ST908 Stochastic Calculus for Finance

Lecture notes by Martin Herdegen[1]

Department of Statistics
University of Warwick

This version: December 2, 2021

[1] I would like to thank the students Rob Buck, Stylianos Demosthenous, Jón Halldórsson, Luke Jeeves, Christopher Jennings, Ivan Milokhin, Juan Carlos Ramirez Cano, Shanaya Sheth, Oscar Spevack, László Udvardi, Jie Yeo and Yating Yuan, for spotting typos and mistakes in previous versions. Moreover, I am grateful to my teaching assistants Nikolaos Constantinou, Joseph Jerome and Osian Shelley for fruitful discussions on the material. Any remaining typos and mistakes are of course mine.

Contents

0 User's guide **4**

1 Conditional expectations **5**
1.1 Elementary conditional expectations . . . 5
1.2 Measure-theoretic conditional expectations . . . 5
1.3 Properties of conditional expectations . . . 8

2 Martingale Theory **12**
2.1 Stochastic processes and filtrations . . . 12
2.2 Martingales, submartingales, and supermartingales . . . 13
2.3 Discrete stochastic integral . . . 16
2.4 Stopping times . . . 18
2.5 The stopping theorem . . . 20
2.6 Maximal inequalities . . . 24
2.7 Doob's supermartingale convergence theorem . . . 25
2.8 L^p-martingale convergence theorem . . . 28

3 Pricing and hedging in discrete time **32**
3.1 Financial market . . . 32
3.2 Self-financing strategies . . . 34
3.3 No-arbitrage and the Fundamental Theorem of Asset Pricing . . . 37
3.4 Valuation of contingent claims . . . 40
3.5 Complete markets and the predictable representation property . . . 43
3.6 Pricing and hedging in the binomial model . . . 45

4 Brownian motion and continuous local martingales **49**
4.1 Definition and existence of Brownian motion . . . 49
4.2 Fundamental properties of Brownian motion . . . 51
4.3 The variation of Brownian motion . . . 52
4.4 Square-integrable martingales and their quadratic variation . . . 55
4.5 Continuous local martingales . . . 58

5 Theory of stochastic integration **63**
5.1 Predictable processes in continuous time . . . 63
5.2 Integral with respect to a continuous finite variation process . . . 65
5.3 Stochastic integral with respect to an L^2-bounded martingale . . . 68
5.4 Stochastic integral with respect to a continuous local martingale . . . 71
5.5 Stochastic integral with respect to a continuous semimartingale . . . 75

6 Stochastic Calculus **78**
6.1 Itô's formula in one dimension . . . 78
6.2 Itô's formula in multiple dimension and the product formula . . . 82
6.3 Bayes' theorem . . . 83
6.4 Girsanov's theorem . . . 86

6.5 Itô's representation theorem . . . 88

7 Stochastic differential equations **90**

7.1 Strong solutions . . . 90
7.2 Weak solutions . . . 95
7.3 Feynman–Kac formula . . . 97

8 Pricing and hedging in continuous time **100**

8.1 Financial market . . . 100
8.2 Admissible strategies . . . 101
8.3 No-arbitrage and the fundamental theorem of asset pricing . . . 104
8.4 Valuation of contingent claims . . . 106
8.5 Complete markets and the predictable representation property . . . 107
8.6 Pricing and hedging in the Black–Scholes model . . . 109

A Fundamental concepts of Probability Theory **115**

A.1 Probability spaces . . . 115
A.2 Random variables . . . 119
A.3 Distribution of random variables: univariate case . . . 122
A.4 Distribution of random variables: multivariate case . . . 125
A.5 Conditional probabilities . . . 126
A.6 Independence . . . 127
A.7 Expectation . . . 130
A.8 L^p-spaces . . . 136
A.9 Variance and covariance . . . 138
A.10 The inequalities of Markov and Jensen . . . 139
A.11 Product spaces and Fubini's theorem . . . 141
A.12 Convergence of random variables . . . 143
A.13 Uniform integrability and the dominated convergence theorem . . . 146
A.14 The laws of large numbers . . . 148
A.15 The law of iterated logarithm and the central limit theorem . . . 150

B The usual conditions **151**

0 User's guide

These lecture notes provide a self-contained introduction to stochastic calculus in discrete and continuous time as well as some applications to mathematical finance. They *assume* a sound knowledge of *measure-theoretic probability theory*. A self-contained summary of fundamental concepts of probability theory can be found in Appendix A. We also refer to the excellent textbook by Klenke [9].

Most sections contain some exercises. The stars indicate their level of difficulty (✫☆☆ easy, ★✫☆ medium, ★★✫ hard). They are intended to help the reader to check if they understand the respective material.

1 Conditional expectations

Perhaps *the* central object of mathematical finance is a *martingale*. In order to define and study them, we first need to introduce the key concept of *conditional expectations*.

1.1 Elementary conditional expectations

If $(\Omega, \mathcal{F}, P)$ is a probability space and $A \in \mathcal{F}$ with $P[A] > 0$, the conditional probability of $B \in \mathcal{F}$ given A is given by $P[B \mid A] := \frac{P[A \cap B]}{P[A]}$ and $P[\cdot \mid A]$ is again an probability measure on $(\Omega, \mathcal{F})$; see Section A.5 for details.

The following definition looks at the corresponding *conditional expectation*.

Definition 1.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space, X a P-integrable $\mathcal{F}$-measurable random variable, and $A \in \mathcal{F}$ with $P[A] > 0$. Then the (elementary) *conditional expectation of* X *given* A is defined as

$$E^P[X \mid A] := E^{P[\cdot \mid A]}[X] = \frac{E^P[X \mathbf{1}_A]}{P[A]}. \tag{1.1}$$

Note that the equality in (1.1) follows for simple random variables by the definition of $P[\cdot \mid A]$ and linearity of the expectation. For nonnegative random variables, it follows by monotone convergence, and for general random variables it follows by splitting X into its positive and negative part and linearity of the expectation.

Exercise 1.2. ★☆☆ Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X \sim \mathcal{N}(0, 1)$ a standard normal random variable. Calculate $E[X \mid X < 0]$.

1.2 Measure-theoretic conditional expectations

The elementary conditional expectation considers conditioning on *single events*. The measure-theoretic conditional expectation lifts this idea to conditioning on whole σ-algebras. Before giving the precise definition, let us explain the underlying idea. Consider a random variable X on some probability space $(\Omega, \mathcal{F}, P)$ and assume that the (unknown) state of the world is $\omega \in \Omega$.

If we have *full information*, i.e., we know $\mathcal{F}$, we can *observe* each event $A \in \mathcal{F}$, i.e., we can assert if $\omega \in A$ or $\omega \notin A$. In particular, as $\{X = x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$, we can *fully observe* X. By contrast, if we only have *partial information*, i.e., we are given a *sub-σ-algebra* $\mathcal{G} \subset \mathcal{F}$, we can only observe an event A if $A \in \mathcal{G}$. In particular, as $\{X = x\}$ may not be in $\mathcal{G}$, we may no longer be able to fully observe X. The extreme case is that we have *trivial information*, i.e., $\mathcal{G} = \{\emptyset, \Omega\}$, so we can only assert that $\omega \in \Omega$. So how we can make a *best prognosis* for X if we have only partial information, i.e., if we are given a sub-σ-algebra $\mathcal{G}$. For trivial information, i.e., $\mathcal{G} = \{\emptyset, \Omega\}$, this best prognosis is the expectation $E[X]$. The conditional expectation generalises the concept of expectation to general partial information.

Definition 1.3. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra, and X a random variable with either $X \geq 0$ or $X \in \mathcal{L}^1(P)$. Any *random variable* Y with either $Y \geq 0$ (if $X \geq 0$) or $Y \in \mathcal{L}^1(P)$ (if $X \in \mathcal{L}^1(P)$) such that

(1) Y is $\mathcal{G}$-measurable,

(2) $E[Y\mathbf{1}_A] = E[X\mathbf{1}_A]$ for all $A \in \mathcal{G}$,

is called (a version of) *the conditional expectation of* X *given* $\mathcal{G}$, and we write $Y = E[X \,|\, \mathcal{G}]$.

The random variable Y in Definition 1.3 is to be interpreted as the best prognosis for X given the information $\mathcal{G}$. The *measurability property* (1) ensures that Y only uses the information given in $\mathcal{G}$, and the *averaging property* (2) ensures that Y is indeed the best prognosis.

The following result gives existence and uniqueness of conditional expectations.

Theorem 1.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra, and X a random variable with $X \geq 0$ or $X \in \mathcal{L}^1(P)$.*

(a) *The conditional expectation $E[X \,|\, \mathcal{G}]$ exists.*

(b) *The conditional expectation is P-a.s. unique, i.e., if Y and Y' are random variables with $Y, Y' \geq 0$ (if $X \geq 0$) or $Y, Y' \in \mathcal{L}^1(P)$ (if $X \in \mathcal{L}^1(P)$) satisfying properties* (a) *and* (b) *in Definition 1.3, then $Y = Y'$ P-a.s.*

Proof. We only prove part (b) and only in the case that $X \in \mathcal{L}^1(P)$. For a proof of part (a) in the case that $X \in \mathcal{L}^1(P)$, see [9, Theorem 8.12].

