^2$ there is a decomposition $X_t = M_t + A_t$ where $(M_t)_{t\geq 0}$ is a martingale and $(A_t)_{t\geq 0}$ is a bounded variation process. Such a decomposition is a mathematical form of the idea that a process X_t is the sum of a (rapidely varying) noise component and a (slowly varying) trend component.

8.17 Problem. (*easy*) Let $(X_t)_{t\geq 0}$ be any process with independent increments such that $E(X_t) = 0$ and $E(X_t^2) = t$. Show that assertions 8.15 and 8.16 are valid for $(X_t)_{t\geq 0}$.

8.18 Theorem. *The process $\exp(aW_t - a^2t/2)$ is a martingale with respect to the internal history of the driving Wiener process $(W_t)_{t\geq 0}$.*

Proof: Use $e^{aW_t} = e^{a(W_t - W_s)}e^{aW_s}$ to obtain

$$E(e^{aW_t}|\mathcal{F}_s) = E(e^{a(W_t - W_s)})e^{aW_s} = e^{a^2(t-s)/2}e^{aW_s}$$

□

The process

$$\mathcal{E}(W)_t := \exp(W_t - t/2)$$

is called the *exponential martingale* of $(W_t)_{t\geq 0}$.

UNDER CONSTRUCTION: MAXIMAL INEQUALITY

8.19 Problem. (*intermediate*) Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. Let $Q \sim P$ be an equivalent probability measure. Denote $P_t := P|\mathcal{F}_t$ and $Q_t := Q|\mathcal{F}_t$.

(a) Show that $P_t \sim Q_t$ for every $t \geq 0$.

(b) Show that $\dfrac{dQ_t}{dP_t} = E\Big(\dfrac{dQ}{dP}\Big|\mathcal{F}_t\Big)$ for every $t \geq 0$.

(c) Show that the process $\Big(\dfrac{dQ_t}{dP_t}\Big)$ is a positive martingale such that $\dfrac{dQ_t}{dP_t} > 0$ P-a.e. for every $t \geq 0$.

(d) Prove the so-called „Bayes-formula":

$$E_Q(X|\mathcal{F}_t) = \frac{E_P(X\frac{dQ}{dP}|\mathcal{F}_t)}{E_P(\frac{dQ}{dP}|\mathcal{F}_t)}$$

whenever $X \geq 0$ or $X \in L_1(Q)$.

8.20 Problem. (*intermediate*) Let $(W_t)_{t\geq 0}$ be a Wiener process on a probability space $(\Omega, \mathcal{F}, P)$ and $(\mathcal{F}_t)_{t\geq 0}$ its internal history. Define

$$\frac{dQ_t}{dP_t} := e^{aW_t - a^2 t/2},\ t \geq 0.$$

(a) Show that there is a uniquely determined probability measure $Q|\mathcal{F}_\infty$ such that $Q|\mathcal{F}_t = Q_t$, $t \geq 0$.

(b) Show that $Q|\mathcal{F}_\infty$ is equivalent to $P|\mathcal{F}_\infty$.

(c) Show that $\widetilde{W}_t := W_t - at$ is a Wiener process under Q.

Hint: Prove that for $s < t$

$$E_Q\left(e^{\lambda(\widetilde{W}_t - \widetilde{W}_s)}\middle|\mathcal{F}_s\right) = e^{\lambda^2(t-s)/2}.$$

8.4 Stopping times

Let $(X_t)_{t\geq 0}$ be a right continuous adapted process such that $X_0 = 0$ and for some $a > 0$ let

$$\tau = \inf\{t \geq 0 : X_t \geq a\}$$

The random variable τ is called a *first passage time*: It is the time when the process hits the level a for the first time. By right continuity of the paths we have

$$\tau \leq t \ \Leftrightarrow\ \max_{s\leq t} X_s \geq a \tag{9}$$

Thus, we have $(\tau \leq t) \in \mathcal{F}_t$ for all $t \geq 0$.

8.21 Problem. (*easy for mathematicians*) Prove (9).

8.22 Definition. A random variable $\tau : \Omega \to [0, \infty]$ is called a *stopping time* if $(\tau \leq t) \in \mathcal{F}_t$ for all $t \geq 0$.

8.23 Problem. (*intermediate for mathematicians*) Show that every bounded stopping time τ is limit of a decreasing sequence of bounded stopping times each of which has only finitely many values.

Hint: Let $T = \max \tau$. Define $\tau_n = k/2^n$ whenever $(k-1)/2^n < \tau \le k/2^n$, $k = 0, \ldots, T2^n$.

Let $(M_t)_{t\ge 0}$ be a martingale. Then we have $E(M_t) = E(M_0)$ for every $t \ge 0$. This can be extended to stopping times. The following is the simplest form of the famous optional stopping theorem.

8.24 Theorem. *(Optional stopping theorem) Let $(M_t)_{t\ge 0}$ be a martingale with right-continuous paths and let τ be a bounded stopping time. Then $E(M_\tau) = E(M_0)$.*

Proof: Assume that $\tau \le T$. Let $\tau_n \downarrow \tau$ where τ_n are bounded stopping times with finitely many values. Then it follows from the discrete version of the optional stopping theorem that $E(M_{\tau_n}) = E(M_0)$. Clearly we have $M_{\tau_n} \to M_\tau$. Since $E(M_{\tau_n}) = E(M_T|\mathcal{F}_{\tau_n})$ the sequence (M_{τ_n}) is uniformly integrable and the assertion follows. □

8.25 Problem. (*intermediate*) Let $(X_t)_{t\ge 0}$ be an integrable process adapted to a filtration $(\mathcal{F}_t)_{t\ge 0}$. Show that (a) implies (b):
(a) $E(X_\sigma) = E(X_0)$ for all bounded stopping times σ.
(b) $(X_t)_{t\ge 0}$ is a martingale.
Hint: For $s < t$ and $F \in \mathcal{F}_s$ define

$$\tau := \begin{cases} s & \text{whenever } \omega \in F, \\ t & \text{whenever } \omega \notin F. \end{cases}$$

Then τ is a stopping time. From $E(X_t) = E(X_\tau)$ the assertion follows.

First passage times of the Wiener process

As an application of the optional stopping theorem we derive the distribution of first passage times of the Wiener process.

8.26 Theorem. *Let $(W_t)_{t\ge 0}$ be a Wiener process and for $a > 0$ and $b \in \mathbb{R}$ define*

$$\tau_{a,b} := \inf\{t : W_t \ge a + bt\}$$

Then we have

$$E(e^{-\lambda\tau_{a,b}} 1_{(\tau_{a,b}<\infty)}) = e^{-a(b+\sqrt{b^2+2\lambda})}, \ \lambda \ge 0$$

Proof: Applying the optional stopping theorem to the exponential martingale of the Wiener process we get

$$E(e^{\theta W_\tau - \theta^2\tau/2}) = 1$$

for every $\theta \in \mathbb{R}$ and every bounded stopping time τ. Therefore this equation is true for $\tau_n := \tau_{a,b} \cap n$ for every $n \in \mathbb{N}$. We note that (use 8.5)

$$e^{\theta W_{\tau_n} - \theta^2\tau_n/2} \xrightarrow{P} e^{\theta W_{\tau_{a,b}} - \theta^2\tau_{a,b}/2} 1_{(\tau_{a,b}<\infty)}$$

Applying the dominated convergence theorem it follows (at least for sufficiently large θ) that

$$E(e^{\theta W_{\tau_{a,b}}-\theta^2\tau_{a,b}/2}1_{(\tau_{a,b}<\infty)})=1$$

The rest are easy computations. Since $W_{\tau_{a,b}}=a+b\tau_{a,b}$ we get

$$E(e^{(\theta b-\theta^2/2)\tau_{a,b}}1_{(\tau_{a,b}<\infty)})=e^{-ab}$$

Putting $\lambda:=-\theta b+\theta^2/2$ proves the assertion. □

8.27 Problem. (*advanced*)
Fill in the details of the proof of 8.26.

8.28 Problem. (*easy*)
In the following problems treat the cases $b>0$, $b=0$ and $b<0$ separately.
(a) Find $P(\tau_{a,b}<\infty)$.
(b) Find $E(\tau_{a,b})$.

8.29 Problem. (*easy*)
(a) Does the assertion of the optional sampling theorem hold for the martingale $(W_t)_{t\geq 0}$ and $\tau_{a,b}$?
(b) Does the assertion of the optional sampling theorem hold for the martingale W_t^2-t and $\tau_{a,b}$?

8.30 Problem. (*intermediate*)
(a) Show that $P(\tau_{0,b}=0)=1$ for every $b>0$. (Consider $E(e^{-\lambda\tau_{a_n,b}})$ for $a_n\downarrow 0$.) Give a verbal interpretation of this result.
(b) Show that $P(\max_t W_t=\infty,\ \min_t W_t=-\infty)=1$.
(c) Conclude from (a) that almost all paths of $(W_t)_{t\geq 0}$ infinitely often cross every horizontal line.

From 8.26 we obtain the distribution of the first passage times.

8.31 Corollary. *Let $\tau_{a,b}$ be defined as in 8.26. Then*

$$P(\tau_{a,b}\leq t)=1-\Phi\Big(\frac{a+bt}{\sqrt{t}}\Big)+e^{-2ab}\Phi\Big(\frac{-a+bt}{\sqrt{t}}\Big),\ t\geq 0$$

Proof: Let $G(t):=P(\tau_{a,b}\leq t)$ and let $F_{a,b}(t)$ denote the right hand side of the asserted equation. We want to show that $F_{a,b}(t)=G(t)$, $t\geq 0$. For this we will apply the uniqueness of the Laplace transform. Note that 8.26 says that

$$\int_0^\infty e^{-\lambda t}dG(t)=e^{-a(b+\sqrt{b^2+2\lambda})},\ t\geq 0$$

Therefore, we have to show that

$$\int_0^\infty e^{-\lambda t} dF_{a,b}(t) = e^{-a(b+\sqrt{b^2+2\lambda})}, \ t \geq 0$$

This is done by the following simple calculations. First, it is shown that

$$F_{a,b}(t) = \frac{1}{\sqrt{2\pi}} \int_0^t \frac{a}{s^{3/2}} \exp\Big(-\frac{a^2}{2s} - \frac{b^2 s}{2} - ab \Big)\, ds$$

(This is done by calculating the derivatives of both sides.) Then it follows that

$$e^{ab} \int_0^t e^{-\lambda s} dF_{a,b}(s) = e^{a\sqrt{b^2+2\lambda}} F_{a,\sqrt{b^2+2\lambda}}(t)$$

Putting $t = \infty$ the assertion follows. □

8.32 Problem. (*requires calculation skills*)
Fill in the details of the proof of 8.31.

8.33 Problem. (*easy*)
Find the distribution of $\max_{s \leq t} W_s$.

The following problems are concerned with first passage times for two horizontal boundaries. Let $c, d > 0$ and define

$$\sigma_{c,d} = \inf\{t : W_t \notin (-c, d)\}$$

8.34 Problem. (*easy*)
(a) Show that $\sigma_{c,d}$ is a stopping time.
(b) Show that $P(\sigma_{c,d} < \infty) = 1$.

For $\sigma_{c,d}$ the application of the optional sampling theorem is straighforward since $|W_t| \leq \max\{c, d\}$ for $t \leq \sigma_{c,d}$.

8.35 Problem. (*easy*)
Find the distribution of $W_{\sigma_{c,d}}$.
Hint: Note that $E(W_{\sigma_{c,d}}) = 0$ (why ?) and remember that $W_{\sigma_{c,d}}$ has only two different values.
Solution: $P(W_{\sigma_{c,d}} = -c) = \frac{d}{c+d}$, $P(W_{\sigma_{c,d}} = d) = \frac{c}{c+d}$

8.36 Problem. (*easy*)
Find $E(\sigma_{c,d})$.
Hint: Note that $E(W^2_{\sigma_{c,d}}) = E(\sigma_{c,d})$ (why ?).
Solution: $E(\sigma_{c,d}) = cd$.

8.37 Discussion. Distribution of $\sigma_{c,d}$

The distribution of the stopping time $\sigma_{c,d}$ is a more complicated story. It is easy to obtain the Laplace transforms. Obtaining probabilistic information requires much more analytical efforts.

For reasons of symmetry we have

$$A := \int_{W_{\sigma_{c,d}}=-c} e^{-\theta^2\sigma_{c,d}/2} dP = \int_{W_{\sigma_{d,c}}=c} e^{-\theta^2\sigma_{d,c}/2} dP$$

and

$$B := \int_{W_{\sigma_{c,d}}=d} e^{-\theta^2\sigma_{c,d}/2} dP = \int_{W_{\sigma_{d,c}}=-d} e^{-\theta^2\sigma_{d,c}/2} dP$$

From

$$1 = E\Big(e^{\theta W_{\sigma_{c,d}} - \theta^2\sigma_{c,d}/2}\Big) \quad \text{and} \quad 1 = E\Big(e^{\theta W_{\sigma_{d,c}} - \theta^2\sigma_{d,c}/2}\Big)$$

we obtain a system of equations for A and B leading to

$$A = \frac{e^{\theta d} - e^{-\theta d}}{e^{\theta(c+d)} - e^{-\theta(c+d)}} \quad \text{and} \quad B = \frac{e^{\theta c} - e^{-\theta c}}{e^{\theta(c+d)} - e^{-\theta(c+d)}}$$

This implies

$$E(e^{-\lambda\sigma_{c,d}}) = \frac{e^{-c\sqrt{2\lambda}} + e^{-d\sqrt{2\lambda}}}{1 + e^{-(c+d)\sqrt{2\lambda}}}$$

Expanding this into an infinite geometric series and applying

$$\int_0^\infty e^{-\lambda t} dF_{a,0}(t) = e^{-a\sqrt{2\lambda}},\ t \geq 0$$

we could obtain an infinite series expansion of the distribution of $\sigma_{c,d}$.
(Further reading: Karatzas-Shreve [15], section 2.8.)

The reflection principle

Let $(W_t)_{t\geq 0}$ be a Wiener process and let $(\mathcal{F}_t)_{t\geq 0}$ be its internal history.

Let $s > 0$ and consider the process $X_t := W_{s+t} - W_s$, $t \geq 0$. Since the Wiener process has independent increments the process $(X_t)_{t\geq 0}$ is independent of $\mathcal{F}_s$. Moreover, it is easy to see that $(X_t)_{t\geq 0}$ is a Wiener process. Let us give an intuitive interpretation of these facts.

Assume that we observe the Wiener process up to time s. Then we know the past $\mathcal{F}_s$ and the value W_s at time s. What about the future ? How will the process behave for $t > s$? The future variation of the process after time s is give by $(X_t)_{t\geq 0}$. From the remarks above it follows that the future variation is that of a Wiener process which is independent of the past. The common formulation of this fact is: At every time $s > 0$ the Wiener process starts afresh.

8.38 Problem. (*easy*)
Show that the process $X_t := W_{s+t} - W_s$, $t \geq 0$ is a Wiener process for every $s \geq 0$.

There is a simple consequence of the property of starting afresh at every time s. Note that

$$W_t = \begin{cases} W_t & \text{whenever } t \leq s \\ W_s + (W_t - W_s) & \text{whenever } t > s \end{cases}$$

Define the corresponding process reflected at time s by

$$\widetilde{W_t} = \begin{cases} W_t & \text{whenever } t \leq s \\ W_s - (W_t - W_s) & \text{whenever } t > s \end{cases}$$

Then it is clear that $(W_t)_{t\geq 0}$ and $(\widetilde{W_t})_{t\geq 0}$ have the same distribution. This assertion looks rather harmless and self-evident. However, it becomes a powerful tool when it is extended to stopping times.