To establish (b) for $X \in \mathcal{L}^1(P)$, we show more generally that the conditional expectation is *monotone*, i.e., if $X \leq X'$ P-a.s. with $X, X' \in \mathcal{L}^1(P)$ then $E[X \,|\, \mathcal{G}] \leq E[X' \,|\, \mathcal{G}]$ P-a.s. For $X = X'$, this gives uniqueness of $E[X \,|\, \mathcal{G}]$.

To show monotonicity of conditional expectations, let $X \leq X'$ P-a.s. with $X, X' \in \mathcal{L}^1(P)$ and $Y, Y' \in \mathcal{L}^1(P)$ be a random variables that satisfy properties (1) and (2) in Definition 1.3 with respect to X and X', respectively. Set

$$A := \{Y > Y'\}.$$

Then $A \in \mathcal{G}$ by the measurability property (1) of conditional expectations. Since $X \leq X'$ P-a.s., the averaging property (2) of conditional expectations together with monotonicity of (ordinary) expectations gives

$$E[Y\mathbf{1}_A] = E[X\mathbf{1}_A] \leq E\left[X'\mathbf{1}_A\right] = E\left[Y'\mathbf{1}_A\right].$$

On the other hand $Y\mathbf{1}_A \geq Y'\mathbf{1}_A$ by definition of A and so, we may conclude that $Y\mathbf{1}_A = Y'\mathbf{1}_A$ P-a.s. But this implies that $P[A] = 0$, and so $Y \leq Y'$. $\square$

The following example shows that the measure theoretic conditional expectation is indeed a generalisation of the elementary conditional expectation from Definition 1.1.

Example 1.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A_1, \dots, A_N$ an $\mathcal{F}$-measurable partition of Ω with $P[A_n] > 0$ for each $n \in \{1, \dots, N\}$. Set

$$\mathcal{G} := \sigma(\{A_1, \dots, A_N\}).$$

It is not difficult to check that $A \in \mathcal{G}$ if and only if there are indices $n_1, \dots, n_K \in \{1, \dots, N\}$ such that $A = \bigcup_{k=1}^K A_{n_k}$. Moreover, one can show that a random variable $Y : \Omega \to \mathbb{R}$ is $\mathcal{G}$-measurable if and only if it can be written as

$$Y = \sum_{n=1}^N y_n \mathbf{1}_{A_n}, \tag{1.2}$$

for $y_1, \dots, y_N \in \mathbb{R}$, i.e., Y is constant on each A_n.

If $X : \Omega \to \mathbb{R}$ is an integrable $\mathcal{F}$-measurable random variable, then

$$Y := \sum_{n=1}^N E\left[X \mid A_n\right] \mathbf{1}_{A_n} \tag{1.3}$$

is (a version of) the conditional expectation of X given $\mathcal{G}$. Indeed, it follows from (1.2) that Y satisfies the measurability property (1) of a conditional expectation. To check that Y also satisfies the averaging property (2), let $A \in \mathcal{G}$. Then there are indices $n_1, \dots, n_K \in \{1, \dots, N\}$ such that $A = \bigcup_{k=1}^K A_{n_k}$. Note that by definition of Y and Definition 1.1 of the elementary conditional expectation, for each $k \in \{1, \dots, K\}$,

$$E\left[Y \mathbf{1}_{A_{n_k}}\right] = E\left[E\left[X \mid A_{n_k}\right] \mathbf{1}_{A_{n_k}}\right] = E\left[X \mid A_{n_k}\right] P\left[A_{n_k}\right] = E\left[X \mathbf{1}_{A_{n_k}}\right].$$

This and linearity of the expectation give

$$E\left[Y \mathbf{1}_A\right] = \sum_{k=1}^K E\left[Y \mathbf{1}_{A_{n_k}}\right] = \sum_{k=1}^K E\left[X \mathbf{1}_{A_{n_k}}\right] = E\left[X \mathbf{1}_A\right].$$

This ends the example.

Exercise 1.6. ★☆☆ Let $\Omega := \{1, 2, \dots, 6\}$, $\mathcal{F} := 2^\Omega$ and P be the discrete uniform distribution on Ω. Let $A_1 := \{1, 2, 3\}$, $A_2 = \{4, 5, 6\}$ and set $\mathcal{G} := \sigma(\{A_1, A_2\})$. Moreover define $X : \Omega \to \mathbb{R}$ by $X(\omega) := \omega$. Calculate $\mathcal{G}$ and $E\left[X \mid \mathcal{G}\right]$.

Our next result considers the special case that the sub-σ-algebra $\mathcal{G}$ is generated by a random variable (cf. DefinitionA.10(b)). It relies on the *factorisation lemma* for measurable maps; for a proof see [9, Corollary 1.97].

Lemma 1.7. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X a random variable with $X \geq 0$ or $X \in L^1(P)$. Let $(\Omega', \mathcal{F}')$ be another measurable space and $Z : \Omega \to \Omega'$ an $\mathcal{F} - \mathcal{F}'$-measurable map. Define*

$$E[X \mid Z] := E[X \mid \sigma(Z)].$$

Then there exists a $\mathcal{F}' - \mathcal{B}_{\bar{\mathbb{R}}}$-measurable function $h : \Omega' \to \bar{\mathbb{R}}$ such that

$$E[X \mid Z] = h(Z) \;\; P\text{-a.s.}$$

In the situation of Lemma 1.7, we sometimes write $E[X \mid Z = z]$ for $h(z)$. This is a slight abuse of notation. Correct would be

$$h(z) = E[X \mid Z]\Big|_{Z=z}.$$

Exercise 1.8. ★☆☆ Let X be an exponentially distributed random variable with rate parameter $\lambda = 1$ on some probability space $(\Omega, \mathcal{F}, P)$. Compute $E\left[X^2 \,\middle|\, X = 5\right]$.

1.3 Properties of conditional expectations

In this section, we study some key properties of conditional expectation.

Theorem 1.9. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra, and X a random variable with $X \geq 0$ or $X \in \mathcal{L}^1(P)$. Then we have the following properties:*

(a) *If $\mathcal{G} = \{\emptyset, \Omega\}$, then $E[X \mid \mathcal{G}] = E[X]$.*

(b) $E[E[X \mid \mathcal{G}]] = E[X]$.

(c) *If X is $\mathcal{G}$-measurable, then $E[X \mid \mathcal{G}] = X$ P-a.s.*

(d) *If X_1 and X_2 are integrable random variables and $a, b \in \mathbb{R}$, then*

$$E[aX_1 + bX_2 \mid \mathcal{G}] = aE[X_1 \mid \mathcal{G}] + bE[X_2 \mid \mathcal{G}] \quad P\text{-a.s.}$$

(e) *If X_1 and X_2 are integrable random variables with $X_1 \geq X_2$ P-a.s., then*

$$E[X_1 \mid \mathcal{G}] \geq E[X_2 \mid \mathcal{G}] \quad P\text{-a.s.}$$

If in addition $P[X_1 > X_2] > 0$, then $P[E[X_1 \mid \mathcal{G}] > E[X_2 \mid \mathcal{G}]] > 0$.

(f) *If $\mathcal{H} \subset \mathcal{G}$ is a sub-σ-algebra of $\mathcal{G}$, then*

$$E[E[X \mid \mathcal{G}] \mid \mathcal{H}] = E[X \mid \mathcal{H}] \quad P\text{-a.s.}$$

(g) *If Z is $\mathcal{G}$-measurable and ZX is integrable, then*

$$E[ZX \mid \mathcal{G}] = ZE[X \mid \mathcal{G}] \quad P\text{-a.s.}$$

(h) *If X is independent of $\mathcal{G}$, then*

$$E[X \mid \mathcal{G}] = E[X] \quad P\text{-a.s.}$$

(i) *If $f : \mathbb{R} \to \mathbb{R}$ is convex and $f(X) \geq 0$ or $E[|f(X)|] < \infty$,*

$$E[f(X) \mid \mathcal{G}] \geq f(E[X \mid \mathcal{G}]) \quad P\text{-a.s.}$$

(j) *If $X_1, X_2, \ldots$ and X are $[0, \infty]$-valued random variables with $X_1 \leq X_2 \leq \cdots$ P-a.s. and $\lim_{n\to\infty} X_n = X$ P-a.s., then*

$$\lim_{n\to\infty} E[X_n \mid \mathcal{G}] = E[X \mid \mathcal{G}] \quad P\text{-a.s.}$$

(k) *If $X_1, X_2, \ldots$ are $[0, \infty]$-valued random variables, then*

$$E\left[\liminf_{n\to\infty} X_n \,\middle|\, \mathcal{G}\right] \leq \liminf_{n\to\infty} E[X_n \mid \mathcal{G}] \quad P\text{-a.s.}$$

(l) *If $X_1, X_2, \ldots$ and X are random variables with $\lim_{n\to\infty} X_n = X$ P-a.s. and there is $Y \in \mathcal{L}^1(P)$ with $|X_n| \leq Y$ P-a.s. for each $n \in \mathbb{N}$, then $X \in \mathcal{L}^1(P)$ and*[2]

$$\lim_{n\to\infty} E[X_n \mid \mathcal{G}] = E[X \mid \mathcal{G}] \quad P\text{-a.s.}$$

In the above theorem, property (d) is referred to as *linearity* of conditional expectations, property (e) is referred to as *monotonicity* of conditional expectations, property (f) is referred to as *tower property* of conditional expectations, property (g) is referred to as *pull-out property* of conditional expectations, property (h) is referred to as *independence property* of conditional expectations, property (i) is referred to as *Jensen's inequality* for conditional expectations, property (j) is referred to as the *conditional monotone convergence theorem*, property (k) is referred to as the *conditional lemma of Fatou*, and property (l) is referred to as the *conditional dominated convergence theorem.*

Proof of Theorem 1.9. We only establish parts (a), (e), (f), (h), (j), and (l). The other parts are left as an exercise.

[2] Note that the almost sure convergence is wrong if we only assume that the $(X_n)_{n\in\mathbb{N}}$ are uniformly integrable.

(a). Since $E[X]$ is a constant, it is $\mathcal{G}$-measurable, and so the measurability property is satisfied. To check the averaging property, let $A \in \mathcal{G}$. If $A = \emptyset$, then

$$E[E[X]\mathbf{1}_A] = E[X\mathbf{1}_A],$$

and if $A = \Omega$, then

$$E[E[X]\mathbf{1}_A] = E[E[X]] = E[X] = E[X\mathbf{1}_A].$$

(e) We have established the weak inequality already in the proof of Theorem 1.4. To establish the additional claim, assume that $P[X_1 > X_2]$, which together with $X_1 \geq X_2$ P-a.s. implies that $E[X_1] > E[X_2]$. Seeking a contradiction, suppose that $E[X_1 \,|\, \mathcal{G}] = E[X_2 \,|\, \mathcal{G}]$ P-a.s. Then by the averaging property with $A = \Omega \in \mathcal{G}$, we obtain

$$E[X_1] = E[X_1\mathbf{1}_A] = E[E[X_1 \,|\, \mathcal{G}]\mathbf{1}_A] = E[E[X_2 \,|\, \mathcal{G}]\mathbf{1}_A] = E[X_2\mathbf{1}_A] = E[X_2].$$

But this is a contradiction to $E[X_1] > E[X_2]$.

(f) $E[X \,|\, \mathcal{H}]$ is trivially $\mathcal{H}$-measurable, and so the measurability property is satisfied. To check the averaging property, let $A \in \mathcal{H}$. Then $A \in \mathcal{G}$. By applying the the averaging property of conditional expectation first for $\mathcal{H}$ and then for $\mathcal{G}$, we obtain

$$E[E[X \,|\, \mathcal{H}]\mathbf{1}_A] = E[X\mathbf{1}_A] = E[E[X \,|\, \mathcal{G}]\mathbf{1}_A].$$

(h) Since $E[X]$ is a constant, it is trivially $\mathcal{G}$-measurable, and so the measurability property is satisfied. To the check the averaging property, let $A \in \mathcal{G}$. By independence of $\mathcal{G}$ and X it follows that $\mathbf{1}_A$ and X are independent. Hence, we have

$$E[E[X]\mathbf{1}_A] = E[X]E[\mathbf{1}_A] = E[X\mathbf{1}_A].$$

(j) For convenience set $Y_n := E[X_n \,|\, \mathcal{G}]$ for $n \in \mathbb{N}$. Then $0 \leq Y_1 \leq Y_2 \leq \cdots$ P-a.s. by part (e). Set $Y := \liminf_{n\to\infty} Y_n$. Then Y is $\mathcal{G}$-measurable and $\lim_{n\to\infty} Y_n = Y$ P-a.s. It suffices to show that $Y = E[X \,|\, \mathcal{G}]$ P-a.s. We only need to check the averaging property. So let $A \in \mathcal{G}$. Since $X_1\mathbf{1}_A \leq X_2\mathbf{1}_A \leq \cdots$ P-a.s. and $\lim_{n\to\infty} X_n\mathbf{1}_A = X\mathbf{1}_A$ P-a.s. as well as $Y_1\mathbf{1}_A \leq Y_2\mathbf{1}_A \leq \cdots$ P-a.s. and $\lim_{n\to\infty} Y_n\mathbf{1}_A = Y\mathbf{1}_A$ P-a.s., we obtain by the averaging property for the X_n and the monotone convergence theorem,

$$E[Y\mathbf{1}_A] = \lim_{n\to\infty} E[Y_n\mathbf{1}_A] = \lim_{n\to\infty} E[X_n\mathbf{1}_A] = E[X\mathbf{1}_A].$$

(l) For $n \in \mathbb{N}$, set $Z_n := \sup_{k\geq n} |X_k - X|$. Then $0 \leq Z_n \leq 2Y$ P-a.s. for $n \in \mathbb{N}$ and $(Z_n)_{n\in\mathbb{N}}$ is nonincreasing with $\lim_{n\to\infty} Z_n = 0$ P-a.s. Moreover, the dominated convergence theorem gives

$$\lim_{n\to\infty} E[Z_n] = 0 \tag{1.4}$$

Set $\tilde{Y} := E[Y \mid \mathcal{G}]$, $\tilde{Z}_n := E[Z_n \mid \mathcal{G}]$ for $n \in \mathbb{N}$, and $\tilde{Z} := \liminf_{n\to\infty} Z_n$. By monotonicity of conditional expectations, we have $0 \leq \tilde{Z}_n \leq 2\tilde{Y}$ P-a.s. for each $n \in \mathbb{N}$ and $(\tilde{Z}_n)_{n\in\mathbb{N}}$ is P-a.s. nonincreasing sequence with $\lim_{n\to\infty} \tilde{Z}_n = \tilde{Z}$ P-a.s. Fatou's lemma together with the part (b) and (1.4) gives

$$E\left[\tilde{Z}\right] = E\left[\liminf_{n\to\infty} \tilde{Z}_n\right] \leq \liminf_{n\to\infty} E\left[\tilde{Z}_n\right] = \liminf_{n\to\infty} E\left[\tilde{Z}_n\right] = 0$$

As $\tilde{Z} \geq 0$, it follows that $\tilde{Z} = 0$ P-a.s. This implies that $\lim_{n\to\infty} \tilde{Z}_n = 0$ P-a.s. Finally, by linearity of conditional expectations, Jensen's inequality for conditional expectations (for the convex function $x \mapsto |x|$), and the definition of Z_n, we have for each $n \in \mathbb{N}$,

$$|E[X_n \mid \mathcal{G}] - E[X \mid \mathcal{G}]| = |E[X_n - X \mid \mathcal{G}]| \leq E[|X_n - X| \mid \mathcal{G}] \leq E[Z_n \mid \mathcal{G}] = \tilde{Z}_n \ \ P\text{-a.s.}$$

Combining this with the above yields the result. □

Exercise 1.10. ★☆☆ Let $(\Omega, \mathcal{F}, P)$ be a probability space and X_1, X_2, X_3 independent and integrable random variables. Set $\mathcal{G} := \sigma(X_1, X_2)$. Calculate $E[X_1 X_3 \mid \mathcal{G}]$. Explain which properties of conditional expectations you are using.

2 Martingale Theory

It is not exaggerated to say that *martingales* are the heart of Mathematical Finance. In this chapter, we study the discrete time theory of martingales up to the celebrated martingale convergence theorems.

2.1 Stochastic processes and filtrations

We being our discussion by introducing the key notions of a *stochastic process* and a *filtration*.

Definition 2.1. Let $(\Omega, \mathcal{F})$ be a measurable space, $\mathcal{I} \subset [0, \infty)$ a (nonempty) index set, and $(S, \mathcal{S})$ another measurable space. A family $X = (X_t)_{t \in \mathcal{I}}$ of S-valued $\mathcal{F}$-$\mathcal{S}$-measurable maps is called a *stochastic process* on $(\Omega, \mathcal{F})$ with *index set* (or time set) $\mathcal{I}$ and *state space* $(S, \mathcal{S})$. If $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, X is called a *real-valued* stochastic process.

Typical examples for $\mathcal{I}$ are $\{0, \dots, T\}$ ("finite discrete time"), $\mathbb{N}_0$ ("infinite discrete time"), $[0, T]$ ("finite continuous time"), and $\mathbb{R}_+$ ("infinite continuous time").

The basic definition of a stochastic process does not say anything about the *flow of information.* To model the latter, we assume that the *information* available at time t is described by a σ-algebra $\mathcal{F}_t$. As information increases over time, it is naturally to assume that $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leq t$.

Definition 2.2. Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{I} \subset \mathbb{R}_+$ an index set. A family of σ-algebras $\mathbb{F} := (\mathcal{F}_t)_{t \in \mathcal{I}}$ with $\mathcal{F}_t \subset \mathcal{F}$ for all $t \in \mathcal{I}$ is called a *filtration* on $(\Omega, \mathcal{F})$ if

$$\mathcal{F}_s \subset \mathcal{F}_t \quad \text{for all } s, t \in \mathcal{I} \text{ with } s \leq t.$$

In this case ,the triple $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}})$ is called a *filtered measurable space.* Moreover, if P is a probability measure on $(\Omega, \mathcal{F})$, then the quadruple $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}}, P)$ is called a *filtered probability space.*

Definition 2.3. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}})$ be a filtered measurable space. A stochastic process $X = (X_t)_{t \in \mathcal{I}}$ on $(\Omega, \mathcal{F})$ is said to be *adapted to the filtration* $\mathbb{F}$ (or short $\mathbb{F}$-adapted) if X_t is $\mathcal{F}_t$-measurable for all $t \in \mathcal{I}$.