8.39 Theorem. *(Reflection principle)*
Let τ be any stopping time and define

$$\widetilde{W_t} = \begin{cases} W_t & \textit{whenever } t \leq \tau \\ W_\tau - (W_t - W_\tau) & \textit{whenever } t > \tau \end{cases}$$

Then the distributions of $(W_t)_{t\geq 0}$ and $(\widetilde{W_t})_{t\geq 0}$ are equal.

Proof: Let us show that the single random variables W_t and $\widetilde{W_t}$ have equal distributions. Equality of the finite dimensional marginal distributions is shown in a similar manner.

We have to show that for any bounded continuous function f we have $E(f(W_t)) = E(f(\widetilde{W_t}))$. For obvious reasons we need only show

$$\int_{\tau<t} f(W_t)\, dP = \int_{\tau<t} f(\widetilde{W_t})\, dP$$

which is equivalent to

$$\int_{\tau<t} f(W_\tau + (W_t - W_\tau))\, dP = \int_{\tau<t} f(W_\tau - (W_t - W_\tau))\, dP$$

The last equation is obviously true for stopping times with finitely many values. The common approximation argument then proves the assertion. □

8.40 Problem. (*advanced*)
To get an idea of how the full proof of the reflection principle works show $E(f(W_{t_1}, W_{t_2})) = E(f(\widetilde{W}_{t_1}, \widetilde{W}_{t_2}))$ for $t_1 < t_2$ and bounded continuous f.
Hint: Distinguish between $\tau < t_1$, $t_1 \leq \tau < t_2$ and $t_2 \leq \tau$.

The reflection principle offers an easy way for obtaining information on first passage times.

8.41 Theorem. *Let* $M_t := \max_{s \le t} W_s$. *Then*

$$P(M_t \ge y, W_t < y - x) = P(W_t > y + x), \quad t > 0,\, y > 0,\, x \ge 0$$

Proof: Let $\tau := \inf\{t : W_t \ge y\}$ and $\tilde{\tau} := \inf\{t : \widetilde{W}_t \ge y\}$. Then

$$\begin{aligned} P(M_t \ge y, W_t < y - x) &= P(\tau \le t, W_t < y - x) \\ &= P(\tilde{\tau} \le t, \widetilde{W}_t < y - x) \\ &= P(\tau \le t, W_t > y + x) \\ &= P(W_t > y + x) \end{aligned}$$

□

8.42 Problem. (*easy*)
Use 8.41 to find the distribution of M_t.

8.43 Problem. (*intermediate*)
Find $P(W_t < z, M_t < y)$ when $z < y$, $y > 0$.

8.5 Augmentation

For technical reasons which will become clear later the internal history of the Wiener process is slightly too small. It is convenient to increase the σ-fields of the internal history in a way that does not destroy the basic properties of the underlying process. This procedure is called augmentation.

8.44 Definition. Let $\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$ and define

$$\overline{\mathcal{F}}_t := \{F \in \mathcal{F}_\infty : P(F \triangle G) = 0 \text{ for some } G \in \mathcal{F}_{t+}\}$$

Then $(\overline{\mathcal{F}}_t)_{t \ge 0}$ is the *augmented filtration*.

8.45 Problem. (*intermediate*)
Show that the augmented filtration is really a filtration.

8.46 Corollary. *Let* $(W_t)_{t \ge 0}$ *be a Wiener process. Then the increments* $W_t - W_s$ *are independent of* $\overline{\mathcal{F}}_s^W$, $s \ge 0$.

Proof: (Outline) It is easy to see that

$$E(e^{a(W_t - W_s)} | \mathcal{F}_{s+}^W) = e^{(t-s)/2}$$

This implies

$$E(1_F e^{a(W_t-W_s)}) = P(F)E(e^{a(W_t-W_s)})$$

for every $F \in \mathcal{F}^W_{s+}$. From the totality of exponentials property (see the proof of 13.11) it follows that $W_t - W_s$ is independent of $\mathcal{F}^W_{s+}$. It is clear that this carries over to $\overline{\mathcal{F}}_s$. □

8.47 Problem. (*intermediate for mathematicians*)
Fill in the details of the proof of 8.46,

8.48 Theorem. *Let $(\mathcal{F}_t)_{t\ge0}$ be a filtration. Then the augmented filtration is right-continuous, i.e.*

$$\overline{\mathcal{F}}_t = \bigcap_{s>t} \overline{\mathcal{F}}_s$$

Proof: It is clear that $\subseteq$ holds. In order to prove $\supseteq$ let $F \in \bigcap_{s>t} \overline{\mathcal{F}}_s$. We have to show that $F \in \overline{\mathcal{F}}_t$.

For every $n \in \mathbb{N}$ there is $G_n \in \mathcal{F}_{t+1/n}$ such that $P(F \triangle G_n) = 0$. Define

$$G := \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} G_n = \bigcap_{m=K}^{\infty} \bigcup_{n=m}^{\infty} G_n \in \mathcal{F}_{t+1/K} \text{ for all } K \in \mathbb{N}.$$

Then $G \in \mathcal{F}_{t+}$ and $P(G \triangle F) = 0$. □

One says that a filtration satisfies the *"usual conditions"* if it is right-continuous and contains all negligible sets of $\mathcal{F}_\infty$. The internal history of the Wiener process does not satisfy the usual conditions. However, every augmented filtration satisfies the usual conditions. Thus, 8.46 and 8.48 show that every Wiener process has independent increments with respect to a filtration that satisfies the usual conditions. When we are dealing with a Wiener process we may suppose that the underlying filtration satisfies the usual conditions.

8.49 Problem. (*easy*)
Show that the assertions of 8.15, 8.16 and 8.18 are valid for the augmented internal history of the Wiener process.

Let us illustrate the convenience of filtrations satisfying the usual conditions by a further result. For some results on stochastic integrals it will be an important point that martingales are cadlag. A general martingale need not be cadlag. We will show that a martingale has a cadlag modification if the filtration satisfies the usual conditions.

8.50 Theorem. *Let $(X_t)_{t\ge0}$ be a martingale w.r.t. a filtration satisfying the usual conditions. Then there is a cadlag modification of $(X_t)_{t\ge0}$.*

Proof: (Outline. Further reading: Karatzas-Shreve, [15], Chapter 1, Theorem 3.13.)

We begin with path properties which are readily at hand: There is a set $A \in \mathcal{F}_\infty$, satisfying $P(A) = 1$ and such that the restricted process $(X_t)_{t \in \mathbb{Q}}$ has paths with right and left limits for every $\omega \in A$. This is a consequence of the upcrossings inequality by Doob. See Karatzas-Shreve, [15], Chapter 1, Proposition 3.14, (i).

It is now our goal to modify the martingale in such a way that it becomes cadlag. The idea is to define

$$X_t^+ := \lim_{s \downarrow t, s \in \mathbb{Q}} X_s, \quad t \ge 0.$$

on A and $X_t^+ := 0$ elsewhere. It is easy to see that the paths of $(X_t^+)_{t \ge 0}$ are cadlag. Since $(\mathcal{F}_t)_{t \ge 0}$ satisfies the usual conditions it follows that $(X_t^+)_{t \ge 0}$ is adapted. We have to show that $(X_t^+)_{t \ge 0}$ is a modification of $(X_t)_{t \ge 0}$, i.e. $X_t = X_t^+$ P-a.e. for all $t \ge 0$.

Let $s_n \downarrow t$, $(s_n) \subseteq \mathbb{Q}$. Then $X_{s_n} = E(X_{s_1}|\mathcal{F}_{s_n})$ is uniformly integrable which implies $X_{s_n} \stackrel{L^1}{\to} X_t^+$. From $X_t = E(X_{s_n}|\mathcal{F}_t)$ we obtain $X_t = E(X_t^+|\mathcal{F}_t) = X_t^+$ P-a.e. □

8.6 More on stopping times

The interplay between stopping times and adapted processes is at the core of stochastic analysis. In this section we try to provide a lot of information for reasons of later reference. We will state most of the assertions as exercises with hints if necessary. Throughout the section we assume tacitely that the filtration satisfies the usual conditions.

Further reading: Karatzas-Shreve, [15], Chapter 1, section 1.2.

Let τ be a stopping time. The intuitive meaning of $(\tau \le t) \in \mathcal{F}_t$ is as follows: At every time t it can be decided whether $\tau \le t$ or not.

8.51 Problem. (*intermediate*)
Show that τ is a stopping time iff $(\tau < t) \in \mathcal{F}_t$ for every $t \ge 0$.

8.52 Problem. (*easy*)
Let σ, τ and τ_n be stopping times.
(a) Then $\sigma \cap \tau$, $\sigma \cup \tau$ and $\sigma + \tau$ are stopping times.
(b) $\tau + \alpha$ for $\alpha \ge 0$ and $\lambda\tau$ for $\lambda \ge 1$ are stopping times.
(c) $\sup_n \tau_n$, $\inf_n \tau_n$ are stopping times.

Let $(X_t)_{t \ge 0}$ be a process adapted to a filtration $(\mathcal{F}_t)_{t \ge 0}$ and let $A \subseteq \mathbb{R}$. Define

$$\tau_A = \inf\{t : X_t \in A\}$$

Then τ_A is called the *hitting time* of the set A.

8.53 Remark. The question, for which sets A a hitting time τ_A is a stopping time, is completely solved. The solution is as follows.

We may assume that $P|\mathcal{F}_\infty$ is complete, i.e. that all subsets of negligible sets are added to $\mathcal{F}_\infty$. The whole theory developed so far is not affected by such a completion. We could assume from the beginning that our probability space is complete. The reason why we did not mention this issue is simple: We did not need completeness so far.

However, the most general solution of the hitting time problem needs completeness. The following is true:

If $P|\mathcal{F}_\infty$ is complete and if the filtration satisfies the usual conditions then the hitting time of every Borel set is a stopping time.

For further comments see Jacod-Shiryaev, [14], Chapter I, 1.27 ff.

For particular cases the stopping time property of hitting times is easy to prove.

8.54 Theorem. *Assume that $(X_t)_{t\geq 0}$ has right-continuous paths and is adapted to a filtration which satisfies the usual conditions.*
(a) Then τ_A is a stopping time for every open set A.
(b) If $(X_t)_{t\geq 0}$ has continuous paths then τ_A is a stopping time for every closed set A.

Proof: (a) Note that

$$\tau < t \;\Leftrightarrow\; X_s \in A \text{ for some } s < t$$

Since A is open and $(X_t)_{t\geq 0}$ has right-continuous paths it follows that

$$\tau < t \;\Leftrightarrow\; X_s \in A \text{ for some } s < t,\, s \in \mathbb{Q}$$

(b) Let A be closed and let (A_n) be open neighbourhoods of A such that $\bar{A}_n \downarrow A$. Define $\tau := \lim_{n\to\infty} \tau_{A_n} \leq \tau_A$ which exists since $\tau_{A_n} \uparrow$. We will show that $\tau = \tau_A$.

Since $\tau_{A_n} \leq \tau_A$ we have $\tau \leq \tau_A$. By continuity of paths we have $X_{\tau_{A_n}} \to X_\tau$ whenever $\tau < \infty$. Since $X_{\tau_{A_n}} \in \bar{A}_n$ it follows that $X_\tau \in A$ whenever $\tau < \infty$. This implies $\tau_A \leq \tau$. □

We need a notion of the past of a stopping time.

8.55 Problem. (*intermediate*) A stochastic interval is an interval whose boundaries are stopping times.
(a) Show that the indicators of stochastic intervals are adapted processes.
Hint: Consider $1_{(\tau,\infty)}$ and $1_{[\tau,\infty)}$.
(b) Let τ be a stopping time and let $F \subseteq \Omega$. Show that the process $1_F 1_{[0,\tau)}$ is adapted iff $F \cap (\tau \leq t) \in \mathcal{F}_t$ for all $t \geq 0$.
(c) Let $\mathcal{F}_\tau := \{F : F \cap (\tau \leq t) \in \mathcal{F}_t,\ t \geq 0\}$. Show that $\mathcal{F}_\tau$ is a σ-field.

8.56 Definition. Let τ be a stopping time. The σ-field $\mathcal{F}_\tau$ is called the *past* of τ.

The intuitive meaning of the past of a stopping time is as follows: An event F is in the past of τ if at every time t the occurrence of F can be decided provided that $\tau \leq t$.

Many of the subsequent assertions can be understood intuitively if this interpretation is kept in mind.

8.57 Problem. (*advanced*)
Let σ and τ be stopping times.
(a) If $\sigma \le \tau$ then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.
(b) $\mathcal{F}_{\sigma\cap\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.
(c) The sets $(\sigma < \tau)$, $(\sigma \le \tau)$ and $(\sigma = \tau)$ are in $\mathcal{F}_\sigma \cap \mathcal{F}_\tau$.
Hint: Start with proving $(\sigma < \tau) \in \mathcal{F}_\tau$ and $(\sigma \le \tau) \in \mathcal{F}_\tau$.
(d) Show that every stopping time σ is $\mathcal{F}_\sigma$-measurable.
(e) Let $\tau_n \downarrow \tau$. Show that $\mathcal{F}_\tau = \bigcap_{n=1}^\infty \mathcal{F}_{\tau_n}$.

There is a fundamental rule for iterated conditional expectations with respect to pasts of stopping times.

8.58 Theorem. *Let Z be an integrable or nonnegative random variable and let σ and τ be stopping times. Then*

$$E(E(Z|\mathcal{F}_\sigma)|\mathcal{F}_\tau) = E(Z|\mathcal{F}_{\sigma\cap\tau})$$

Proof: The proof is bit tedious and therefore many textbooks pose it as exercise problem (see Karatzas-Shreve, [15], Chapter 1, 2.17). Let us give more detailed hints.

We have to start with showing that

$$F \cap (\sigma < \tau) \in \mathcal{F}_{\sigma\cap\tau} \quad \text{and} \quad F \cap (\sigma \le \tau) \in \mathcal{F}_{\sigma\cap\tau} \text{ whenever } F \in \mathcal{F}_\sigma$$

Note that the nontrivial part is to show $\in \mathcal{F}_\tau$. The trick is to observe that on $(\sigma \le \tau)$ we have $(\tau \le t) = (\tau \le t) \cap (\sigma \le t)$.

The second step is based on the first step and consists in showing that

$$1_{(\sigma\le\tau)} E(Z|\mathcal{F}_\sigma) = 1_{(\sigma\le\tau)} E(Z|\mathcal{F}_{\sigma\cap\tau}) \tag{10}$$

Finally, we prove the assertion separately on $(\sigma \le \tau)$ and $(\sigma \ge \tau)$. For case 1 we apply (10) to the inner conditional expectation. For case 2 we apply (10) to the outer conditional expectation (interchanging the roles of σ and τ). □

We arrive at the most important result on stopping times and martingales. A preliminary technical problem is whether an adapted process stopped at σ is $\mathcal{F}_\sigma$-measurable. Intuitively, this should be true.

8.59 Discussion. Measurability of stopped processes
Let $(X_t)_{t\ge0}$ be an adapted process and σ a stopping time. We ask whether $X_\sigma 1_{(\sigma<\infty)}$ is $\mathcal{F}_\sigma$-measurable.

It is easy to prove the assertion for right-continuous processes with the help of 8.23. This would be sufficient for the optional stopping theorem below. However, for stochastic integration we want to be sure that the assertion is also valid for left-continuous processes. This can be shown in the following way.

Define
$$X_t^n := n \int_0^t X_s e^{n(s-t)} ds$$
Then $(X_t^n)_{t\geq 0}$ are continuous adapted processes such that $X_t^n \to X_t$ provided that $(X_t)_{t\geq 0}$ has left-continuous paths. Since the assertion is true for (X_t^n) it carries over to (X_t).

8.60 Theorem. *(Optional stopping theorem) Let $(M_t)_{t\geq 0}$ be a right continuous martingale. If σ is a bounded stopping time and τ is any stopping time then*
$$E(M_\sigma|\mathcal{F}_\tau) = M_{\sigma\cap\tau}$$

Proof: The proof is based on the following auxiliary assertion: Let τ be a bounded stopping time and let $M_t := E(Z|\mathcal{F}_t)$ for some integrable random variable Z. Then $M_\tau = E(Z|\mathcal{F}_\tau)$.

Let τ be a stopping time with finitely many values $t_1 < t_2 < \ldots < t_n$. Then
$$M_{t_n} - M_\tau = \sum_{k=1}^{n} (M_{t_k} - M_{t_{k-1}}) 1_{(\tau \leq t_{k-1})}$$
(Prove it on $(\tau = t_{j-1})$). It follows that $E((M_{t_n} - M_\tau)1_F) = 0$ for every $F \in \mathcal{F}_\tau$. This proves the auxiliary assertion for stopping times with finitely many values. The extension to arbitrary bounded stopping times is done by 8.23.

Let $T = \sup \sigma$. The assertion of the theorem follows from
$$E(M_\sigma|\mathcal{F}_\tau) = E(E(M_T|\mathcal{F}_\sigma)|\mathcal{F}_\tau) = E(M_T|\mathcal{F}_{\sigma\cap\tau}) = M_{\sigma\cap\tau}.$$
□

We finish this section with two consequences of the optional stopping theorem which are fundamental for stochastic integration.

8.61 Corollary. *Let τ be any stopping time. If $(M_t)_{t\geq 0}$ is a martingale then $(M_{\tau\cap t})_{t\geq 0}$ is a martingale, too.*

8.62 Problem. (*easy*)
Prove 8.61.

8.63 Corollary. *Let $(M_t)_{t\geq 0}$ be a martingale. Let $\sigma \leq \tau$ be stopping times and let Z be $\mathcal{F}_\sigma$-measurable and bounded. Then $Z(M_{\tau\cap t} - M_{\sigma\cap t})$ is a martingale, too.*

8.64 Problem. (*intermediate*)
Prove 8.63.
Hint: Apply 8.25.

UNDER CONSTRUCTION: EXTENSION TO SUB- AND SUPERMARTINGALES

8.7 The Markov property

We explain and discuss the Markov property at hand of the Wiener process.

When we calculate conditional expectations given the past $\mathcal{F}_s$ of a stochastic process $(X_t)_{t\ge 0}$ then from the general point of view conditional expectations $E(X_t|\mathcal{F}_s)$ are $\mathcal{F}_s$-measurable, i.e. they depend on any $X_u,\ u \le s$. But when we were dealing with special conditional expectations given the past of a Wiener process then we have got formulas of the type

$$E(W_t|\mathcal{F}_s) = W_s, \quad E(W_t^2|\mathcal{F}_s) = W_s^2 + (t-s), \quad E(e^{aW_t}|\mathcal{F}_s) = e^{aW_s + a^2(t-s)/2}$$

These conditional expectations do not use the whole information available in $\mathcal{F}_s$ but only the value W_s of the Wiener process at time s.

8.65 Theorem. *Let $(W_t)_{t\ge 0}$ be a Wiener process and $(\mathcal{F}_t)_{t\ge 0}$ its internal history. Then for every P-integrable function Z which is $\sigma(\bigcup_{u\ge s} \mathcal{F}_u)$-measurable we have*

$$E(Z|\mathcal{F}_s) = \phi(W_s)$$

where ϕ is some measurable function.

Proof: For the proof we only have to note that the system of functions

$$e^{a_1 W_{s+h_1} + a_2 W_{s+h_2} + \cdots + a_n W_{s+h_n}}, \quad h_i \ge 0, n \in \mathbb{N},$$

is total in $\mathcal{L}^2(\sigma(\bigcup_{u\ge s} \mathcal{F}_u))$. □

8.66 Problem. (*intermediate*)
Under the assumptions of 8.65 show that $E(Z|\mathcal{F}_s) = E(Z|W_s)$.

8.65 is the simplest and basic formulation of the Markov property. It is, however, illuminating to discuss more sophisticated versions of the Markov property. We need some preliminaries.

8.67 Remark. Redundant conditioning
We have to be aware of an important property of conditional expectations.
Let X be $\mathcal{A}$-measurable and let Y be P-integrable and independent of $\mathcal{A}$. Then we know that

$$E(XY|\mathcal{A}) = XE(Y|\mathcal{A}) = XE(Y)$$

The conditional expectation depends on $\mathcal{A}$ only through X. This can be understood intuitively in the following way: Since X is $\mathcal{A}$-measurable the information in $\mathcal{A}$ gives the whole information on X. The rest (i.e. Y) is independent of $\mathcal{A}$. Note, that the equation can be written as follows:

$$E(XY|\mathcal{A}) = \phi \circ X \text{ where } \phi(\xi) = E(\xi Y)$$

This view can be extended to much more general cases:

$$E(f(X,Y)|\mathcal{A}) = \phi \circ X \text{ where } \phi(\xi) = E(f(\xi, Y))$$

(provided that f is sufficiently integrable).

Let us calculate $E(f(W_{s+t})|\mathcal{F}_s)$ where f is bounded and measurable. We have

$$E(f(W_{s+t})|\mathcal{F}_s) = E(f(W_s + (W_{s+t} - W_s))|\mathcal{F}_s)$$

Since W_s is $\mathcal{F}_s$-measurable and $W_{s+t} - W_s$ is independent of $\mathcal{F}_s$ we have

$$E(f(W_{s+t})|\mathcal{F}_s) = \phi \circ W_s \text{ where } \phi(\xi) = E(f(\xi + (W_{s+t} - W_s))) \tag{11}$$

Roughly speaking, conditional expectations simply are expectations depending depending on a parameter slot where the present value of the process has to be plugged in.

8.68 Theorem. *Let $(W_t)_{t\ge 0}$ be a Wiener process and $(\mathcal{F}_t)_{t\ge 0}$ its internal history. Then the conditional distribution of $(W_{s+t})_{t\ge 0}$ given $\mathcal{F}_s$ is the same as the distribution of a process $\xi + \widetilde{W}_t$ where $\xi = W_s$ and $(\widetilde{W}_t)_{t\ge 0}$ is any (other) Wiener process.*

Proof: Extend (11) to functions of several variables. □

8.68 contains that formulation which is known as the ordinary Markov property of the Wiener process. It says that at every time point s the Wiener process starts afresh at the state $\xi = W_s$ as a new Wiener process forgetting everything what happened before time s.

It is a remarkable fact with far reaching consequences that the Markov property still holds if time s is replaced by a stopping time. The essential preliminary step is the following.

8.69 Theorem. *Let τ be any stopping time and define $Q(F) = P(F|\tau < \infty)$, $F \in \mathcal{F}_\infty$. Then the process*

$$X_t := W_{\tau+t} - W_\tau, \quad t \ge 0,$$

is a Wiener process under Q which is independent of $\mathcal{F}_\tau$.

Proof: (Outline) Let us show that

$$\int_F f(W_{\tau+t} - W_\tau)\, dP = P(F)E(f(W_t))$$

when $F \subseteq (\tau < \infty)$, $F \in \mathcal{F}_\tau$ and f is any bounded continuous function. But this is certainly true for stopping times with finitely many values. The common approximation argument proves the equation. Noting that the equation holds for $\tau + s$, $s > 0$, replacing τ, proves the assertion. □

8.70 Problem. (*advanced*)
Fill in the details of the proof of 8.69.

8.71 Theorem. *(Strong Markov property)*
Let $(W_t)_{t\ge 0}$ be a Wiener process and $(\mathcal{F}_t)_{t\ge 0}$ its internal history. Let σ be any stopping time. Then on $(\sigma < \infty)$ the conditional distribution of $(W_{\sigma+t})_{t\ge 0}$ given $\mathcal{F}_\sigma$ is the same as the distribution of a process $\xi + \widetilde{W}_t$ where $\xi = W_\sigma$ and $(\widetilde{W}_t)_{t\ge 0}$ is some (other) Wiener process.

Further reading: Karatzas-Shreve [15], sections 2.5 and 2.6.

Chapter 9

The financial market picture

This chapter gives an overview over the basic concepts of pricing in financial markets. We restrict our view to trading strategies with a finite number of trading times. However, all concepts are formulated such that they are valid also in the general case which will be tractable by stochastic analysis.

During the presentation of stochastic analysis we will refer to the concepts and ideas discussed in this chapter. Thus, the present chapter is the basic motivation for going into the troubles of stochastic analysis.

9.1 Assets and trading strategies

Let $S = (S^0, S^1, \ldots, S^m)$ be a finite set of right-continuous processes modelling the value of tradable assets of a financial market. We consider a finite time horizon $[0, T]$, $T < \infty$. A *trading strategy* in S is a process which determines how many units of each asset are held during a time interval:

$$H_t^k = \sum_{j=1}^n a_{j-1}^k 1_{(\sigma_{j-1}, \sigma_j]}, \quad k = 0, 1, 2, \ldots, n$$

where $0 = \sigma_0 \le \sigma_1 \le \ldots \le \sigma_n = T$ are stopping times and a_{j-1}^k is $\mathcal{F}_{\sigma_{j-1}}$-measurable. Thus, the processes H_t^k are adapted and left-continuous processes. Left-continuity is essential because trading strategies must be *predictable*.

The *market value* of the portfolio $(H_t^k)_k$ at time t is given by

$$V_t = \sum_k H_t^k S_t^k$$

The process (V_t) is the *wealth process* corresponding to the trading strategy.

The trading strategy is called *self-financing* if the changes in the portfolio at σ_j are financed by nothing else than the value of the portfolio:

$$\sum_k H_{\sigma_{j-1}}^k S_{\sigma_j}^k = \sum_k H_{\sigma_j}^k S_{\sigma_j}^k \tag{12}$$

The self-financing property has the consequence that the wealth process can be written as a gambling system like (6). Later we will see that for continuous trading the corresponding representation is that of an integral.

9.1 Theorem. *A trading strategy $(H^k)_k$ is self-financing iff*

$$V_t = V_0 + \sum_k \sum_j H^k_{\sigma_{j-1}} (S^k_{\sigma_j \cap t} - S^k_{\sigma_{j-1} \cap t}) \tag{13}$$

Proof: It is easy to see that 13 implies 12.
Conversely, if the trading strategy is self-financing then for $t \in (\sigma_{j-1}, \sigma_j]$ we have

$$\begin{aligned} V_t - V_{\sigma_{j-1}} &= \sum_k (H^k_{\sigma_{j-1}} S^k_t - H^k_{\sigma_{j-2}} S^k_{\sigma_{j-1}}) \qquad (14) \\ &= \sum_k H^k_{\sigma_{j-1}} (S^k_t - S^k_{\sigma_{j-1}}) \qquad (15) \end{aligned}$$

This formula can be extended to $t \in [0, T]$ by writing it as

$$V_{\sigma_j \cap t} - V_{\sigma_{j-1} \cap t} = \sum_k H^k_{\sigma_{j-1}} (S^k_{\sigma_j \cap t} - S^k_{\sigma_{j-1} \cap t})$$

The assertion follows from $V_t - V_0 = \sum_j (V_{\sigma_j \cap t} - V_{\sigma_{j-1} \cap t})$. □

9.2 Problem. Fill in the details of the proof of 9.1.

The value of assets is measured in terms of a unit of money. The unit of money can be a currency or some other positive value process. It is important to know how the properties of trading strategies behave under a change of the unit of money.

It should be noted that formula (16) only holds for wealth processes of self-financing trading strategies.

9.3 Theorem. *Let N be any positive right-continuous process. A trading strategy $(H^k)_k$ is self-financing for $S = (S^0, S^1, \ldots, S^m)$ iff it is self-financing for $\widetilde{S}$ where $\widetilde{S}^k = S^k/N$. If V and $\widetilde{V}$ are the corresponding wealth processes then*

$$\tilde{V}_t = \frac{V_t}{N_t}, \quad t \geq 0. \tag{16}$$

Proof: The first part follows easily by dividing (12) by N_{σ_j}.
For proving (16) we get from (14) that

$$\widetilde{V}_t - \widetilde{V}_{\sigma_{j-1}} = \frac{V_t}{N_t} - \frac{V_{\sigma_{j-1}}}{N_{\sigma_{j-1}}} \text{ whenever } \sigma_{j-1} < t \leq \sigma_j$$

With $t = \sigma_j$ an induction argument implies $\widetilde{V}_{\sigma_j} = V_{\sigma_j}/N_{\sigma_j}$. This proves the assertion.
□

9.2 Financial markets and arbitrage

A claim at time $t = T$ is any $\mathcal{F}_T$-measurable random variable C. The fundamental problem of mathematical finance is to find a reasonable price x_0 at time $t = 0$ for the claim C.

There are two methods to find a price x_0 for the claim. The *insurance method* is to define x_0 as the expectation under P of the discounted claim. The risk of this kind of pricing is controlled by selling a large number of claims. Then by the LLN the average cost of that set of claims equals x_0. But this works only if the claims are independent. That might be true for insurance but not for financial markets.

The more recent and most important method of pricing is *risk neutral pricing* using hedge strategies. This leads to the concept of a market model.

9.4 Definition. A *financial market* is a set $\mathcal{M}$ of wealth processes (for the moment right-continuous processes) with the following properties:

(1) $\mathcal{M}$ is a vector space, i.e. every linear combination of wealth process in $\mathcal{M}$ is contained in $\mathcal{M}$.

(2) Every self-financing trading strategy based on finitely many wealth processes in $\mathcal{M}$ leads to a wealth process in $\mathcal{M}$.

9.5 Problem. Show that (2) implies (1) in 9.4.

A claim C has a *hedge* in the market $\mathcal{M}$ (is *attainable*) if there is a wealth process $V \in \mathcal{M}$ satisfying $V_T = C$. In this case it looks reasonable to define $V_0 = x_0$. This is called *risk neutral pricing*. However, the risk neutral pricing method is only reasonable if the price does not allow *arbitrage*.

The idea is the following: A price x_0 for a claim C is arbitrage-free if there does not exist a wealth process $V \in \mathcal{M}$ such that $V_0 = x_0$ and $V_T \geq C$, $V_T \neq C$. Alas, such a concept does not work since there are very plausible market models where arbitrage can be achieved with highly risky wealth processes. Such wealth processes have to be excluded from competition.

9.6 Definition. A wealth process $V \in \mathcal{M}$ is called *admissible* if it is bounded from below.

9.7 Definition. A price x_0 is an *arbitrage-free price* for a claim C if there does not exist an admissible wealth process $V \in \mathcal{M}$ such that $V_0 = x_0$ and $V_T \geq C$, $V_T \neq C$.

How can we be sure that risk neutral pricing leads to arbitrage-free prices ?

9.3 Martingale measures

Let $\mathcal{M}$ be a market model. The common answer to the question posed is the existence of a so-called martingale measure.

A generating system of $\mathcal{M}$ is a subset $\mathcal{M}_0 \subseteq \mathcal{M}$ such that every wealth process $V \in \mathcal{M}$ is generated by a trading strategy based on finitely many elements of $\mathcal{M}_0$.

9.8 Definition. A *martingale measure* is a probability measure $Q \sim P$ such that all wealth processes of some generating system are martingales.