If there is no danger of confusion, we often omit the qualifier "to the filtration $\mathbb{F}$".

We illustrate the above definitions by an example, that is also key to the *binomial model* in Mathematical Finance.

Example 2.4. Suppose that $I = \{0, \dots, T\}$, where $T \in \mathbb{N}$ is a time horizon. We want to model a process that in each period (i.e., from $t-1$ to t, $t \in \{1, \dots, T\}$) can make two different moves, which we call "up" and "down" (which in most cases means indeed that the process at time t either goes up or down relative to its value at time $t-1$). Set

$$\Omega := \{1,2\}^T = \Big\{\omega = (x_1, \dots, x_T) : x_1, \dots, x_T \in \{1,2\}\Big\},$$

i.e., each $\omega = (x_1, \dots, x_T)$ describes one *path* of going up or down, where $x_i = 1$ denotes going up and $x_i = 2$ denotes going down at time i. Let $\mathcal{F} := 2^\Omega$.

For $t \in \{1, \dots, T\}$ and $x_1, \dots, x_t \in \{1,2\}$, we set

$$A_{(x_1,\dots,x_t)} := \{\omega = (\tilde{x}_1, \dots, \tilde{x}_T) \in \Omega : \tilde{x}_1 = x_1, \dots, \tilde{x}_t = x_t\}.$$

Then $A_{(x_1,\dots,x_t)}$ denotes all states of the world with "path up to time t" given by $(x_1, \dots, x_t)$. As we can observe at t exactly "all paths up to time t", the information available at time t is given by

$$\mathcal{F}_t := \sigma(\mathcal{A}_t), \quad \text{where} \quad \mathcal{A}_t := \{A_{(x_1,\dots,x_t)} : x_1, \dots, x_t \in \{1,2\}\}.$$

Then $\mathbb{F} := (\mathcal{F}_t)_{t \in \mathcal{I}}$ is a filtration. Indeed, let $1 \leq s < t \leq T$. Then

$$A_{(x_1,\dots,x_s)} = \bigcup_{\tilde{x}_{s+1},\dots,\tilde{x}_t \in \{1,2\}} A_{(x_1,\dots,x_s,\tilde{x}_{s+1},\dots,\tilde{x}_t)},$$

which shows that $\mathcal{A}_s \subset \sigma(\mathcal{A}_t)$. This implies that $\mathcal{F}_s = \sigma(\mathcal{A}_s) \subset \sigma(\mathcal{A}_t) = \mathcal{F}_t$.[3]

Since each $\mathcal{F}_t$ is generated by the partition $\mathcal{A}_t$, a map $Y : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-measurable if and only if Y is constant on each $A_{(x_1,\dots,x_t)}$. It follows that a stochastic process $X = (X_t)_{t \in \mathcal{I}}$ on $(\Omega, \mathcal{F})$ is $\mathbb{F}$-adapted if and only if

$$X_t((x_1, \dots, x_T)) = X_t((\tilde{x}_1, \dots, \tilde{x}_T)) \quad \text{if } x_1 = \tilde{x}_1, \dots, x_t = \tilde{x}_t,$$

i.e., each X_t can "distinguish between paths" only up to time t.

Exercise 2.5. ★☆☆ Consider the setting of Example 2.4. Let $T = 3$. Explicitly write down the σ-algebras $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$. Also give an example of an $\mathbb{F}$-adapted process $X = (X_t)_{t \in \{0,\dots,3\}}$.

2.2 Martingales, submartingales, and supermartingales

If a stochastic process $X = (X_t)_{t \in \mathcal{I}}$ is adapted, then at time s, we are given the information $\mathcal{F}_s$ and so can fully observe X_r for all $r \leq s$. By contrast, we may not be able to fully observe X_t for $t > s$. The special case that X_s gives the best prognosis for X_t, i.e., $X_s = E[X_t \,|\, \mathcal{F}_s]$ P-a.s. leads to the concept of a *martingale*.

Definition 2.6. A real-valued stochastic process $M = (M_t)_{t \in \mathcal{I}}$ on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}}, P)$ is called a $(P, \mathbb{F})$-*martingale* if

[3] Note that trivially $\mathcal{F}_0 = \{\emptyset, \Omega\} \subset \mathcal{F}_t$ for each $t \in \{1, \dots, T\}$.

(1) M is adapted to $\mathbb{F}$;

(2) M is P-integrable, i.e., each M_t is P-integrable;

(3) $E[M_t \,|\, \mathcal{F}_s] = M_s$ P-a.s. for all $s \le t \in \mathcal{I}$.

If "$=$" in (3) is replaced by "$\ge$", M is called a $(P, \mathbb{F})$-*submartingale*, and if it is replaced by "$\le$", M is called a $(P, \mathbb{F})$-*supermartingale*.

If there is no danger of confusion we often omit the qualifier $(P, \mathbb{F})$.

Remark 2.7. (a) In Definition 2.6, property (1) is referred to as *adaptedness*, property (2) is referred to as *integrability*, and property (3), which is the crucial property, is referred to as *martingale property* (or submartingale/supermartingale property, respectively).

(b) In discrete time, i.e., if $\mathcal{I} = \mathbb{N}_0$ or $\mathcal{I} = \{0, \dots, T\}$, the martingale property (3) in Definition 2.6 is equivalent to the formally weaker *one-step martingale property*

(3') $E[M_t \,|\, \mathcal{F}_{t-1}] = M_{t-1}$ P-a.s. for all $t \in \mathcal{I} \setminus \{0\}$.

Indeed, if the martingale property (3) in Definition 2.6 is satisfied, then also the one-step martingale property (3') is satisfied (pick $s = t - 1$). Conversely, if (3') is satisfied, let $0 \le s \le t \in \mathcal{I}$. If $s = t$, then $E[M_t \,|\, \mathcal{F}_s] = E[M_t \,|\, \mathcal{F}_t] = M_t = M_s$ P-a.s. by adaptedness of M and the pull-out property of conditional expectations. Otherwise, there is $n \in \mathbb{N}$ such that $s = t - n$. The tower property of conditional expectations and (3') give

$$\begin{aligned}
E[M_t \,|\, \mathcal{F}_s] &= E[M_t \,|\, \mathcal{F}_{t-n}] = E[E[M_t \,|\, \mathcal{F}_{t-1}] \,|\, \mathcal{F}_{t-n}] = E[M_{t-1} \,|\, \mathcal{F}_{t-n}] \\
&= E[E[M_{t-1} \,|\, \mathcal{F}_{t-2}] \,|\, \mathcal{F}_{t-n}] = E[M_{t-2} \,|\, \mathcal{F}_{t-n}] = \cdots \\
&= E\left[M_{t-(n-1)} \,\middle|\, \mathcal{F}_{t-n}\right] = M_{t-n} = M_s \quad P\text{-a.s.}
\end{aligned}$$

(c) It follows from property (3) in Definition 2.6 and the tower property that $t \mapsto E[M_t]$ is constant. Similarly, $t \mapsto E[M_t]$ is nondecreasing if M is a submartingale and nonincreasing if M is a supermartingale.

We give two important examples of martingales.

Example 2.8. (a) Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{I} := \mathbb{N}_0$, and $X_1, X_2, \dots$ independent $\mathcal{F}$-measurable integrable random variables with mean 0. Set[4]

$$\mathcal{F}_t := \sigma(X_1, \dots, X_t) := \sigma\left(\bigcup_{n=1}^{t} \sigma(X_n)\right), \quad t \in \mathbb{N}_0,$$

[4] Here $\mathcal{F}_0$, the σ-algebra generated by the empty set, is the trivial σ-algebra $\{\emptyset, \Omega\}$.

i.e., $\mathcal{F}_t$ is the smallest σ-algebra for which $X_1, \ldots, X_t$ are measurable. Define the process $M = (M_t)_{t \in \mathbb{N}_0}$ by[5]

$$M_t := \sum_{i=1}^{t} X_i.$$

Then M is a martingale. Indeed, fix $t \in \mathbb{N}_0$. Then M_t is the sum of $\mathcal{F}_t$-measurable and integrable random variables and hence again $\mathcal{F}_t$-measurable and integrable. This gives adaptedness and integrability of M. We proceed to check the (one-step) martingale property (3'). So let $t \in \mathbb{N}$. By linearity, the pull-out property, and the independence property of conditional expectations (using that X_t is independent of $\mathcal{F}_{t-1}$ by Theorem A.37(b)) and the fact that $E[X_t] = 0$, we obtain

$$\begin{aligned} E[M_t \,|\, \mathcal{F}_{t-1}] &= E[M_{t-1} + X_t \,|\, \mathcal{F}_{t-1}] = E[M_{t-1} \,|\, \mathcal{F}_{t-1}] + E[X_t \,|\, \mathcal{F}_{t-1}] \\ &= M_{t-1} + E[X_t] = M_{t-1} \quad P\text{-a.s.} \end{aligned}$$

So M is a martingale.