9.9 Lemma. *If there exists a martingale measure then all admissible wealth processes $V \in \mathcal{M}$ satisfy*

$$E(V_t|\mathcal{F}_s) \le V_s \text{ whenever } s < t.$$

(Admissible wealth processes are "supermartingales").

The proof of this lemma is postponed.

The following theorem is fundamental for the modern theory of pricing in financial markets.

9.10 Theorem. *Let $\mathcal{M}$ be a market model and Q some martingale measure. Let C be an attainable claim with a hedge whose wealth process is a Q-martingale (a martingale hedge). Then $x_0 := E_Q(C)$ is an arbitrage-free price.*

Proof: Let C be a claim and let V be the wealth process of a martingale hedge of the claim. Clearly, we have $x_0 = V_0$.

Let $V^1 \in \mathcal{M}$ be an admissible wealth process such that $V_0^1 = V_0 = x_0$ and $V_T^1 \ge C = V_T$. Then we have

$$E_Q(V_T) = V_0 \quad \text{and} \quad E_Q(V_T^1) \le V_0^1 = V_0$$

Since $V_T^1 - V_T \ge 0$ and $E_Q(V_T^1 - V_T) \le 0$ it follows that $V_T^1 = V_T = C$ Q-a.e. und hence P-a.e. □

Theorem 9.10 shows that the existence of a martingale measure makes risk neutral pricing an easy exercise (at least in theory). Therefore it is important to ask whether we may expect to have martingale measures for arbitrage-free markets. For common financial market models martingale measures exist, as a rule. However, for a mathematician it is a challenge to ask whether the existence of a martingale measure is only some sufficient condition for a market to be arbitrage-free, or it is even necessary. This turns out to be a rather delicate question.

9.11 Definition. A market model $\mathcal{M}$ is *arbitrage-free* if $x_0 = 0$ is an arbitrage-free price of the claim $C = 0$.

9.12 Problem. Show that the existence of a martingale measure implies that the market is arbitrage-free.

It is not true that every arbitrage-free market admits martingale measures.

9.4 Change of numeraire

There is an easy aspect of the existence of martingale measures which is important for practical purposes.

Assume that one of the assets of the market is a *risk-less* (for the moment: non-stochastic) positive asset, e.g. a bank account $N_t = e^{rt}$.

9.13 Problem. In an arbitrage-free market all risk-less wealth processes are proportional.

9.14 Problem. Show: If a martingale measure exists then all risk-less assets of the market are constant.

If a market contains a bank account with positive interest then there cannot exist a martingale measure ! This message sounds a bit disappointing. Fortunately, there is an easy solution of that problem.

Assume that there exists some positive tradable asset N (a *numeraire*). Then we may define this asset as our unit of money. For the market model this amounts to dividing all value processes by N, resulting in a so-called normalized market $\widetilde{\mathcal{M}}$ consisting of the wealth processes given in 9.3. The normalized market has only constant risk-less assets proportional to $1 = N/N$ and therefore does not exclude the existence of a martingale measure. If we try to find martingale measures, first we have to look for numeraires in order to normalize the market.

Summing up, the risk neutral pricing machinery runs as follows:

(1) Find a numeraire N and turn to the normalized market.

(2) Find a martingale measure Q for the normalized market.

(3) If C/N_T has a martingale hedge in the normalized market then define the price to be $x_0 = N_0 E_Q(C/N_T)$.

Chapter 10

Stochastic calculus

10.1 Elementary Integration

Bounded variation

Let $f : [0,T] \to \mathbb{R}$ be any function.

10.1 Definition. The *variation* of f on the interval $[s,t] \subseteq [0,T]$ is

$$V_s^t(f) := \sup \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})|$$

where the supremum is taken over all subdivisions $s = t_0 < t_1 < \ldots < t_n = t$ and all $n \in \mathbb{N}$.

A function f is of *bounded variation* on $[0,T]$ if $V_0^T(f) < \infty$. The set of all functions of bounded variation is denoted by $BV([0,T])$.

10.2 Problem. (*intermediate*)
Let f be differentiable on $[s,t]$ with continuous derivative. Then $f \in BV$ and

$$V_s^t(f) = \int_s^t |f'(u)|du$$

10.3 Problem. (*easy*)
Show that BV is a vector space.

10.4 Problem. (*very easy*)
Show that monotone functions are BV and calculate their variation.

10.5 Problem. (*intermediate*)
Show that any function $f \in BV$ can be written as $f = g - h$ where g, h are increasing and satisfy $V_0^t(f) = g(t) + h(t)$.
Hint: Let $g(t) := (V_0^t(f) + f(t))/2$ and $h(t) := (V_0^t(f) - f(t))/2$.

10.6 Problem. Which BV-functions are Borel-measurable ?

There are continuous functions on compact intervals which are not of bounded variation.

The Cauchy-Stieltjes integral

Let $\mathcal{T}([0,T])$ be the set of all left-continuous stepfunctions on $[0,T]$, i.e. functions of the form

$$f(t) = \sum_{k=1}^{n} a_k 1_{(t_{k-1},t_k]}$$

where $0 = t_0 < t_1 < \ldots < t_n = T$ is some subdivision. If $f \in \mathcal{T}$ and if g is right-continuous then we define

$$\int_0^T f\,dg := \sum_{k=1}^{n} a_k (g(t_k) - g(t_{k-1})) \tag{17}$$

If g is increasing then this definition coincides with $\int f\,d\lambda_g$. In a similar way as for the abstract integral we conclude that 17 is a valid definition being linear both in f and g.

We want to extend the integral to a more general class of functions f. Of course, this could be done along the lines of general integration theory. But we will describe the older and more elementary approach by Cauchy for its formal similarity to Protter's ([19]) definition of the stochastic integral.

Recall that a sequence of subdivisions $0 = t_0 < t_1 < \ldots < t_n = T$ is called Riemannian if $\max |t_k - t_{k-1}| \to 0$.

10.7 Lemma. *Let f be left-continuous with limits from the right (caglad). Then for any Riemannian sequence of subdivisions the sequence of step functions*

$$f_n := \sum_{k=1}^{n} f(t_{k-1}) 1_{(t_{k-1},t_k]}$$

converges uniformly to f, i.e. $||f_n - f||_u \to 0$.

Proof: This is a beginner's lemma if f is continuous on $[0,T]$. If f has infinitely many jumps one has to work a little harder. □

Such step functions on Riemannian sequences of subdivisions can be used to extend the integral to arbitrary caglad-functions due to the following inequality.

10.8 Lemma. *If $f \in \mathcal{T}$ and if g is right-continuous then*

$$\Big| \int_0^T f\,dg \Big| \le ||f||_u V_0^T(g)$$

This lemma implies that the integral is continuous under uniform convergence of the integrands provided that g is of bounded variation. In particular, if (f_n) converges uniformly then the sequence of integrals is also convergent. This leads to the definition of the integral.

10.9 Definition. Let f be caglad and g be cadlag and of bounded variation. Then the *Cauchy-Stieltjes integral* is

$$\int_0^T f\,dg := \lim_{n\to\infty} \int_0^T f_n\,dg$$

where (f_n) is any sequence in $\mathcal{T}$ converging uniformly to f.

10.10 Problem. Show that for increasing g the CS-integral coincides with the abstract integral for λ_g.

For $g(t) = t$ this is the notion of an integral that is taught in schools.

10.11 Remark. Later the notion of the stochastic integral will be defined in the following way (due to Protter [19]). The integrators g are extended to adapted cadlag processes for which the integrals of adapted caglad step-processes have a continuity property similar to the CS-integral. It turns out that not only processes with BV-paths have such a continuity property but also martingales. For such processes (so-called semimartingales) the definition of the integral works for adapted caglad processes. In this way all processes can be used as integrators which can be written as a sum of a cadlag martingale and an adapted cadlag BV-process.

Differential calculus

Let f be caglad on $[0, T]$ and $g \in BV([0, T])$ be cadlag. For notational convenience we define

$$\int_0^t f\,dg := \int_0^T 1_{(0,t]} f\,dg, \quad 0 \le t \le T,$$

and

$$f \bullet g : t \mapsto \int_0^t f\,dg, \quad 0 \le t \le T.$$

10.12 Theorem. *Let f be caglad on $[0, T]$ and $g \in BV([0, T])$ be cadlag. Then the following assertions are true:*

(a) $f \bullet g$ is cadlag.
(b) $f \bullet g$ is of bounded variation since $V(f \bullet g) = |f| \bullet V(g)$.
(c) If g is continuous then $f \bullet g$ is continuous.

Proof: (a) and (c) are due to the fact that cadlag and continuity is inherited under uniform convergence. The equation under (b) is a consequence of the corresponding equation for step-functions. □

Now we turn to the three basic rules of differential calculus. These are the prototypes for the corresponding rules of stochastic calculus. The rules are concerned with the evaluation of

$$\int_0^T f\, d(g \bullet h), \quad \int_0^T f\, d(gh), \quad \int_0^T f\, d(g \circ h)$$

We assume tacitely that all involved functions fulfil those conditions which are required for making the expressions well-defined.

The first rule is *associativity*. Let f and g be caglad and h cadlag and BV. Then

$$\int_0^T f\, d(g \bullet h) = \int_0^T fg\, dh, \quad \text{in short:} \quad d(g \bullet h) = g\, dh$$

This is true by definition for $f = 1_{(0,t]}$ and extends to general f by a straightforward induction argument.

There is an important special case. Let $h(t) = t$ and let G be the primitive of g, i.e. $G' = g$. Since $G = g \bullet h$ we obtain

$$\int_0^T f\, dG = \int_0^T f(s)G'(s)\, ds, \quad \text{in short:} \quad dG(s) = G'(s)\, ds$$

The second rule is the *product rule* which in integral notation is called *integration by parts*. For this let g and h be continuous and BV. Then

$$\int_0^T f\, d(gh) = \int_0^T fg\, dh + \int_0^T fh\, dg, \quad \text{in short:} \quad d(gh) = g\, dh + h\, dg$$

For $f = 1_{(0,t]}$ this means

$$g(t)h(t) = g(0)h(0) + \int_0^t g\, dh + \int_0^t h\, dg$$

This gives well-known formulas if g and h are differentiable.

The proof runs over approxmation by step-functions. Let $0 = t_0 < t_1 < \ldots < t_n = t$ by some subdivision. Define $\Delta g(t_j) := g(t_j) - g(t_{j-1})$ and $\Delta h(t_j)$ similarly. Then

$$g(t)h(t) = g(0)h(0) + \sum_{j=1}^n g(t_{j-1})\Delta h(t_j) + \sum_{j=1}^n h(t_{j-1})\Delta g(t_j) + \sum_{j=1}^n \Delta g(t_j)\Delta h(t_j)$$

The assertion follows since by the BV-property the last term tends to zero for a Riemannian sequence of subdivisions.

At this point we can already see some of the frictions with extending such formulas to the stochastic case. The notorious last term vanishes if at least one of the functions g or h is BV. But if both functions are Wiener paths then the last term tends to the quadratic variation !

The last rule is the *chain rule* or *substitution rule*. Let g be continuously differentiable. Then

$$\int_0^T f\, d(g \circ h) = \int_0^T f\,(g' \circ h)\, dh, \quad \text{in short:} \quad d(g \circ h) = (g' \circ h)\, dh$$

The special case $dg(t) = g'(t)dt$ is the chain rule of ordinary calculus.

There is an elegant proof verifying the formula for arbitrary power functions $g(t) = t^k$ by the product rule, passing to polynomials and finally applying the Weierstrass approximation theorem. A different approach is based on Taylor polynomials where the terms of higher order vanish by the BV-property of h. Both proofs indicate that in the stochastic case the formula will have to be changed by including quadratic variation terms. The result will be the Ito-formula.

10.2 The stochastic integral

Let $(Z_t)_{t\geq 0}$ be any right-continuous adapted process. It is our goal to define an integral

$$\int_0^T H\, dZ, \quad t \geq 0,$$

for left continuous adapted processes $(H_t)_{t\geq 0}$. Let $\mathcal{L}_0$ be the set of all left-continuous adapted processes. There are subsets of $\mathcal{L}_0$ where the definition of the integral is easy.

The integral of stepfunctions

Let $\mathcal{E}_0$ be the set of processes of the form

$$H_t(\omega) = \sum_{j=0}^{n} a_{j-1}(\omega) 1_{(s_{j-1}, s_j]}(t)$$

where $0 = s_0 < s_1 < \ldots < s_n = T$ is a subdivision and a_j is $\mathcal{F}_{s_j}$-measurable for every j. It is easy to see that $(H_t)_{t\geq 0}$ is left-continuous and adapted.

A bit more general is the set $\mathcal{E}$ of processes

$$H_t(\omega) = \sum_{j=0}^{n} a_{j-1}(\omega) 1_{(\sigma_{j-1}, \sigma_j]}(t) \tag{18}$$

where $0 = \sigma_0 < \sigma_1 < \ldots < \sigma_n = T$ is a subdivision and a_j is $\mathcal{F}_{\sigma_j}$-measurable for every j. Again it is obvious that the paths are left-continuous and from 8.55(b) we know that the processes in $\mathcal{E}$ are adapted.

10.13 Problem. Let $H \in \mathcal{E}$ be defined by (18). Show that $1_{(0,t]}H = \sum_{j=1}^n a_{j-1}1_{(\sigma_{j-1}\cap t,\sigma_j\cap t]}$.

For functions in $\mathcal{E}$ it obvious how to define the integral. This can be done pathwise and leads to the following definition:

$$\int_0^t H\,dZ := \int_0^T 1_{(0,t]}H\,dZ = \sum_{j=1}^n a_{j-1}(Z_{\sigma_j\cap t} - Z_{\sigma_{j-1}\cap t})$$

if H is defined by (18). Since for each single path this is an ordinary Stieltjes integral we have immediately the properties:

$$\int_0^t (\alpha H_1 + \beta H_2)\,dZ = \alpha \int_0^t H_1\,dZ + \beta \int_0^t H_2\,dZ \tag{19}$$

$$\int_0^t H d(\alpha Z_1 + \beta Z_2) = \alpha \int_0^t H\,dZ_1 + \beta \int_0^t H\,dZ_2 \tag{20}$$

For notational convenience denote $H \bullet Z : t \mapsto \int_0^t H\,dZ$.

10.14 Theorem. *Let $(M_t)_{t\geq 0}$ be a martingale and let $H \in \mathcal{E}$ be bounded. Then $H \bullet M$ is a martingale.*

Proof: Apply 8.63. □

10.15 Discussion. Financial markets

Let us continue the financial market framework of chapter 9.

First we observe that the representation of self-financing trading strategies in 9.1 can be written as an integral:

$$V_t = V_0 + \sum_k \sum_j H^k_{\sigma_{j-1}}(S^k_{\sigma_j\cap t} - S^k_{\sigma_{j-1}\cap t}) = V_0 + \sum_k \int_0^t H^k_s\,dS^k_s$$

Thus, the self-financing property is characterized by the equation

$$V_t = \sum_k H^k_t S^k_t = \sum_k H^k_0 S^k_0 + \sum_k \int_0^t H^k_s\,dS^k_s \tag{21}$$

If there is a martingale measure Q and if the trading strategy is bounded then the corresponding wealth process is a martingale under Q.

It follows that every claim C which can be hedged by a bounded self-financing trading strategy in $\mathcal{E}$ is a martingale hedge and the pricing formula $x_0 = E_Q(C)$ can be applied.

However, many claims cannot be (exactly) hedged using only finitely many trading times. Therefore for dealing with general claims we have to consider continuous trading strategies. The self-financing property of continuous trading strategies will be defined by a formula like (21) and for this we have to extend our notion of the integral to continuous integrands. If the integrators are processes with paths of bounded variation then the integral extension could be done pathwise like a CS-integral. But if our assets behave like random walks, e.g. driven by a Wiener process, there is no hope to have paths of bounded variation.

Semimartingales

10.16 Definition. A right continuous process $(X_t)_{t\ge 0}$ is a *semimartingale* if for every sequence (H^n) of processes in $\mathcal{E}$ the following condition holds:

$$\sup_{s\le T} |H_s^n| \to 0 \;\Rightarrow\; \sup_{t\le T} \left| \int_0^t H_s^n \, dZ_s \right| \xrightarrow{P} 0$$

The set of all semimartingales is denoted by $\mathcal{S}$.

It will turn out that a reasonable extension process of the stochastic integral can be carried out for integrator processes which are semimartingales. It is therefore important to get an overview over typical processes that are semimartingales. Before we turn to such examples let us study the structure of the set $\mathcal{S}$ of semimartingales.

10.17 Problem. (*intermediate for mathematicians*)

Show that:

(a) The set of semimartingales is a vector space.

(b) If $X \in \mathcal{S}$ then for every stopping time the *stopped process* $X^\tau := (X_{\tau\cap t})_{t\ge 0}$ is a semimartingale.

(c) Let $\tau_n \uparrow \infty$ be a sequence of stopping times such that $X^{\tau_n} \in \mathcal{S}$ for every $n \in \mathbb{N}$. Then $X \in \mathcal{S}$.

Hint: Note that $(X_t \neq X_t^{\tau_n}) \subseteq (\tau_n < t)$.

The concept of semimartingales is only reasonable if it covers adapted cadlag processes with paths of bounded variation. From 10.8 it follows that this is actually the case.

The following result opens the door to stochastic processes like the Wiener process.

10.18 Theorem. *Every square integrable cadlag martingale $(M_t)_{t\geq 0}$ is a semimartingale.*

Proof: Let (H^n) be a sequence in $\mathcal{E}$ such that $||H^n||_u \to 0$. Since $\int_0^t H^n dM$ is a martingale we have by the maximal inequality

$$P\Big(\sup_{s\leq t}\Big|\int_0^s H^n dM\Big| > a\Big) \leq \frac{1}{a^2}E\Big(\Big(\int_0^t H^n dM\Big)^2\Big)$$

For convenience let $M_j := M_{\sigma_j^n \cap t}$. We have

$$\begin{aligned}
E\Big(\Big(\int_0^t H^n dM\Big)^2\Big) &= E\Big(\Big(\sum_{j=1}^n a_{j-1}(M_j - M_{j-1})\Big)^2\Big) \\
&= E\Big(\sum_{j=1}^n a_{j-1}^2(M_j - M_{j-1})^2\Big) \\
&\leq ||H^n||_u^2 E\Big(\sum_{j=1}^n (M_j - M_{j-1})^2\Big) \\
&= ||H^n||_u^2 E\Big(\sum_{j=1}^n (M_j^2 - M_{j-1}^2)\Big) \leq ||H^n||_u^2 E(M_t^2)
\end{aligned}$$

□

Thus, we proved that $(W_t)_{t\geq 0}$is a semimartingale.

10.19 Problem. (*easy*)
Show that $(W_t^2)_{t\geq 0}$ is a semimartingale.

10.20 Problem. (*intermediate*)
Show that every cadlag martingale $(M_t)_{t\geq 0}$ with continuous paths is a semimartingale.
Hint: Let $\tau_n = \inf\{t : |M_t| \geq n\}$ and show that M^{τ_n} is a square integrable martingale for very $n \in \mathbb{N}$.

Summing up, we have shown that every cadlag process which is a sum of a continuous martingale and an adapted process with paths of bounded variation is a semimartingale.

Actually every cadlag martingale is a semimartingale. See Jacod-Shiryaev, [14], Chapter I, 4.17.

Extending the stochastic integral

The extension of the stochastic integral from $\mathcal{E}$ to $\mathcal{L}_0$ is based on the fact that every process in $\mathcal{L}_0$ can be approximated by processes in $\mathcal{E}$. In short, the procedure is as follows. Let X is a semimartingale and let $H \in \mathcal{L}_0$. Consider some sequence (H^n) in $\mathcal{E}$ such that $H^n \to H$ and define

$$\int_0^T H\,dX := \lim_{n\to\infty} \int_0^T H^n\,dX \tag{22}$$

However, in order to make sure that such a definition makes sense one has to consider several mathematical issues.

10.21 Discussion. Foundations of the extension process
The main points of definition (22) are existence and uniqueness of the limit. Let $X \in \mathcal{S}$ and $H \in \mathcal{L}_0$. We follow Protter, [20].

(1) One can always find a sequence $(H^n) \subseteq \mathcal{E}$ such that

$$\sup_{s\le T} |H^n_s - H_s| \xrightarrow{P} 0$$

(2) Semimartingales satisfy

$$(H^n) \subseteq \mathcal{E},\ \sup_{s\le T} |H^n_s| \xrightarrow{P} 0 \ \Rightarrow\ \sup_{t\le T} \left| \int_0^t H^n dX \right| \xrightarrow{P} 0.$$

(This is slightly stronger than the defining property of semimartingales.)

(3) From (2) it follows that for every sequence $(H^n) \subseteq \mathcal{E}$ satisfying (1) the corresponding sequence of stochastic integrals

$$\int_0^T H^n dX$$

is a Cauchy sequence with respect to convergence in probability, uniformly on $[0, T]$. Therefore there exists a process Y such that

$$\sup_{t\le T} \left| \int_0^t H^n dX - Y_t \right| \xrightarrow{P} 0.$$

(4) From (2) it follows that the limiting process Y does not depend on the sequence (H^n).

The preceding discussion shows that there is a well-defined stochastic integral $\int_0^T H\,dX$ whenever $H \in \mathcal{L}_0$ and $X \in \mathcal{S}$. The stochastic integral has a strong continuity property.

10.22 Theorem. *Let X be a semimartingale. Then for every sequence (H^n) of processes in $\mathcal{L}_0$*

$$\sup_{s\le T} |H^n_s| \xrightarrow{P} 0 \ \Rightarrow\ \sup_{t\le T} \left| \int_0^t H^n_s\,dZ_s \right| \xrightarrow{P} 0$$

For deriving (understanding) the basic properties or rules of this stochastic integral we will apply the following approximation result.

10.23 Theorem. *Let X be a semimartingale and $H \in \mathcal{L}_0$. Assume that $0 = t_0^n < t_1^n < \ldots < t_{k_n}^n = t$ is any Riemannian sequence of subdivisions of $[0,t]$. Then*

$$\sup_{s \le t} \Big| \sum_{j=1}^{k_n} H_{t_{j-1}}(X_{t_j} - X_{t_{j-1}}) - \int_0^t H\,dX \Big| \xrightarrow{P} 0$$

Proof: This is not a proof but a comment. The assertion can be proved by our means if the paths of H are continuous. In this case the Riemannian step processes

$$\sum_{j=1}^{k_n} H_{t_{j-1}} 1_{(t_{j-1},t_j]}$$

converge to H in probability, uniformly on compacts, (actually they converge everywhere, uniformly on compacts). If H is only left-continuous then the Riemannian step functions converge to H pointwise, but not necessarily uniformly on compacts. In order to achieve uniform convergence one has to replace the arbitrary deterministic sequence of subdivisions by a particular sequence of subdivisions based on stopping times instead of fixed interval boundaries.

However, the result is true anyway. For BV-processes X it is a consequence of Lebesgue's theorem on dominated convergence. It can also be proved for square integrable martingales X. The universal validity follows from a very deep representation theorem for semimartingales which is not available for us at this stage. Confer Protter, [20], or Jacod-Shiryaev, [14]. □

Let us apply 10.23 for the evaluation of a fundamental special case.

10.24 Theorem. *Let $(W_t)_{t\ge 0}$ be a Wiener process. Then*

$$\int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - t) \tag{23}$$

Proof: Let $0 = t_0 \le t_1 \le \ldots \le t_n = t$ be an interval partition such that $\max |t_j - t_{j-1}| \to 0$ as $n \to \infty$. This implies

$$\sum_{j=1}^{n} W_{t_{j-1}}(W_{t_j} - W_{t_{j-1}}) \xrightarrow{P} \int_0^t W_s\, dW_s$$

On the other hand we have

$$W_t^2 = \sum_{j=1}^n (W_{t_j}^2 - W_{t_{j-1}}^2) =$$
$$= \sum_{j=1}^n (W_{t_j} - W_{t_{j-1}})^2 + 2\sum_{j=1}^n W_{t_{j-1}}(W_{t_j} - W_{t_{j-1}})$$

We know that

$$\sum_{j=1}^n (W_{t_j} - W_{t_{j-1}})^2 \to t \quad (P)$$

This proves the assertion. □

10.25 Problem. How to modify (23) if $(W_t)_{t\ge 0}$ is replaced by some BV-process ?

It is clear that the linearity properties (19) remain valid for the stochastic integral with $H \in \mathcal{L}_0$.

10.26 Problem. Define

$$\int_s^t H\,dX := \int_0^t 1_{(s,\infty)} H\,dW$$

(1) Prove a concatenation property of the stochastic integral.
(2) Show that

$$\int_s^t 1_F H\,dX = 1_F \int_s^t H\,dX \text{ whenever } F \in \mathcal{F}_s.$$

Path properties

UNDER CONSTRUCTION

The Wiener integral

UNDER CONSTRUCTION

10.3 Calculus for the stochastic integral

There are three fundamental rules for calculations with the stochastic integral which correspond to the three rules considered in section 10.1:

(1) the associativity rule,
(2) the integration-by-parts formula,
(3) the chain rule (Ito's formula)

The associativity rule

This rule can be formulated briefly as follows. Let $H, G \in \mathcal{L}_0$ and $X \in \mathcal{S}$. Then

$$H \bullet (G \bullet X) = (HG) \bullet X, \quad \text{in short:} \quad d(G \bullet X) = G\,dX \tag{24}$$

Details are as follows.

10.27 Theorem.
(1) Let $X \in \mathcal{S}$ and $G \in \mathcal{L}_0$. Then $G \bullet X$ is in $\mathcal{S}$.
(2) Let $H \in \mathcal{L}_0$. Then $\int_0^T H\,d(G \bullet X) = \int_0^T HG\,dX$.
Proof: For $H_n \in \mathcal{E}_0$ we have

$$\int_0^T H_n\,d(G \bullet X) = \int_0^T H_n G\,dX$$

If $H_n \to 0$ in an appropriate sense this implies the semimartingale property of $G \bullet X$. If $H_n \to H$ in an appropriate sense the asserted equation follows. □

There is an important consequence of rule (24) which should be isolated.

10.28 Theorem. *Truncation rule*
Let $H \in \mathcal{L}_0$ and $X \in \mathcal{S}$. Then for any stopping time τ

$$\int_0^T 1_{(0,\tau]} H\,dX = \int_0^{T\cap\tau} H\,dX = \int_0^T H\,dX^\tau$$

10.29 Problem. (*intermediate*)
Prove 10.28.
Hint: The first equation follows from the definition of the integral on $\mathcal{E}$. For the second equation note that $1_{(0,\tau]} \bullet X = X^\tau$.

The integration-by-parts formula

We restrict our presentation of the integration-by-parts formula to processes with continuous paths.

Recall the deterministic integration-by-parts formula for continuous BV-functions:

$$f(t)g(t) - f(0)g(0) = \int_0^t f\,dg + \int_0^t g\,df$$

This formula is not true for arbitrary semimartingales. The following is a definition rather than a theorem but is called the *integration by parts* formula.

10.30 Definition. Let X and Y be semimartingales with continuous paths. Define

$$[X,Y]_t := X_t Y_t - X_0 Y_0 - \int_0^t X\,dY - \int_0^t Y\,dX, \quad t \geq 0.$$

This process is called the *quadratic covariation* of X and Y.

It is clear that $[X,Y]$ is well-defined and is a continuous adapted process. The integration by parts formula can be written as

$$\int_0^T H\,d(XY) = \int_0^T HX\,dY + \int_0^T HY\,dX + \int_0^T H\,d[X,Y], \quad H \in \mathcal{L}_0$$

or in short

$$d(XY) = X dY + Y dX + d[X,Y]$$

However, this only makes sense if $[X,Y]$ is a semimartingale. So let us have a closer look onto $[X,Y]$.

10.31 Problem. (*easy*)
Show that $[X,Y]$ is linear in both arguments.

10.32 Theorem. *Let X and Y be semimartingales with continuous paths. For every Riemannian sequence of subdivisions*

$$\sum_{j=1}^n (X_{t_j} - X_{t_{j-1}})(Y_{t_j} - Y_{t_{j-1}}) \xrightarrow{P} [X,Y]_t, \quad t \geq 0.$$

Proof: This follows easily from the definition of $[X,Y]$ when it is approximated by Riemannian sequence of step processes. □

10.33 Problem. (*intermediate*)
Fill in the details of the proof of 10.32.

From 10.32 it follows that $[X,X] =: [X]$ is the quadratic variation of X. This is an increasing process, hence a BV-process and a semimartingale. Moreover, since

$$[X,Y] = \frac{1}{2}([X+Y] - [X-Y])$$

also the quadratic covariation is a BV-process and a semimartingale.

10.34 Problem. (*easy*)
Show: If X and Y are continuous semimartingales then XY is a (continuous) semimartingale, too.

10.35 Problem. (*intermediate*)
Let X be a continuous BV-process and Y, Z continuous semimartingales. Show that:
(a) $[X] = 0$,
(b) $[X, Y] = 0$,
(c) $[X + Y, Z] = [Y, Z]$.

10.36 Problem. (*intermediate*)
Let $X \in \mathcal{S}$ be continuous.
(a) Show that

$$dX^2 = 2XdX + d[X]$$

(b) Find a formula for dX^k, $k \in \mathbb{N}$.
Hint: Use induction on k.

10.37 Problem. (*advanced*)
Show that $[X^\tau, Y] = [X, Y]^\tau$.
Hint: This is intuitively clear from 10.32 and could be made precise by approximating τ by a sequence of stopping times with finitely many values. However, it can be obtained from the definition without any approximation argument. For this note that

$$\int_0^t X^\tau dY = \int_0^{\tau \cap t} X\, dY + X_{\tau \cap t}(Y_t - Y_{\tau \cap t})$$

10.38 Problem. (*intermediate*)
Show that $[H \bullet X, Y] = H \bullet [X, Y]$.
Hint: Prove it for $H = 1_{(\sigma,\tau]}$ where $\sigma \leq \tau$ are stopping times.

10.39 Problem. (*easy*)
Let (W_t) be a Wiener process. Calculate $[H \bullet W]$.

10.40 Problem. (*intermediate*)
Show that $H \bullet W$ is a BV-process iff $H \equiv 0$.

10.41 Problem. (*intermediate*)
A process of the form

$$X_t = x_0 + \int_0^t a_s ds + \int_0^t b_s dW_s$$

is called an *Ito-process*.
Show that a and b are uniquely determined by X.

Ito's formula

Now we turn to the most important and most powerful rule of stochastic analysis.