(b) Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}}, P)$ be a filtered probability space, where $\mathcal{I} \subset \mathbb{R}_+$ is an arbitrary nonempty index set and Z an $\mathcal{F}$-measurable integrable random variable. Then the process $M = (M_t)_{t \in \mathcal{I}}$ defined by

$$M_t := E[Z \,|\, \mathcal{F}_t], \quad t \in \mathcal{I},$$

is a martingale. Indeed, adaptedness and integrability of M follow from the definition of conditional expectations. The martingale property of M follows from the tower property of conditional expectations.

Exercise 2.9. ★☆☆ Let $X_1, X_2, \ldots$ be i.i.d. random variables that are standard normal distributed on some probability space $(\Omega, \mathcal{F}, P)$. Set

$$\mathcal{F}_t := \sigma(X_1, \ldots, X_t) := \sigma\left(\bigcup_{n=1}^{t} \sigma(X_n)\right), \quad t \in \mathbb{N}_0,$$

Define the process $M = (M_t)_{t \in \mathbb{N}_0}$ by $M_t = \sum_{i=1}^{t} X_i^2$. Is M a martingale, sub- or supermartingale? Carefully check properties (1) – (3) in Definition 2.6.

The following result contains some elementary properties of (sub-/super-)martingales. Its proof is left as an exercise.

Lemma 2.10. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}}, P)$ be a filtered probability space and $X = (X_t)_{t \in \mathcal{I}}$ and $Y = (Y_t)_{t \in \mathcal{I}}$ stochastic processes.*

(a) *X is a supermartingale if and only if $-X$ is a submartingale.*

[5] By convention $\sum_{i=1}^{0} X_i := 0$.

(b) *If X and Y are martingales and $a, b \in \mathbb{R}$, then $aX + bY = (aX_t + bY_t)_{t \in \mathcal{I}}$ is again a martingale.*

(c) *If X and Y are submartingales and $a, b \geq 0$, then $aX + bY = (aX_t + bY_t)_{t \in \mathcal{I}}$ is again a submartingale.*

(d) *If X and Y are supermartingales and $a, b \geq 0$, then $aX + bY = (aX_t + bY_t)_{t \in \mathcal{I}}$ is again a supermartingale.*

We proceed to show that a convex function of a martingale is a *submartingale* – provided that it is integrable.

Theorem 2.11. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function and $M = (M_t)_{t \in \mathcal{I}}$ a martingale on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}}, P)$. For $t \in \mathcal{I}$, set $X_t := \varphi(M_t)$ and assume that $E[|\varphi(M_t)|] < \infty$. Then $X = (X_t)_{t \in \mathcal{I}}$ is a submartingale.*

Proof. Adaptedness follows from the fact that M is adapted and φ is measurable. Integrability follows by definition. To establish the submartingale property, let $0 \leq s < t \in \mathcal{I}$. Then by the conditional version of Jensen's inequality (Theorem 1.9(i)) and the martingale property for M we have

$$X_s = \varphi(M_s) = \varphi(E[M_t \mid \mathcal{F}_s]) \leq E[\varphi(M_t) \mid \mathcal{F}_s] = E[X_t \mid \mathcal{F}_s] \quad P\text{-a.s.} \qquad \square$$

2.3 Discrete stochastic integral

In this section, we study stochastic integrals in discrete time. To motivate this concept, suppose that the stochastic process $X = (X_t)_{t \in \mathbb{N}_0}$ describes the (discounted) value of a stock over time, i.e., X_0 describes the stock price today, X_1 describes the (discounted) value of the stock tomorrow, etc. Suppose we want to invest into this stock. If we describe our holdings from time $t-1$ to t by H_t, then our (discounted) gains/losses from trading during time $t-1$ to t is

$$H_t X_t - H_t X_{t-1} = H_t(X_t - X_{t-1}).$$

If we denote by $G_t(H)$ our cumulative (discounted) gains/losses from trading the strategy $H = (H_t)_{t \in \mathbb{N}}$ up to time t, then

$$G_t(H) := \sum_{k=1}^{t} H_k(X_k - X_{k-1}), \quad t \in \mathbb{N}_0.$$

Consider now the stochastic process $H = (H_t)_{t \in \mathbb{N}_0}$, and think about its measurability properties. Since we have to decide on H_t (our holding from time $t-1$ to t) already at time $t-1$ and we cannot look into the future, we have to require that H_t is $\mathcal{F}_{t-1}$-measurable.

The above discussion motivates the following definitions.

Definition 2.12. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0})$ be a filtered measurable space. A stochastic process $H = (H_t)_{t\in\mathbb{N}_0}$ on $(\Omega, \mathcal{F})$ is said to be *predictable* for the filtration $\mathbb{F}$ (or short $\mathbb{F}$-predictable) if H_0 is a constant and H_t is $\mathcal{F}_{t-1}$-measurable for all $t \in \mathbb{N}$.

If there is no danger of confusion, we often omit the qualifier "to the filtration $\mathbb{F}$".

Definition 2.13. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0})$ be a filtered measurable space, $X = (X_t)_{t\in\mathbb{N}_0}$ a real-valued $\mathbb{F}$-adapted process and $H = (H_t)_{t\in\mathbb{N}_0}$ a real-valued $\mathbb{F}$-predictable process. Then the real-valued process $H \bullet X = ((H \bullet X)_t)_{t\in\mathbb{N}}$ defined by

$$(H \bullet X)_t := \sum_{k=1}^{t} H_k(X_k - X_{k-1})$$

is called the *discrete stochastic integral* of H with respect to X.

In slight abuse of notation, we often write $H \bullet X_t$ for $(H \bullet X)_t$.

Remark 2.14. For the notion of predictability, time 0 plays a special role, since there is no σ-algebra $\mathcal{F}_{-1}$. Also note that in the definition of the discrete stochastic integral, the value of H_0 is irrelevant. There are two ways to deal with this issues: One can either require predictable processes to start at time 1 or require that H_0 is constant and note that it does not play a role. Both approaches have their advantages and disadvantages. We have chosen the second option for consistency with continuous time, where the first option is no longer available.

It is not difficult to check that the stochastic integral is again an adapted process. The following result shows among others that if X is a martingale and H is bounded, then the stochastic integral $H \bullet X$ is again a martingale.

Theorem 2.15. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space, $M = (M_t)_{t\in\mathbb{N}_0}$ a real-valued adapted process and $H = (H_t)_{t\in\mathbb{N}_0}$ a real-valued predictable process that is P-a.s. bounded, i.e., there is $C > 0$ such that $|H_t| \leq C$ P-a.s. for all $t \in \mathbb{N}_0$.*

(a) *If M is a martingale, then the stochastic integral $H \bullet M$ is again a martingale.*

(b) *If M is a submartingale and H is nonnegative, then $H \bullet M$ is a again a submartingale.*

(c) *If M is a supermartingale and H is nonnegative, then $H \bullet M$ is a again a supermartingale.*

Proof. We only establish (a), the proofs of (b) and (c) are similar and left as an exercise.

First, we check adaptedness of $H \bullet M$. So fix $t \in \mathbb{N}_0$. If $t = 0$, then $H \bullet M_0 = 0$, which is $\mathcal{F}_0$-measurable. If $t \geq 1$, then $H \bullet M_t = \sum_{k=1}^{t} H_k(M_k - M_{k-1})$ is $\mathcal{F}_t$-measurable, because each H_k, M_k, M_{k-1} for $k \in \{1, \ldots, t\}$ is $\mathcal{F}_t$-measurable by adaptedness of M and predictability of H.

Next, we check that $H \bullet M$ is integrable. Since H is P-a.s.-bounded, there is a constant $C > 0$ such that $|H_t| \leq C$ P-a.s. for all $t \in \mathbb{N}_0$. Fix $t \in \mathbb{N}_0$. If $t = 0$, then $H \bullet M_0 = 0$ is trivially P-integrable. If $t \geq 1$, then by the triangle inequality, we obtain

$$|H \bullet M_t| = \left| \sum_{k=1}^{t} H_k(M_k - M_{k-1}) \right| \leq \sum_{k=1}^{t} |H_k||M_k - M_{k-1}| \leq C \left(\sum_{k=1}^{t} |M_k| + |M_{k-1}| \right)$$

Since each M_k is integrable, so is the finite sum $\sum_{k=1}^{t} |M_k| + |M_{k-1}|$. It follows that $|H \bullet M_t|$ is integrable.

Finally, to check the martingale property, fix $t \geq 1$. Using that $H \bullet M_{t-1}$ and H_t are $\mathcal{F}_{t-1}$-measurable an M is an $\mathbb{F}$-martingale, we obtain

$$\begin{aligned} E\left[H \bullet M_t \,|\, \mathcal{F}_{t-1}\right] &= E\left[H \bullet M_{t-1} + H_t(M_t - M_{t-1}) \,|\, \mathcal{F}_{t-1}\right] \\ &= H \bullet M_{t-1} + H_t E\left[M_t - M_{t-1} \,|\, \mathcal{F}_{t-1}\right] \\ &= H \bullet M_{t-1} + 0 = H \bullet M_{t-1} \;\; P\text{-a.s.} \end{aligned}$$

□

Exercise 2.16. ★☆☆ Prove part (b) of Theorem 2.15.

2.4 Stopping times

We proceed to define the key notion of a *stopping time* which formalises the concept of a time depending on some random outcome.