10.42 Theorem. *Ito's formula*
Let $X \in \mathcal{S}$ be continuous and let $\phi : \mathbb{R} \to \mathbb{R}$ be twice differentiable with continuous derivatives. Then

$$\phi(X_t) = \phi(X_0) + \int_0^t \phi'(X_s)\, dX_s + \frac{1}{2}\int_0^t \phi''(X_s)\, d[X]_s$$

Proof: The assertion is true for polynomials. Since smooth functions can be approximated uniformly by polynomials in such a way that also the corresponding derivatives are approximated the assertion follows. □

10.43 Problem. (*easy*)
Show that Ito's formula is true if ϕ is a polynomial.
Hint: Start with powers $\phi(x) = x^k$.

10.44 Problem. (*very easy*)
State 10.42 in terms of differentials.

10.45 Problem. (*easy*)
Calculate dW_t^a, $a > 0$.

10.46 Problem. (*advanced*)
Use Ito's formula to find a recursion formula for $E(W_t^k)$, $k \in \mathbb{N}$.

10.47 Problem. (*easy*)
Calculate $de^{\alpha W_t}$.

10.48 Definition. Let $X \in \mathcal{S}$ be continuous. Then

$$\mathcal{E}(X) = e^{X-[X]/2}$$

is called the *stochastic exponential* of X.

10.49 Problem. (*intermediate*)
Let $X \in \mathcal{S}$ be continuous and $Y := \mathcal{E}(X)$. Show that

$$Y_t = Y_0 + \int_0^t Y_s dX_s, \quad \text{in short:} \quad dY = Y dX$$

There is a subtle point to discuss. Consider some positive continuous semimartingale X and a function like $\phi(x) = \log(x)$ or $\phi(x) = 1/x$. Then we may consider $\phi(X)$

since it is well-defined and real-valued. But Ito's formula cannot be applied in that version we have proved it. The reason for this difficulty is due to the fact that the range of X is not contained in a compact interval where ϕ can be approached uniformly by polynomials.

10.50 Problem. (*advanced*)
Let X be a positive continuous semimartingale.
(a) Show that Ito's formula holds for $\phi(x) = \log(x)$ and for $\phi(x) = 1/x$.
Hint: Let $\tau_n = \min\{t \geq 0 : X_t \geq 1/n\}$. Apply Ito's formula to X^{τ_n} and let $n \to \infty$.
(b) Show that $\phi(X)$ is a semimartingale.

10.51 Problem. (*intermediate*)
Let X be a continuous positive semimartingale. Find $\int_0^t 1/X_s^k \, dX_s$, $k \in \mathbb{N}$.

10.52 Problem. (*intermediate*)
Show that every positive continuous semimartingale X can be written as a stochastic exponential $\mathcal{E}(L)$.
Hint: Note that $dX = X\, dL$ implies $dL = 1/X\, dX$.

10.53 Theorem. *Ito's formula*
Let $X, Y \in \mathcal{S}$ be continuous and let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be twice differentiable with continuous derivatives. Then

$$\begin{aligned}\phi(X_t, Y_t) &= \phi(X_0, Y_0) \\ &\quad + \int_0^t \phi_1'(X_s, Y_s)\, dX_s + + \int_0^t \phi_2'(X_s, Y_s)\, dY_s \\ &\quad + \frac{1}{2}\int_0^t \phi_{11}''(X_s, Y_s)\, d[X]_s + \int_0^t \phi_{12}''(X_s, Y_s)\, d[X,Y]_s + \frac{1}{2}\int_0^t \phi_{22}''(X_s, Y_s)\, d[Y]_s\end{aligned}$$

Proof: The assertion is true for polynomials. Since smooth functions can be approximated uniformly by polynomials in such a way that also the corresponding derivatives are approximated the assertion follows. □

10.54 Problem. (*advanced*)
Show that Ito's formula is true if $\phi : \mathbb{R}^2 \to \mathbb{R}$ is a polynomial.
Hint: Start with powers $\phi(x, y) = x^k y^l$.

10.55 Problem. (*very easy*)
State 10.53 in terms of differentials.

10.56 Problem. (*intermediate*)
State Ito's formula for $\phi(x, t)$.
Hint: Apply 10.53 to $Y_t = t$.

10.57 Problem. (*easy*)
Use 10.53 to derive the differential equation for the stochastic differential.

Chapter 11

Applications to financial markets

11.1 Self-financing trading strategies

Consider a financial market model $\mathcal{M} = (X, Y)$ which is generated by two semi-martingales X and Y. Continuing the discussion 10.15 leads to the following definition.

11.1 Definition. A trading strategy (H^X, H^Y) (consisting of left-continuous adapted processes) is *self-financing* if

$$H_t^X X_t + H_t^Y Y_t = H_0^X X_0 + H_0^Y Y_0 + \int_0^t H^X dX + \int_0^t H^Y dY$$

or in other words

$$d(H^X X) + d(H^Y Y) = H^X dX + H^Y dY$$

The property of being self-financing is a very strong property which narrows the set of available wealth processes considerably. Let us illustrate this fact at hand of continuous models.

Assume that the market model and the trading strategy are continuous. Then from the integration by parts formula we have

$$\begin{aligned} d(H^X X) + d(H^Y Y) &= H^X dX + H^Y dY \\ &+ X dH^X + Y dH^Y + d[X, H^X] + d[Y, H^Y] \end{aligned}$$

If the trading strategy is self-financing then the expression on the second line vanishes.

In financial calculations it is often convenient to change the unit of money. The most simple example is discounting by a fixed interest rate. But there are also important applications where the „numeraire" is a stochastic process. It seems to be intuitively clear that such a change of numeraire should have no influence on the trading strategy and should not destroy the self-financing property. The following theorem shows that this is actually true.

11.2 Theorem. *Assume that the market model and the trading strategy are continuous and*

$$dV := d(H^X X) + d(H^Y Y) = H^X dX + H^Y dY$$

Let Z be a continuous semimartingale. Then

$$d(VZ) = d(H^X XZ) + d(H^Y YZ) = H^X d(XZ) + H^Y d(YZ)$$

Proof: The first equality is obvious. The second follows from

$$\begin{aligned} d(VZ) &= ZdV + VdZ + d[Z,V] \\ &= ZH^X dX + ZH^Y dY + H^X XdZ + H^Y YdZ + d[Z,V] \\ &= H^X(ZdX + XdZ) + H^Y(ZdY + YdZ) + d[Z,V] \\ &= H^X(d(XZ) - d[X,Z]) + H^Y(d(YZ) - d[Y,Z]) + d[Z,V] \\ &= H^X d(XZ) + H^Y d(YZ) + d[Z,V] - H^X d[X,Z] - H^Y d[Y,Z] \end{aligned}$$

□

Assume now that X is a positive continuous semimartingale. Applying the preceding result to $Z = 1/X$ we obtain

$$V_t/X_t = H_t^X + H_t^Y(Y_t/X_t) = V_0/X_0 + \int_0^t H^Y d(Y/X)$$

11.3 Problem.
Show that any wealth process V satisfying

$$V_t/X_t = H_t^X + H_t^Y(Y_t/X_t) = V_0/X_0 + \int_0^t H d(Y/X)$$

for some continuous adapted process H is a self-financing wealth process. Find the corresponding trading strategy.

11.2 Markovian wealth processes

Let $\mathcal{M} = (X^1, X^2, \dots, X^n)$ be a financial market model consisting of Ito-processes

$$dX_t^i = \mu_{it}\, dt + \sigma_{it}\, dW_t.$$

This is a so-called one-factor model since only one Wiener process is responsible for random fluctuations. We assume that the processes (σ_{it}) are positive.

Let V be a wealth process generated by a self-financing trading strategy. The wealth process is called Markovian if there exists a function $f(x,t)$, $x \in \mathbb{R}^n$, $t \geq 0$, such that $V_t = f(X_t, t)$ where $X_t = (X_t^1, X_t^2, \dots, X_t^n)$.

We will show that if the function $f(x,t)$ is smooth then it satisfies necessarily partial differential equations which for special cases are known as Black-Scholes equations.

To begin with we note that the self-financing property implies the existence of a trading strategy $(\phi^1, \phi^2, \dots, \phi^n)$ such that

$$\begin{aligned} f(X_t,t) &= f(X_0,0) + \sum_{i=1}^n \int_0^t \phi_s^i dX_s^i \\ &= \sum_{i=1}^n \phi_t^i X_t^i \end{aligned}$$

On the other hand the Ito-formula gives

$$\begin{aligned} f(X_t,t) = f(X_0,0) &+ \sum_{i=1}^n \int_0^t f_{x_i}(X_s,s) dX_s^i \\ &+ \int_0^t f_t(X_s,s)ds + \frac{1}{2}\sum_{i,j}\int_0^t f_{x_i x_j}(X_s,s)\sigma_{is}\sigma_{js}ds \end{aligned}$$

Both representations are Ito-processes which are equal if both the dW_t-part and the dt-part coincide. The equality of the dW_t-part gives $\phi_t^i = f_{x_i}(X_t,t)$ and thus the first partial differential equation:

$$f(X_t,t) = \sum_{i=1}^n f_{x_i}(X_t,t) X_t^i$$

Comparing the dt-part gives the second partial differential equation:

$$f_t(X_t,t) + \frac{1}{2}\sum_{i,j} f_{x_i x_j}(X_t,t)\sigma_{it}\sigma_{jt} = 0$$

In former times wealth processes have been calculated by solving these partial differential equations by analytical or numerical methods.

11.3 The Black-Scholes market model

The simplest mathematical model of a financial asset is the model of a bank account (B_t) with fixed interest $r > 0$:

$$B_t = B_0 e^{rt} \;\Leftrightarrow\; dB_t = rB_t\,dt \tag{25}$$

Denoting $R_t := rt$ the bank account follows the differential equation

$$dB_t = B_t\, dR_t$$

Stochastic models for financial assets are often based on a stochastic model for the rendite process (R_t).

Assume that $R_t = \mu t + \sigma W_t$ where (W_t) is a Wiener process. If this („generalized Wiener process") is a model of the rendite of an asset (S_t) then it follows that

$$dS_t = S_t\, dR_t = \mu S_t\, dt + \sigma S_t\, dW_t \tag{26}$$

This is a stochastic differential equation. The number $\sigma > 0$ is called the volatility of the asset.

11.4 Problem. Show that $S_t = S_0 e^{(\mu-\sigma^2/2)dt+\sigma W_t}$ is a solution of (26).

11.5 Definition. A *Black-Scholes model* is a market model which is generated by two assets (B_t, S_t) following equations (25) and (26).

Let us give an overview over the available wealth processes in the Black-Scholes model. We begin with smooth Markovian wealth processes.

The Black-Scholes equation

We are going to apply 11.2. A Black-Scholes model consists of the assets $X_t^1 = e^{rt}$ and $X_t^2 = S_t$ where $\sigma_{1t} = 0$ and $\sigma_{2t} = \sigma S_t$. Let $f(X_t^1, X_t^2, t)$ be a self-financing wealth process and define

$$g(x,t) = f(e^{rt}, x, t)$$

The Black-Scholes equation is the partial differential equation for the function $g(x,t)$.

Note that

$$g_t = f_t + f_{x_1} r e^{rt} \quad \text{and} \quad g_x = f_{x_2}$$

Since

$$g = f_{x_1} e^{rt} + f_{x_2} x = f_{x_1} e^{rt} + g_x x$$

we obtain

$$g_t = f_t + r(g - g_x x)$$

From 11.2 we know that

$$f_t = -\frac{1}{2} f_{x_2 x_2} \sigma_2^2 = -\frac{1}{2} g_{xx} \sigma^2 x^2$$

This leads to the famous Black-Scholes equation

$$g_t + \frac{1}{2} g_{xx} \sigma^2 x^2 + r g_x x = rg \tag{27}$$

The market price of risk

Much more insight into the structure of wealth processes is obtained in a different way. We are now going to apply 11.1.

Let (V_t) be a positive wealth process. Then it can be written as

$$dV_t = \mu_t^V V_t dt + \sigma_t^V V_t dW_t$$

Let $\overline{S} := S/B$ and $\overline{V} := V/B$. From the integration by parts formula it follows that

$$d\overline{S}_t = (\mu - r)\overline{S}_t dt + \sigma \overline{S}_t dW_t$$

and

$$d\overline{V}_t = (\mu_t^V - r)\overline{V}_t dt + \sigma_t^V \overline{V}_t dW_t$$

On the other hand we know from 11.1 that for some process τ_t

$$d\overline{V}_t = \tau_t d\overline{S}_t = \tau_t((\mu - r)\overline{S}_t dt + \sigma \overline{S}_t dW_t)$$

This implies

$$\tau_t \overline{S}_t = \frac{\sigma_t^V}{\sigma}\overline{V}_t \quad \text{and} \quad \frac{\mu - r}{\sigma} = \frac{\mu^V - r}{\sigma^V}$$

11.6 Theorem. *Let $\lambda := (\mu - r)/\sigma$ (the „market price of risk”). Then a wealth process*

$$dV_t = \mu_t^V V_t dt + \sigma_t^V V_t dW_t$$

is self-financing iff

$$\mu_t^V - r = \lambda \sigma_t^V, \ t \geq 0. \tag{28}$$

11.7 Problem. Prove 11.6.

11.8 Problem. Show that for a Markovian wealth process equations (27) and (28) are equivalent.

Chapter 12

Stochastic differential equations

12.1 Introduction

A (Wiener driven) *stochastic differential equation* is an equation of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)\, dW_t$$

where $(W_t)_{t\geq 0}$ is a Wiener process and $b(t, x)$ and $\sigma(t, x)$ are given functions. The problem is to find a process $(X_t)_{t\geq 0}$ that satisfies the equation. Such a process is then called a solution of the differential equation.

Note that the differential notation is only an abbreviation for the integral equation

$$X_t = x_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s$$

There are three issues to be discussed for differential equations:

(1) Theoretical answers for existence and uniqueness of solutions.
(2) Finding analytical expressions for solutions.
(3) Calculating solutions by numerical methods.

We will focus on analytical expressions for important but easy special cases. However, let us indicate some issues which are important from the theoretical point of view.

For stochastic differential equations even the concept of a solution is a subtle question. We have to distinguish between weak and strong solutions, even between weak and strong uniqueness. It is not within the scope of this text to give precise definitions of these notions. But the idea can be described in an intuitive way.

A *strong solution* is a solution where the driving Wiener process (and the underlying probability space) is fixed in advance and the solution $(X_t)_{t\geq 0}$ is a function of this given driving Wiener process. A *weak solution* is an answer to the question: Does there exist a probability space where a process $(X_t)_{t\geq 0}$ and a Wiener process $(W_t)_{t\geq 0}$ exist such that the differential equation holds ?

When we derive analytical expressions for solutions we will derive strong solutions. In particular for linear differential equations (to be defined below) complete formulas for strong solutions are available.

There is a general theory giving sufficient conditions for existence and uniqueness of non-exploding strong solutions. Both the proofs and the assertions of this theory are quite similar to the classical theory of ordinary differential equations. We refer to Hunt-Kennedy [12] and Karatzas-Shreve [15].

Let us introduce some terminology.

Any stochastic differential equation is *time homogeneous* if $b(t,x) = b(x)$ and $\sigma(t,x) = \sigma(x)$.

A *linear differential equation* is of the form

$$dX_t = (a_0(t) + a_1(t)X_t)dt + (\sigma_0(t) + \sigma_1(t)X_t)dW_t$$

It is a *homogeneous* linear differential equation if $a_0(t) = \sigma_0(t) = 0$.

The simplest homogeneous case is

$$dX_t = \mu X_t dt + \sigma X_t dW_t$$

which corresponds to the Black Scholes model. The constant σ is called the *volatility* of the model. If the volatility is time dependent then it is a *local volatility* model.