Definition 2.17. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathcal{I}})$ be a filtered measurable space. A map $\tau : \Omega \to \mathcal{I} \cup \{\infty\}$ is called a *stopping time for the filtration* $\mathbb{F}$ (or short an $\mathbb{F}$-stopping time) if

$$\{\tau \leq t\} \in \mathcal{F}_t, \quad \text{for all } t \in \mathcal{I}.$$

If there is no danger of confusion, we often omit the qualifier "for the filtration $\mathbb{F}$".

Remark 2.18. In discrete time, i.e., if $\mathcal{I} = \mathbb{N}_0$ or $\mathcal{I} = \{0, \ldots, T\}$, there is an alternative (and sometimes easier) criterion to decide if a map $\tau : \Omega \to \mathcal{I} \cup \{\infty\}$ is a stopping time: It is equivalent check that

$$\{\tau = t\} \in \mathcal{F}_t, \quad \text{for all } t \in \mathcal{I}. \tag{2.1}$$

Indeed, if τ is a stopping time and $t \geq 1$, then $\{\tau \leq t\} \in \mathcal{F}_t$ and $\{\tau \leq t - 1\} \in \mathcal{F}_{t-1} \subset \mathcal{F}_t$ by the definition of a stopping time. Hence,

$$\{\tau = t\} = \{\tau \leq t\} \cap \{\tau \leq t - 1\}^c \in \mathcal{F}_t.$$

Moreover, $\{\tau = 0\} = \{\tau \leq 0\} \in \mathcal{F}_0$ by the definition of a stopping time.

Conversely if (2.1) is satisfied and $t \in \mathcal{I}$, then

$$\{\tau \le t\} = \bigcup_{k=0}^{t} \{\tau = k\} \in \mathcal{F}_t,$$

and hence τ is a stopping time.

We proceed to give an important example of a stopping time.

Example 2.19. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0})$ be a filtered measurable space and $X = (X_t)_{t\in\mathbb{N}_0}$ an adapted process with values in the state space $(S, \mathcal{S})$. Let $A \in \mathcal{S}$. Then the *first hitting time* of A, defined by

$$\tau_A(\omega) := \inf\{t \in \mathbb{N}_0 : X_t(\omega) \in A\},$$

is a stopping time.[6] Indeed, let $t \in \mathbb{N}_0$. Then using that X is adapted to the filtration $\mathbb{F}$, we obtain

$$\{\tau_A = t\} = \bigcap_{k=0}^{t-1} \{X_k \notin A\} \cap \{X_t \in A\} \in \mathcal{F}_t.$$

Now the claim follows from Remark 2.18.

Remark 2.20. The *last hitting time*, defined by

$$L_A(\omega) := \sup\{t \in \mathbb{N}_0 : X_t(\omega) \in A\},$$

is in general not a stopping time.

Exercise 2.21. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0})$ be a filtered measurable space and $S = (S_t)_{t\in\mathbb{N}_0}$ an real-valued adapted process, where S_t models the price of a stock at day t. Assume that $S_0 = 100$. Which of the following is a stopping time? Give a formal proof.

(a) ✫☆☆ The first time that S is larger than 200, i.e.,

$$\tau_1 := \inf\{t \ge 0 : S_t > 200\}.$$

(b) ★✫☆ The first time that S is larger than the previous maximum, i.e.,

$$\tau_2 := \inf\{t \ge 1 : S_t > \max_{k\in\{0,\dots,t-1\}} S_k\}.$$

(c) ★✫☆ The first time that S reaches the absolute maximum, i.e.,

$$\tau_3 := \inf\{t \ge 0 : S_t = \sup_{k\in\mathbb{N}_0} S_k\}.$$

[6] Here, we use the standard convention that $\inf \emptyset = +\infty$.

The following result shows that minima and maxima of stopping times are again stopping times. Its proof is left as an exercise.

Proposition 2.22. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathcal{I}})$ be a filtered measurable space and σ and τ stopping times. Then the following maps are again stopping times:*

(a) $\sigma \wedge \tau := \min(\sigma, \tau)$.

(b) $\sigma \vee \tau := \max(\sigma, \tau)$.

(c) $\sigma + \tau$ *if in addition* $\mathcal{I} = \mathbb{N}_0$ *or* $\mathcal{I} = \mathbb{R}_+$.

We proceed to assign to each stopping time τ a σ-algebra $\mathcal{F}_\tau$.

Definition 2.23. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathcal{I}})$ be a filtered measurable space and τ an $\mathbb{F}$-stopping time. Set

$$\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty : A \cap \{\tau \le t\} \in \mathcal{F}_t \text{ for all } t \in \mathcal{I}\},$$

where $\mathcal{F}_\infty := \sigma\left(\bigcup_{t\in\mathcal{I}} \mathcal{F}_t\right)$.[7]

Intuitively, $\mathcal{F}_\tau$ contains all events that can be observed at time τ. One can check that $\mathcal{F}_\tau$ is indeed a σ-algebra.

The following result shows among others that $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$ for $\sigma \le \tau$, which is a generalisation of $\mathcal{F}_s \subset \mathcal{F}_t$ for deterministic $s \le t$. Its proof is left as an exercise.

Proposition 2.24. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathcal{I}})$ be a filtered measurable space and σ and τ stopping times. Then*

$$\mathcal{F}_\sigma \cap \{\sigma \le \tau\} \subset \mathcal{F}_\tau.$$

In particular, $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$ *if* $\sigma \le \tau$.

2.5 The stopping theorem

In this section, we establish the stopping theorem for sub-/super-/martingales, which is one of the key results of martingale theory.

We begin our discussion by linking a stopping time to a stochastic process X.

Definition 2.25. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathcal{I}})$ be a filtered measurable space, $X = (X_t)_{t\in\mathcal{I}}$ a stochastic process with values in some state space $(S, \mathcal{S})$, and τ an $\mathbb{F}$-stopping time.

(a) If $\tau < \infty$ or there is an S-valued random variable X_∞ that is $\mathcal{F}_\infty$-$\mathcal{S}$-measurable, the map $X_\tau : \Omega \to \mathcal{S}$, defined by

$$X_\tau(\omega) := X_{\tau(\omega)}(\omega),$$

is called the *value of X at time τ*.

[7] Note that this implies that if $\tau = \infty$, then $\mathcal{F}_\tau = \mathcal{F}_\infty$.

(b) The process $X^\tau := (X^\tau_t)_{t\in\mathcal{I}}$, defined by

$$X^\tau_t := X_{t\wedge\tau}, \quad t \in \mathcal{I},$$

is called the *process X stopped at time τ*.

Exercise 2.26. ★☆☆ Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X_1, X_2, \ldots$ i.i.d. Bernoulli random variables with parameter $p \in (0,1)$. Define the process M by $M_t := \sum_{k=1}^t X_k$ and set $\mathcal{F}_t := \sigma(X_1, \ldots, X_k)$ and $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}$. Show that $\tau := \inf\{t \geq 0 : M_t = 2020\}$ is an $\mathbb{F}$-stopping time and calculate M_τ. (*Hint:* Use the strong law of large numbers to deduce that τ is P-a.s. finite.)

The following result shows that if X is adapted and $\mathcal{I}$ is countable, X_τ is an $\mathcal{F}_\tau$-measurable random variable and the stopped process X^τ is again adapted. Its proof is left as an exercise.

Proposition 2.27. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0})$ be a filtered measurable space, $X = (X_t)_{t\in\mathbb{N}_0}$ an $\mathbb{F}$-adapted process with values in some state space $(S, \mathcal{S})$, and τ an $\mathbb{F}$-stopping time. Then:*

(a) *X_τ is $\mathcal{F}_\tau$-measurable.*[8]

(b) *The stopped process process X^τ is again adapted to the filtration $\mathbb{F}$.*[9]

Remark 2.28. The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous) paths, i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

Exercise 2.29. ★★☆ Prove Proposition 2.27.

The following result is the cornerstone of the stopping theorem.

Lemma 2.30. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0})$ be a filtered measurable space, $X = (X_t)_{t\in\mathbb{N}_0}$ an $\mathbb{F}$-adapted process and τ an $\mathbb{F}$-stopping time. Then the process $H = (H_t)_{t\in\mathbb{N}_0}$ defined by*

$$H_t := \mathbf{1}_{\{t\leq\tau\}}$$

is $\mathbb{F}$-predictable and

$$H \bullet X_t = X_{\tau\wedge t} - X_0, \quad t \in \mathbb{N}_0.$$

Proof. H is predictable because $H_0 = \mathbf{1}_{\{0\leq\tau\}} = 1$ is a constant and for $t \geq 1$, each H_t is $\mathcal{F}_{t-1}$-measurable because $\{t \leq \tau\} = \{\tau \leq t-1\}^c \in \mathcal{F}_{t-1}$. Moreover, the definition of the stochastic integral gives

$$H \bullet X_t = \sum_{k=1}^{t} H_k(X_k - X_{k-1}) = \sum_{k=1}^{t\wedge\tau}(X_k - X_{k-1}) = X_{t\wedge\tau} - X_0. \qquad \square$$

[8] We implicitly assume that either $\tau < \infty$ or there is an S-valued $\mathcal{F}_\infty$-$\mathcal{S}$-measurable random variable X_∞.

[9] It follows from part (a), that X^τ is also adapted to the filtration $\mathbb{F}^\tau := (\mathcal{F}_{t\wedge\tau})_{t\in\mathbb{N}_0}$.