There are plenty of linear differential equations used in the theory of stochastic interest rates. If (B_t) denotes a process that is a model for a bank account with stochastic interest rate then

$$r_t := \frac{B_t'}{B_t} \Leftrightarrow B_t = B_0 e^{\int_0^t r_s ds}$$

is called the *short rate*. Popular short rate models are the *Vasicek model*

$$dr_t = a(b - r_t)dt + \sigma dW_t$$

and the *Hull-White model*

$$dr_t = (\theta(t) - a(t)r_t)dt + \sigma(t)dW_t$$

12.2 The abstract linear equation

Let Y and Z be any continuous semimartingales. The abstract homogeneous linear equation is

$$dX_t = X_t dY_t$$

and its solution is known to us as

$$X_t = x_0 e^{Y_t - [Y]_t/2} = x_0 \mathcal{E}(Y_t)$$

This is the recipe to solve any homogeneous linear stochastic differential equation. There is nothing more to say about it at the moment.

12.1 Problem. (*easy*)
Solve $dX_t = a(t)X_t dt + \sigma(t)X_t dW_t$.

Things become more interesting when we turn to the general inhomogeneous equation

$$dX_t = X_t dY_t + dZ_t$$

There is an explicit expression for the solution but it is much more illuminating to memorize the approach how to arrive there.

The idea is to write the equation as

$$dX_t - X_t dY_t = dZ_t$$

and to find an integrating factor that transforms the left hand side into a total differential.

Let $dA_t = A_t dY_t$ and multiply the equation by $1/A_t$ giving

$$\frac{1}{A_t} dX_t - \frac{X_t}{A_t} dY_t = \frac{1}{A_t} dZ_t \tag{29}$$

Note that

$$d\frac{1}{A_t} = -\frac{1}{A_t} dY_t + \frac{1}{A_t} d[Y]_t$$

Then

$$\begin{aligned} d\Big(\frac{1}{A_t} X_t\Big) &= \frac{1}{A_t} dX_t + X_t d\frac{1}{A_t} + d\Big[\frac{1}{A_t}, X_t\Big] \\ &= \frac{1}{A_t} dX_t - \frac{X_t}{A_t} dY_t + \frac{X_t}{A_t} d[Y]_t - \frac{1}{A_t} d[Y,X]_t \\ &= \frac{1}{A_t} dX_t - \frac{X_t}{A_t} dY_t - \frac{1}{A_t} d[Y,Z]_t \end{aligned}$$

Thus, the left hand side of (29) differs from a total differential by a known BV-function. We obtain

$$d\Big(\frac{1}{A_t} X_t\Big) = \frac{1}{A_t} dZ_t - \frac{1}{A_t} d[Y,Z]_t$$

leading to

$$X_t = A_t\Big(x_0 - \int_0^t \frac{1}{A_s} d[Y,Z]_s + \int_0^t \frac{1}{A_s} dZ_s\Big) \tag{30}$$

Note that the solution is particularly simple if either Y or Z are BV-processes.

12.2 Problem. (*intermediate*)
Fill in and explain all details of the derivation of (30).

12.3 Wiener driven models

The Vasicek model is

$$dX_t = (\nu - \mu X_t)dt + \sigma dW_t$$

For $\nu = 0$ the solution is called the Ornstein-Uhlenbeck process.

The Vasicek model is a special case of the inhomogeneous linear equation for

$$dY_t = -\mu dt \quad \text{and} \quad dZ_t = \nu dt + \sigma dW_t$$

Therefore the integrating factor is $A_t = e^{-\mu t}$ and the solution is obtained as in the case of an ordinary linear differential equation.

12.3 Problem. (*advanced*)
Show that the solution of the Vasicek equation is

$$X_t = e^{-\mu t}x_0 + \frac{\nu}{\mu}(1 - e^{-\mu t}) + \sigma \int_0^t e^{-\mu(t-s)} dW_s$$

12.4 Problem. (*advanced*)
Derive the following properties of the Vasicek model:

(a) The process $(X_t)_{t\ge 0}$ is a Gaussian process (i.e. all joint distribution are normal distributions).

(b) Find $E(X_t)$ and $\lim_{t\to\infty} E(X_t)$.

(c) Find $V(X_t)$ and $\lim_{t\to\infty} V(X_t)$.

(d) Find $\text{Cov}(X_t, X_{t+h})$ and $\lim_{t\to\infty} \text{Cov}(X_t, X_{t+h})$.

12.5 Problem. (*advanced*)
Let $X_0 \sim N\left(\frac{\nu}{\mu}, \frac{\sigma^2}{2\mu}\right)$. Explore the mean and covariance structure of a Vasicek model starting with X_0.

Let us turn to models that are not time homogneous.

12.6 Problem. (*intermediate*)
The *Brownian bridge*:

(a) Find the solution of

$$dX_t = -\frac{1}{1-t} X_t dt + dW_t,\ 0 \le t < 1.$$

(b) Show that $(X_t)_{t\ge 0}$ is a Gaussian process. Find the mean and the covariance structure.

(c) Show that $X_t \to 0$ if $t \to 1$.

12.7 Problem. (*intermediate*)
Find the solution of the Hull-White model:

$$dX_t = (\theta(t) - a(t)X_t)dt + \sigma(t)dW_t$$

Finally, let us consider a nonlinear model.

12.8 Problem. (*advanced*)
Let $Z_t = \mathcal{E}(\mu t + \sigma W_t)$.
(a) For $a > 0$ find the differential equation of

$$X_t := \frac{Z_t}{1 + a\int_0^t Z_s ds}$$

(b) What about $a < 0$?

Chapter 13

Martingales and stochastic calculus

13.1 Martingale properties of the stochastic integral

Facts

Let $(M_t)_{t\geq 0}$ be a continuous square integrable martingale, i.e. $E(M_t^2) < \infty$, $t \geq 0$. We would like to know for which $H \in \mathcal{L}_0$

$$H \bullet M : t \mapsto \int_0^t H \, dM$$

is a square martingale.

There are two main results in this section.

13.1 Theorem. *For any continuous square integrable martingale $(M_t)_{t\geq 0}$ the process $M_t^2 - [M]_t$ is a martingale.*

13.2 Theorem. *Let $(M_t)_{t\geq 0}$ be a continuous square integrable martingale and $H \in \mathcal{L}_0$. Then $H \bullet M$ is a square integrable martingale for $t \in [0, T]$ iff $E([H \bullet M]_T) < \infty$.*

We will outline the proofs at the end of the section. At this point we attempt to understand the assertions and their consequences.

First we note that 13.1 is known to us for the Wiener process. Thus, it is a generalization of a familiar structure.

For a better understanding of 13.2 we note that

$$[H \bullet M]_T = \int_0^T H_s^2 d[M]_s$$

Therefore $\int_0^t H_s dM_s$, $t \leq T$, is a square integrable martingale iff

$$E\Big(\int_0^T H_s^2 d[M]_s \Big) < \infty$$

Thus, we have to check the P-integrability of a Stieltjes integral. For Wiener driven martingales this is even an ordinary Lebesgue integral.

If the condition is satisfied then by 13.1 it follows that

$$\Big(\int_0^t H_s dM_s\Big)^2 - \int_0^t H_s^2 d[M]_s,\ t \le T,$$

is a martingale which means that

$$E\Big(\Big(\int_0^t H_s dM_s\Big)^2\Big) = E\Big(\int_0^t H_s^2 d[M]_s\Big)$$

This is one of the most important identities of stochastic analysis. It was the original starting point of the construction of the stochastic integral and it is still the starting point of further extensions of the stochastic integral to larger entities than $\mathcal{L}_0$.

By the way, what we did (Protter's [19] approach "without tears") is the stochastic counterpart of the Cauchy-Stieltjes integral. The most general version of the stochastic integral (being not the subject of this text) could be considered as the stochastic counterpart of abstract (Lebesgue) integration theory.

Let us mention that the assertion of 13.1 is related to 13.2 by

$$M_t^2 - [M]_t = M_0^2 + 2\int_0^t M_s dM_s$$

This implies that $M \bullet M$ is a martingale. However, it is not necessarily a square integrable martingale !

13.3 Problem. (*intermediate*)

Show that every continuous square integrable martingale of bounded variation is necessarily constant.

13.4 Problem. (*intermediate*)

Let $(M_t)_{t\ge 0}$ be a continuous square integrable martingale. If (A_t) is a continuous adapted process of bounded variation such that $M_t^2 - A_t$ is a martingale, then $A_t = [M]_t$.

Proofs

Now, let us turn to the proofs of 13.1 and 13.2. For warming up we provide some straightforward facts.

13.5 Problem. (*advanced*)

For every continuous martingale $(M_t)_{t\ge 0}$ we have $E(M_t^2) \ge E([M]_t)$.
Hint: Show that for any subdivision $0 = t_0 < t_1 < \ldots < t_n = t$

$$E(M_t^2) = E\Big(\sum_{j=1}^n (M_{t_j} - M_{t_{j-1}})^2\Big)$$

and apply Fatou's lemma to an appropriate subsequence of a Riemannian sequence.

13.6 Problem. (*easy*)
Prove the "only if" part in 13.2.

13.7 Problem. (*intermediate*)
Suppose you know that for every continuous square integrable martingale $(M_t)_{t\geq 0}$ the equation $E(M_t^2) = E([M]_t)$, $t \geq 0$, is true. Show that this implies that $M_t^2 - [M]_t$ is even a martingale.
Hint: Apply 8.25.

Next we prove a preliminary assertion. The proof isolates some arguments which are related to the martingale structure.

13.8 Lemma. *Let $(M_t)_{t\geq 0}$ be a continuous square integrable martingale. Then $H \bullet M$ is a square integrable martingale for every bounded $H \in \mathcal{L}_0$.*

Proof: It is sufficient to show that $E(\int_0^t H\, dM) = 0$.

Let $0 = t_0 < t_1 < \ldots < t_n = t$ be a the n-th element of a Riemannian sequence of subdivisions and define

$$H_n = \sum_{j=1}^n H_{t_{j-1}} 1_{(t_{j-1}, t_j]}$$

Then $E(\int_0^t H_n dM) = 0$ and $\int_0^t H_n dM \xrightarrow{P} \int_0^t H\, dM$. It remains to show that $E([\int_0^t H_n dM]^2)$ is bounded. For this, note that

$$\begin{aligned} E\Big(\Big[\int_0^t H_n dM\Big]^2\Big) &= E\Big(\Big[\sum_{j=1}^n H_{t_{j-1}}(M_{t_j} - M_{t_{j-1}})\Big]^2\Big) \\ &= \sum_{j=1}^n E(H_{t_{j-1}}^2 (M_{t_j} - M_{t_{j-1}})^2) \\ &\leq C \sum_{j=1}^n E((M_{t_j} - M_{t_{j-1}})^2) = C(E(M_t^2) - E(M_0^2)) \end{aligned}$$

□

Now we are in the position to prove 13.1.

Proof: (of Theorem 13.1) For a bounded martingale the assertion follows from the integration by parts formula and 13.8. For proving the general case it is sufficient to show that $E(M_t^2) = E([M]_t)$.

Recall that for any stopping time τ the identity $[M^\tau]_t = [M]_{t\cap\tau}$ holds. Let

$$\tau_n = \inf\{t : |M_t| \geq n\}$$

Then it follows that

$$E(M^2_{t\cap\tau_n}) = E((M^{\tau_n}_t)^2) = E([M^{\tau_n}]_t) = E([M]_{t\cap\tau_n})$$

Letting $n \to \infty$ it is clear that $E([M]_{t\cap\tau_n}) \to E([M]_t)$. The corresponding convergence of the left hand side follows from $M^2_{t\cap\tau} = E(M_t|\mathcal{F}_\tau)^2 \le E(M^2_t|\mathcal{F}_\tau)$. □

The following assertion is a continuation of 13.8.

13.9 Lemma. *Let $(M_t)_{t\ge 0}$ be a continuous square integrable martingale. Then $E((H \bullet M)^2_t) = E([H \bullet M]_t)$ for every bounded $H \in \mathcal{L}_0$.*

Proof: With the aid of 13.1 it follows that

$$E((M_t - M_s)^2|\mathcal{F}_s) = E([M]_t - [M]_s|\mathcal{F}_s)$$

Then the equation array of the proof of 13.8 can be improved to

$$\begin{aligned} E\Big(\Big[\int_0^t H_n dM\Big]^2\Big) &= E\Big(\Big[\sum_{j=1}^n H_{t_{j-1}}(M_{t_j} - M_{t_{j-1}})\Big]^2\Big) \\ &= \sum_{j=1}^n E(H^2_{t_{j-1}}(M_{t_j} - M_{t_{j-1}})^2) \\ &= \sum_{j=1}^n E(H^2_{t_{j-1}}([M]_{t_j} - [M]_{t_{j-1}})) = E\Big(\int_0^t H^2_n d[M]\Big) \end{aligned}$$

This is extended to bounded $H \in \mathcal{L}_0$ by routine arguments. □

Proof: (of Theorem 13.2) We need only prove the "if"-part and for this it is sufficient to prove that 13.9 extends to arbitrary $H \in \mathcal{L}_0$.

Let $\tau_n := \inf\{t : |H_t| \ge n\}$. Then (by left-continuity !) H^{τ_n} is bounded and tends to H. Lemma 13.9 can be applied to H^{τ_n} and the assertion is proved again by routine arguments. □

13.2 Martingale representation

Let $(W_t)_{t\ge 0}$ be a Wiener process. We know that

$$\int_0^t H_s dW_s,\ t \ge 0,$$

is a square integrable martingale iff

$$E\Big(\int_0^t H^2_s ds\Big) < \infty,\ t \ge 0.$$

Now, in this special case there is a remarkable converse: Each square integrable martingale arises in this way !

We have to be a bit more modest: If we confine ourselves (as we have done so far) to $H \in \mathcal{L}_0$ (left-continuous adapted processes) then all square integrable martingales can only be approximated which arbitrary precision by stochastic integrals. We will comment this point later.

The martingale representation fact is an easy consequence of the following seemingly simpler assertion:

Each random variable $C \in L^2(\mathcal{F}_t)$ (each "claim") can be (approximately) written as a stochastic integral ("hedged" by a self-financing strategy).

Let us introduce some simplifying terminology.

13.10 Definition. A set $\mathcal{C}$ of random variables in $L^2(\mathcal{F}_t)$ is called *dense* if for every $C \in L^2(\mathcal{F}_t)$ there is a sequence $C_n \subseteq \mathcal{C}$ such that $E((C_n - C)^2) \to 0$.

A set $\mathcal{C}$ of random variables in $L^2(\mathcal{F}_t)$ is called *total* if the linear hull of $\mathcal{C}$ is dense.

Thus, we want to prove

13.11 Theorem. *The set of all integrals $\int_0^t H\,dW$ with $H \in \mathcal{L}_0$ and $E(\int_0^t H_s^2 ds) < \infty$ is dense in $L^2(\mathcal{F}_t)$.*

Proof: The starting point is that $\mathcal{F}_t$ is generated by $(W_s)_{s \le t}$ and therefore also by $(e^{W_s})_{s \le t}$. Therefore an obvious dense set consists of the functions

$$\phi(e^{W_{s_1}}, e^{W_{s_2}}, \ldots, e^{W_{s_n}}),$$

where ϕ is some continuous function with compact support and $s_1, s_2, \ldots, s_n$ is some finite subset of $[0, t]$. Every continuous function can be approximated uniformly by polynomials (Weierstrass' theorem) and polynomials are linear combinations of powers. Thus, we arrive at a total set consisting of

$$\exp\Big(\sum_{j=1}^n k_j W_{s_j}\Big)$$

which after reshuffling can be written as

$$\exp\Big(\sum_{j=1}^n a_{j-1}(W_{s_j} - W_{s_{j-1}})\Big) = \exp\Big(\int_0^t f(s)\,dW_s\Big) \tag{31}$$

for some bounded left-continuous (step) function $f : [0, t] \to \mathbb{R}$. It follows that the set of functions (differring from (31) by constant factors)

$$G_t = \exp\Big(\int_0^t f(s)\,dW_s - \frac{1}{2}\int_0^t f^2(s)\,ds\Big)$$

is total when f varies in the set of all bounded left-continuous (step) functions $f : [0,t] \to \mathbb{R}$.