We are now in a position to establish the stopping theorem.

Theorem 2.31 (Stopping theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space, $X = (X_t)_{t\in\mathbb{N}_0}$ a (sub-/super-)martingale, and τ a stopping time. Then*

(a) *The stopped process $X^\tau = (X^\tau_t)_{t\in\mathbb{N}_0}$ is again a (sub-/super-)martingale.*

(b) *If τ is P-a.s. bounded and σ is another stopping time with $\sigma \leq \tau$ P-a.s., then X_τ is P-integrable and*

$$\begin{aligned} E\left[X_\tau \,|\, \mathcal{F}_\sigma\right] &= X_\sigma \;\; P\text{-a.s.} && \text{if } X \text{ is a martingale},\\ E\left[X_\tau \,|\, \mathcal{F}_\sigma\right] &\geq X_\sigma \;\; P\text{-a.s.} && \text{if } X \text{ is a submartingale},\\ E\left[X_\tau \,|\, \mathcal{F}_\sigma\right] &\leq X_\sigma \;\; P\text{-a.s.} && \text{if } X \text{ is a supermartingale}. \end{aligned}$$

Remark 2.32. (a) The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous) paths, i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

(b) Part (b) of Theorem 2.31 is often referred to as *optional sampling.*

Proof of Theorem 2.31. We only consider the case that X is a martingale. The case of sub-/ supermartingales is similar.

(a) It follows from Lemma 2.30 that $X^\tau = H \bullet X + X_0$ for the bouded predictable process $H = (H_t)_{t\in\mathbb{N}_0}$ given by $H_t = \mathbf{1}_{\{t\leq\tau\}}$. The claim follows from Theorem 2.15 and Lemma 2.10(b), noting that the constant process X_0 is trivially a martingale.

(b) Since τ is P-a.s. bounded, there is $N \in \mathbb{N}_0$ such that $\tau \leq N$ P-a.s. Hence $X_\tau = X^\tau_N$ P-a.s., and $X^\tau_N \in \mathcal{L}^1(P)$ by the fact that X^τ is a martingale by part (a). Thus, X_τ is integrable and it suffices to show that

$$E\left[X^\tau_N \,|\, \mathcal{F}_\sigma\right] = X_\sigma \;\; P\text{-a.s.}$$

We use the definition of conditional expectations. The measurability property is trivially satisfied. To check the averaging property, let $A \in \mathcal{F}_\sigma$. Then the definition of $\mathcal{F}_\sigma$ and the fact that $\{\sigma = n\}$ is $\mathcal{F}_n$-measurable for each $n \in \mathbb{N}_0$ give

$$A \cap \{\sigma = n\} = A \cap \{\sigma \leq n\} \cap \{\sigma = n\} \in \mathcal{F}_n, \quad n \in \mathbb{N}_0.$$

Using this, the fact that X^τ is an $\mathbb{F}$-martingale, the fact that $\sigma \leq \tau \leq N$ P-a.s., and the averaging property of conditional expectations, we obtain

$$\begin{aligned} E\left[X_\sigma \mathbf{1}_A\right] = E\left[X^\tau_\sigma \mathbf{1}_A\right] = E\left[\sum_{n=0}^N X^\tau_\sigma \mathbf{1}_A \mathbf{1}_{\{\sigma=n\}}\right] = \sum_{n=0}^N E\left[X^\tau_\sigma \mathbf{1}_A \mathbf{1}_{\{\sigma=n\}}\right] = \sum_{n=0}^N E\left[X^\tau_n \mathbf{1}_A \mathbf{1}_{\{\sigma=n\}}\right] \\ = \sum_{n=0}^N E\left[X^\tau_N \mathbf{1}_A \mathbf{1}_{\{\sigma=n\}}\right] = E\left[\sum_{n=0}^N X^\tau_N \mathbf{1}_A \mathbf{1}_{\{\sigma=n\}}\right] = E\left[X^\tau_N \mathbf{1}_A\right]. \end{aligned}$$

□

We illustrate the stopping theorem by discussing the famous Gambler's ruin problem.

Example 2.33 (Gambler's ruin)**.** Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, P)$ where $P[X_1 = 1] = \frac{1}{2}$ and $P[X_1 = -1] = \frac{1}{2}$. Set $\mathcal{F}_t := \sigma(X_1, \dots, X_t)$ for $t \in \mathbb{N}$. Let $x \in \mathbb{Z}$ and define the process $S = (S_t)_{t\in\mathbb{N}_0}$ by

$$S_t := x + \sum_{i=1}^{t} X_i.$$

Then S is a martingale by Example 2.8(a) and Lemma 2.10(b). It is called the *symmetric random walk* with initial value x. We can interpret S as the evolution of the wealth of a gambler, who starts with initial capital x and in each round tosses a fair coin. For "heads" the gambler gains 1 and for "tails" the gambler loses 1.

Let $a, b \in \mathbb{Z}$ with $a < x < b$. Let $\tau_{a,b}$ be the first hitting time of the set $\{a, b\}$, i.e.,

$$\tau_{a,b} = \inf\{t \in \mathbb{N} : S_t \in \{a, b\}\}.$$

We can interpret a as the point, where the "bank" stops the game, and the gambler is ruined, and b as the point, where the gambler has "earned enough" so that she can stop playing. Set

$$r(x) := P[S_{\tau_{a,b}} = a],$$

i.e., $r(x)$ is the probability that the gambler is "ruined".

We want to find $r(x)$. To this end, we use the stopping theorem. It follows from Theorem 2.31(b) (with $\tau = \tau_{a,b} \wedge n$ and $\sigma = 0$) that

$$E\left[S_{\tau_{a,b}\wedge n}\right] = E\left[S_0\right] = x, \quad n \in \mathbb{N}.$$

Using that $\tau_{a,b} < \infty$ P-a.s. by the law of iterated logarithm (Theorem A.94), we obtain

$$\lim_{n\to\infty} S_{\tau_{a,b}\wedge n} = S_{\tau_{a,b}} \quad P\text{-a.s.}$$

Moreover, using that $|S_{\tau_{a,b}\wedge n}| \leq |a| \vee |b|$ for all $n \in \mathbb{N}$, we get by dominated convergence,

$$E\left[S_{\tau_{a,b}}\right] = \lim_{n\to\infty} E\left[S_{\tau_{a,b}\wedge n}\right] = x, \quad n \in \mathbb{N}.$$

Now using that $S_{\tau_{a,b}} = a$ with probability $r(x)$ and $S_{\tau_{a,b}} = b$ with probability $1 - r(x)$, we obtain

$$r(x)a + (1 - r(x))b = x \quad \Leftrightarrow \quad r(x) = \frac{b - x}{b - a}.$$

2.6 Maximal inequalities

In this section, we prove two fundamental inequalities for the maximum of a martingale.

We begin by the *Doob maximal inequality*, for the (unilateral) maximum process of a submartingale. The proof, which uses the stopping theorem, is left as an exercise.

Lemma 2.34 (Doob's maximal inequality). *Let $X = (X_t)_{t\in\mathbb{N}_0}$ be a submartingale on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$. Then for each $\lambda > 0$.*

$$P\left[\max_{k\in\{0,\dots,n\}} X_k \geq \lambda\right] \leq \frac{1}{\lambda} E\left[X_n \mathbf{1}_{\{\max_{k\in\{0,\dots,n\}} X_k \geq \lambda\}}\right]. \tag{2.2}$$

Remark 2.35. The same result (with max replaced by sup) holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous paths), i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

Next, we define the (bilateral) maximum process of a stochastic process.

Definition 2.36. Let $X = (X_t)_{t\in\mathbb{N}_0}$ be a real-valued stochastic process on some measurable space $(\Omega, \mathcal{F})$. Define the *(bilateral) maximum process* $X^* = (X^*_t)_{t\in\mathbb{N}_0}$ by

$$X^*_t := \max_{s\in\{0,\dots,t\}} |X_s|.$$

Moreover set $X^*_\infty := \sup_{s\in\mathbb{N}_0} |X_s| = \lim_{t\to\infty} X^*_t$.

Exercise 2.37. Let $X = (X_t)_{t\in\mathbb{N}_0}$ be a real-valued stochastic process on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$.

(a) ★✮☆ Show that if X is $\mathbb{F}$-adapted, then so is X^*.

(b) ✮☆☆ Show that if X is P-integrable, then so is X^*.[10]

We now state and prove the famous *Doob's L^p-inequality.*

Theorem 2.38 (Doob's L^p-inequality). *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space and $X = (X_t)_{t\in\mathbb{N}_0}$ a martingale. Then for any $p > 1$ and $t \in \mathbb{N}_0$,*

$$E\left[(X^*_t)^p\right] \leq \left(\frac{p}{p-1}\right)^p E\left[|X_t|^p\right]. \tag{2.3}$$

Remark 2.38. The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous) paths, i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

[10] Warning: This result is false in continuous time.

Proof of Theorem 2.38. Fix $p > 1$ and $t \in \mathbb{N}_0$. We may assume without loss of generality that $P[X_t^* > 0] > 0$.