Recall that $(G_s)_{s\le t}$ is a square integrable martingale and satisfies

$$G_t = 1 + \int_0^t G\,d(f \bullet W) = \int_0^t G_s f(s) dW_s$$

From 13.1 it follows that

$$E\Big(\int_0^t G_s^2 f^2(s)\,ds\Big) < \infty.$$

Therefore, the set of integrals

$$\int_0^t H_s dW_s \text{ where } H \in \mathcal{L}_0 \quad \text{and} \quad E\Big(\int_0^t H_s^2\,ds\Big) < \infty$$

is total and by linearity of the integral even dense. □

UNDER CONSTRUCTION:
Extension to predictable processes.
Representation of martingales.

13.3 Levy's theorem

UNDER CONSTRUCTION

13.4 Exponential martingale and Girsanov's theorem

UNDER CONSTRUCTION

Chapter 14

Pricing of claims

UNDER CONSTRUCTION

Part III

Appendix

Chapter 15

Foundations of modern analysis

Futher reading: Dieudonné, [8].

15.1 Sets and functions

15.1 Problem. (*easy*) Prove de Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c$$

15.2 Problem. (*intermediate*) Prove de Morgan's laws:

$$\Big(\bigcup_{i\in\mathbb{N}} A_i\Big)^c = \bigcap_{i\in\mathbb{N}} A_i^c, \quad \Big(\bigcap_{i\in\mathbb{N}} A_i\Big)^c = \bigcup_{i\in\mathbb{N}} A_i^c$$

cartesian products, rectangles

Let X and Y be non-empty sets.

A *function* $f : X \to Y$ is a set of pairs $(x, f(x)) \in X \times Y$ such that for every $x \in X$ there is exactly one $f(x) \in Y$. X is the *domain* of f and Y is the *range* of f.

A function $f : X \to Y$ is *injective* if $f(x_1) = f(x_2)$ implies $x_1 = x_2$. It is surjective if for every $y \in Y$ there is $x \in X$ such that $f(x) = y$. If a function is injective and surjective then it is *bijective*.

If $A \subseteq X$ then $f(A) := \{f(x) : x \in A\}$ is the *image* of A under f. If $B \subseteq Y$ then $f^{-1}(B) := \{x : f(x) \in B\}$ is the *inverse image* of B under f.

15.3 Problem. (*easy*)

Show that:

(a) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$.

(b) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

(c) $f^{-1}(B^c) = (f^{-1}(B))^c$
(d) Extend (a) and (b) to families of sets.

15.4 Problem. (*easy*)
Show that:
(a) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.
(b) $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$.
(c) Give an example where inequality holds in (b).
(d) Show that for injective functions equality holds in (b).
(e) Extend (a) and (b) to families of sets.

15.5 Problem. (*easy*)
Show that:
(a) $f(f^{-1}(B)) = f(X) \cap B$
(b) $f^{-1}(f(A)) \supseteq A$

Let $f : X \to Y$ and $g : Y \to Z$. Then the composition $g \circ f$ is the function from X to Z such that $(g \circ f)(x) = g(f(x))$.

15.6 Problem. (*easy*)
Let $f : X \to Y$ and $g : Y \to Z$. Show that $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$, $C \subseteq Z$.

15.2 Sequences of real numbers

The set $\mathbb{R}$ of real numbers is well-known, at least regarding its basic algebraic operations. Let us talk about topological properties of $\mathbb{R}$.

The following is not intended to be an introduction to the subject, but a checklist which should be well understood or otherwise an introductory textbook has to be consulted.

A (open and connected) neighborhood of $x \in \mathbb{R}$ is an open interval (a, b) which contains x. Note that neighborhoods can be very small, i.e. can have any length $\epsilon > 0$.

Let us start with sequences. An (infinite) sequence is a function form $\mathbb{N} \to \mathbb{R}$, denoted by $n \mapsto x_n$, for short (x_n), where $n = 1, 2, \ldots$. When we say that an assertion holds for almost all x_n then we mean that it is true for all x_n, beginning with some index N, i.e. for x_n with $n \geq N$ for some N.

A number $x \in \mathbb{R}$ is called a limit of (x_n) if every neighborhood of x contains almost all x_n. In other words: The sequence (x_n) converges to x: $\lim_{n\to\infty} x_n = x$ or $x_n \to x$. A sequence can have at most one limit since two different limits could be put into disjoint neighborhoods.

A fundamental property of $\mathbb{R}$ is the fact that any bounded increasing sequence has a limit which implies that every bounded monotone sequence has a limit. This is not a theorem but the completeness axiom. It is an advanced mathematical construction to show that there exists $\mathbb{R}$, i.e. a set having the familiar properties of real numbers including completeness.

An increasing sequences (x_n) which is not bounded is said to diverge to ∞ ($x_n \uparrow \infty$), i.e. for any a we have $x_n > a$ for almost all x_n. Thus, we can summarize: An increasing sequence either converges to some real number (iff it is bounded) or diverges to ∞ (iff it is unbounded). A similar assertion holds for decreasing sequences.

A simple fact which is an elementary consequence of the order structure says that every sequence has a monotone subsequence.

Putting terms together we arrive at a very important assertion: Every bounded sequence (x_n) has a convergent subsequence. The limit of a subsequence is called an accumulation point of the original sequence (x_n). In other words: Every bounded sequence has at least one accumulation point. An accumulation point x can also be explained in the follwing way: Every neighborhood of x contains infinitely many x_n, but not necessarily almost all x_n. A sequence can have many accumulation points, and it is not necessarily bounded to have accumulation points. A sequence has a limit iff it is bounded and has only one accumulation point, which then is necessarily the limit.

There is a popular criterion for convergence of a sequence which is related to the assertion just stated. Call a sequence (x_n) a Cauchy-sequence if there exist arbitrarily small intervals containing almost all x_n. Cleary every convergent sequence is a Cauchy-sequence. But also the converse is true in view of completeness. Indeed, every Cauchy-sequence is bounded and can have at most one accumulation point. By completeness it has at least one accumulation point, and is therefore convergent.

15.3 Real-valued functions

UNDER CONSTRUCTION

15.4 Banach spaces

Let V be a vector space.

15.7 Definition. A norm on V is a function $v \mapsto ||v||$, $v \in V$, satisfying the following conditions:
(1) $||v|| \geq 0, \quad ||v|| = 0 \;\Leftrightarrow\; v = o$,
(2) $||v + w|| \leq ||v|| + ||w||,\; v, w \in V$,
(3) $||\lambda v|| \leq |\lambda|\,||v||,\; \lambda \in \mathbb{R},\, v \in V$.

A pair $(V, ||.||)$ consisting of a vector space V and a norm $||.||$ is a normed space.

15.8 Example.

(1) $V = \mathbb{R}$ is a a normed space with $||v|| = |v|$.

(2) $V = \mathbb{R}^d$ is a normed space under several norms. E.g.

$$||\mathbf{v}||_1 = \sum_{i=1}^{d} |v_i|, \quad ||\mathbf{v}||_2 = \Big(\sum_{i=1}^{d} v_i^2\Big)^{1/2} \text{ (Euclidean norm)}, \quad ||\mathbf{v}||_\infty = \max_{1 \le i \le d} |v_i|$$

(3) Let $V = C([0,1])$ be the set of all continuous functions $f : [0,1] \to \mathbb{R}$. This is a vector space. Popular norms on this vector space are

$$||f||_\infty = \max_{0 \le s \le 1} |f(s)|$$

and

$$||f||_1 = \int_0^1 |f(s)|\, ds$$

The distance of two elements of V is defined to be

$$d(v, w) := ||v - w||$$

This function has the usual properties of a dstance, in particular satisfies the triangle inequality. A set of the form

$$B(v, r) := \{w \in V : ||w - v|| < r\}$$

is called an open ball around v with radius r. As sequence $(v_n) \subseteq V$ is convergent with limit v if $||v_n - v|| \to 0$.

A sequence (v_n) is a Cauchy-sequence if there exist arbitrarily small balls containing almost all members of the sequence, i.e.

$$\forall \epsilon > 0 \ \exists N(\epsilon) \in \mathbb{N} \text{ such that } ||v_n - v_m|| < \epsilon \text{ whenever } n, m \ge N(\epsilon)$$

15.9 Definition. A normed space is a Banach space if it is complete, i.e. if every Cauchy sequence is convergent.

It is clear that $\mathbb{R}$ and $\mathbb{R}^d$ are complete under the usual norms. Actually they are complete under any norm. The situation is completely different with infinite dimensional normed spaces.

15.10 Problem. (*easy for mathematicians*)
Show that $C([0,1])$ is complete under $||.||_\infty$.

15.11 Problem. (*easy for mathematicians*)
Show that $C([0,1])$ is not complete under $||.||_1$.

The latter fact is one of the reasons for extending the notion and the range of the elementary integral.

15.5 Hilbert spaces

A special class of normed spaces are inner product spaces. Let V be a vector space.

15.12 Definition. An inner product on V is a function $(v, w) \mapsto < v, w >$, $v, w \in V$, satisfying the following conditions:
(1) $(v, w) \mapsto < v, w >$ is linear in both variables,
(2) $< v, v) \geq 0$, $< v, v >= 0 \;\Leftrightarrow\; v = o$.
A pair $(V, < ., . >)$ consisting of a vector space V and an inner product $< ., . >$ is an inner product space.

An inner product gives rise to a norm according to

$$||v|| :=< v, v >^{1/2},\; v \in V.$$

15.13 Problem. (*easy*)
Show that $||v|| :=< v, v >^{1/2}$ is a norm.

15.14 Example.
(1) $V = \mathbb{R}$ is an inner product space with $< v, w >= vw$. The corresponding norm is $||v|| = |v|$.
(2) $V = \mathbb{R}^d$ is an inner product space with

$$< \boldsymbol{v}, \boldsymbol{w} >= \sum_{i=1}^{d} v_i w_i$$

The corresponding norm is $||v||_2$.
(3) Let $V = C([0, 1])$ be the set of all continuous functions $f : [0, 1] \to \mathbb{R}$. This is an inner product space with

$$< f, g >= \int_0^1 f(s)g(s)\, ds$$

The corresponding norm is

$$||f||_2 = \Big(\int_0^1 f(s)^2\, ds \Big)^{1/2}$$

15.15 Definition. An inner product space is a Hilbert space if it is complete under the norm defined by the inner product.

15.16 Problem. (*easy for mathematicians*)
Show that $C([0, 1])$ is not complete under $||.||_2$.

Inner product spaces have a geometric structure which is very similar to that of $\mathbb{R}^d$ endowed with the Euclidean inner product. In particular, the notions of orthogonality and of projections are available on inner product spaces. The existence of orthogonal projections depends on completeness, and therefore requires Hilbert spaces.

15.17 Problem. (*intermediate*)
Let C be a closed convex subset of an Hilbert space $(V, < .,. >)$ and let $v \notin C$. Show that there exists $v_0 \in C$ such that

$$||v - v_0|| = \min\{||v - w|| : w \in C\}$$

Hint: Let $\alpha := \inf\{||v - w|| : w \in C\}$ and choose a sequence $(w_n) \subseteq C$ such that $||v - w_n|| \to \alpha$. Apply the parallelogram equality to show that (w_n) is a Cauchy sequence.

Bibliography

[1] Heinz Bauer. *Probability theory. Translated from the German by Robert B. Burckel.* de Gruyter Studies in Mathematics. 23. Berlin: Walter de Gruyter. xv, 523 p. , 1996.

[2] Heinz Bauer. *Measure and integration theory. Transl. from the German by Robert B. Burckel.* de Gruyter Studies in Mathematics. 26. Berlin: de Gruyter. xvi, 230 p. , 2001.

[3] Tomasz R. Bielecki and Marek Rutkowski. *Credit risk: Moldelling, valuation and hedging.* Springer Finance. Berlin: Springer. xviii, 500 p. , 2002.

[4] Tomas Bjoerk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 2004.

[5] Pierre Bremaud. *Point processes and queues. Martingale dynamics.* Springer Series in Statistics. New York - Heidelberg - Berlin: Springer- Verlag. XIX, 354 p. , 1981.

[6] Pierre Brémaud. *An introduction to probabilistic modeling.* Undergraduate Texts in Mathematics. New York etc.: Springer-Verlag. xvi, 207 p. , 1988.

[7] Pierre Brémaud. *Markov chains. Gibbs fields, Monte Carlo simulation, and queues.* Texts in Applied Mathematics. New York, NY: Springer. xviii, 444 p. , 1999.

[8] Jean Dieudonné. *Foundations of modern analysis. Enlarged and corrected printing.* New York-London: Academic Press. XV, 387 p. , 1969.

[9] Michael U. Dothan. *Prices in financial markets.* New York etc.: Oxford University Press. xv, 342 p. , 1990.

[10] Edwin Hewitt and Karl Stromberg. *Real and abstract analysis. A modern treatment of the theory of functions of a real variable. 3rd printing.* Graduate Texts in Mathematics. 25. New York - Heidelberg Berlin: Springer-Verlag. X, 476 p. , 1975.

[11] John C. Hull. *Options, futures, and other derivatives. 5th ed.* Prentice-Hall International Editions. Upper Saddle River, NJ: Prentice Hall. xxi, 744 p. , 2003.

[12] P.J. Hunt and J.E. Kennedy. *Financial derivatives in theory and practice. Revised ed.* Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons. xxi, 437 p. , 2004.

[13] Albrecht Irle. *Financial mathematics. The evaluation of derivatives. (Finanzmathematik. Die Bewertung von Derivaten) 2., überarbeitete und erweiterte Auflage.* Teubner Studienbücher Mathematik. Stuttgart: Teubner. 302 S. , 2003.

[14] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes. 2nd ed.* Grundlehren der Mathematischen Wissenschaften. 288. Berlin: Springer., 2003.

[15] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics, 113. New York etc.: Springer-Verlag. xxiii, 470 p. , 1991.

[16] Ioannis Karatzas and Steven E. Shreve. *Methods of mathematical finance.* Applications of Mathematics. Berlin: Springer. xv, 407 p. , 1998.

[17] Marek Musiela and Marek Rutkowski. *Martingale methods in financial modelling. 2nd ed.* Stochastic Modelling and Applied Probability 36. Berlin: Springer. xvi, 636 p. , 2005.

[18] Salih N. Neftci. *Introduction to the mathematics of financial derivatives. 2nd ed.* Orlando, FL: Academic Press. xxvii, 527 p. , 2000.

[19] Philip Protter. Stochastic integration without tears (with apology to P. A. Meyer). *Stochastics*, 16:295–325, 1986.

[20] Philip E. Protter. *Stochastic integration and differential equations. 2nd ed.* Applications of Mathematics 21. Berlin: Springer. xiii, 2004.

[21] A.N. Shiryaev. *Probability. Transl. from the Russian by R. P. Boas. 2nd ed.* Graduate Texts in Mathematics. 95. New York, NY: Springer-Verlag. xiv, 609 p. , 1995.

[22] Paul Wilmott. *Paul Wilmott on Quanitative Finance, Volume One*. John Wiley and Sons, 2000.

[23] Paul Wilmott. *Paul Wilmott on Quanitative Finance, Volume Two*. John Wiley and Sons, 2000.