First, since $|X|$ is a submartingale by Theorem 2.11, it follows from Lemma 2.34 that for any $\lambda \geq 0$,

$$P\left[X_t^* \geq \lambda\right] \leq \frac{1}{\lambda} E\left[|X_t| \mathbf{1}_{\{X_t^* \geq \lambda\}}\right]. \tag{2.4}$$

Fix $K > 0$. Using Fubini's theorem, (2.4) and again Fubini's theorem, we obtain

$$\begin{aligned}
E\left[(X_t^* \wedge K)^p\right] &= E\left[\int_0^{X_t^* \wedge K} p\lambda^{p-1} \,\mathrm{d}\lambda\right] = E\left[\int_0^K p\lambda^{p-1} \mathbf{1}_{\{X_t^* \geq \lambda\}} \,\mathrm{d}\lambda\right] \\
&= \int_0^K p\lambda^{p-1} P[X_t^* \geq \lambda] \,\mathrm{d}\lambda \leq \int_0^K p\lambda^{p-2} E\left[|X_t| \mathbf{1}_{\{X_t^* \geq \lambda\}}\right] \,\mathrm{d}\lambda \\
&= E\left[|X_t| \int_0^K p\lambda^{p-2} \mathbf{1}_{\{X_t^* \geq \lambda\}} \,\mathrm{d}\lambda\right] = E\left[|X_t| \int_0^{X_t^* \wedge K} p\lambda^{p-2} \,\mathrm{d}\lambda\right] \\
&= \frac{p}{p-1} E\left[|X_t| (X_t^* \wedge K)^{p-1}\right].
\end{aligned}$$

Now Hölder's inequality (Theorem A.62) with exponents p and $q = \frac{p}{p-1}$ yields

$$E\left[(X_t^* \wedge K)^p\right] \leq \frac{p}{p-1} E\left[|X_t|^p\right]^{\frac{1}{p}} E\left[(X_t^* \wedge K)^p\right]^{\frac{p-1}{p}}. \tag{2.5}$$

Now raising (2.5) to the p-th power and then dividing by $E\left[(X_t^* \wedge K)^p\right]^{p-1}$ (which is in $(0, K]$ by the fact that $P[X_t^* > 0] > 0$), we obtain

$$E\left[(X_t^* \wedge K)^p\right] \leq \left(\frac{p}{p-1}\right)^p E\left[|X_t|^p\right]. \tag{2.6}$$

Now the result follows from the Monotone Convergence Theorem, by letting $K \to \infty$. $\square$

We note an important corollary. Its proof is left as an exercise.

Corollary 2.39. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{N}_0}, P)$ be a filtered probability space and $X = (X_t)_{t \in \mathbb{N}_0}$ a martingale. Let $p > 1$ and suppose that X is bounded in L^p, i.e., $\sup_{t \in \mathbb{N}_0} E\left[|X_t|^p\right] < \infty$. Then $X_\infty^* \in \mathcal{L}^p$ and*

$$E\left[(X_\infty^*)^p\right] \leq \left(\frac{p}{p-1}\right)^p \sup_{t \geq 0} E\left[|X_t|^p\right]. \tag{2.7}$$

Exercise 2.40. ★☆☆ Prove Corollary 2.39.

2.7 Doob's supermartingale convergence theorem

In this section, we study the limiting behaviour of supermartingales.

At the heart of the (super)-martingale convergence theorems lies the so-called *upcrossing inequality.*

Lemma 2.41 (Upcrossing inequality). *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space and $(X_t)_{t\in\mathbb{N}_0}$ a supermartingale. Fix $a, b \in \mathbb{R}$ with $a < b$. Define the stopping times $(\sigma_k)_{k\in\mathbb{N}}$ and $(\tau_k)_{k\in\mathbb{N}_0}$ recursively by $\tau_0 := 0$ and*

$$\sigma_k(\omega) := \inf\{t \geq \tau_{k-1}(\omega) : X_t(\omega) \leq a\}.$$
$$\tau_k(\omega) := \inf\{t \geq \sigma_k(\omega) := X_t(\omega) \geq b\}.$$

For $t \in \mathbb{N}$, denote by $U_t^{a,b} := \sup\{k \in \mathbb{N}_0 : \tau_k \leq t\}$ the number of upcrossings *of X over $[a, b]$ until time t. Then*

$$E\left[U_t^{a,b}\right] \leq \frac{1}{b-a} E\left[(X_t - a)^-\right].$$

Proof. Define the process $H = (H_t)_{t\in\mathbb{N}_0}$ by

$$H_t := \sum_{k=1}^{\infty} \mathbf{1}_{\{\sigma_k < t \leq \tau_k\}}, \quad t \in \mathbb{N}_0. \tag{2.8}$$

By the fact that $\sigma_k, \tau_k \geq k$ for all $k \in \mathbb{N}$, the sum in (2.8) is always finite. More precisely, for each fixed $t \in \mathbb{N}$,

$$H_s = \sum_{k=1}^{t} \mathbf{1}_{\{\sigma_k < s \leq \tau_k\}} = \sum_{k=1}^{t} \left(\mathbf{1}_{\{s \leq \tau_k\}} - \mathbf{1}_{\{s \leq \sigma_k\}}\right), \quad s \in \{0, \dots, t\}. \tag{2.9}$$

Note that there is at most one k such that $\mathbf{1}_{\{\sigma_k < s \leq \tau_k\}} \neq 0$. Thus, H is bounded by 1. Also, (2.9) together with Lemma 2.30 imply that H is predictable. Moreover, the fact that the stochastic integral is linear together with Lemma 2.30 give

$$\begin{aligned} H \bullet X_t &= \sum_{k=1}^{t} \left(\mathbf{1}_{\{\sigma_k < \cdot \leq \tau_k\}} \bullet X\right)_t = \sum_{k=1}^{t} \left(\left(\mathbf{1}_{\{\cdot \leq \tau_k\}} - \mathbf{1}_{\{\cdot \leq \sigma_k\}}\right) \bullet X\right)_t \\ &= \sum_{k=1}^{t} \left((X_{\tau_k \wedge t} - X_0) - (X_{\sigma_k \wedge t} - X_0)\right) = \sum_{k=1}^{t} \left(X_{\tau_k \wedge t} - X_{\sigma_k \wedge t}\right). \end{aligned}$$

Now using the the definition of $U_t^{a,b}$, we obtain

$$H \bullet X_t = \sum_{k=1}^{t} \left(X_{\tau_k \wedge t} - X_{\sigma_k \wedge t}\right) = \sum_{k=1}^{U_t^{a,b}} \left(X_{\tau_k} - X_{\sigma_k}\right) + \left(X_t - X_{t \wedge \sigma_{U_t^{a,b}+1}}\right).$$

To estimate the last term, we note that by the the definition of $\sigma_{U_t^{a,b}+1}$,

$$X_t - X_{t\wedge\sigma_{U_t^{a,b}+1}} = \begin{cases} X_t - X_t = 0 \geq -(X_t - a)^- & \text{if } \sigma_{U_t^{a,b}+1} > t, \\ X_t - X_{\sigma_{U_t^{a,b}+1}} \geq X_t - a \geq -(X_t - a)^- & \text{if } \sigma_{U_t^{a,b}+1} \leq t. \end{cases}$$

This implies that

$$H \bullet X_t = \sum_{k=1}^{U_t^{a,b}} (X_{\tau_k} - X_{\sigma_k}) + \left(X_t - X_{t \wedge \sigma_{U_t^{a,b}+1}} \right) \geq U_t^{a,b}(b-a) - (X_t - a)^-.$$

Finally, since $H \bullet X$ is a supermartingale with $H \bullet X_0 = 0$ by Theorem 2.15, we have

$$0 \geq E\left[H \bullet X_t\right] \geq (b-a)E\left[U_t^{a,b}\right] - E\left[(X_t - a)^-\right].$$

Rearranging this equation establishes the claim. □

Using the upcrossing inequality, we proceed to establish Doob's supermartingale convergence theorem. To this end, recall that $\mathcal{F}_\infty := \sigma(\bigcup_{t \in \mathbb{N}_0} \mathcal{F}_t)$.

Theorem 2.42 (Doob's supermartingale convergence theorem). *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{N}_0}, P)$ be a filtered probability space and $(X_t)_{t \in \mathbb{N}_0}$ a supermartingale with $\sup_{t \in \mathbb{N}_0} E\left[X_t^-\right] < \infty$. Then $(X_t)_{t \in \mathbb{N}_0}$ converges P-a.s. to a $\mathcal{F}_\infty$-measurable random variable $X_\infty \in \mathcal{L}^1(P)$.*

Remark 2.43. The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous) paths, i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

Proof of Theorem 2.42. First, note that $\sup_{t \in \mathbb{N}} E\left[X_t^-\right] < \infty$ implies that $(X_t)_{t \in \mathbb{N}_0}$ is bounded in L^1. Indeed, using that $|x| = x + 2x^-$ for each $x \in \mathbb{R}$ and X is a supermartingale, we get

$$E\left[|X_t|\right] = E\left[X_t\right] + 2E\left[X_t^-\right] \leq E\left[X_0\right] + 2E\left[X_t^-\right], \quad t \in \mathbb{N}_0.$$

Taking the supremum over $t \in \mathbb{N}_0$ gives

$$\sup_{t \in \mathbb{N}_0} E\left[|X_t|\right] \leq E\left[X_0\right] + 2 \sup_{t \in \mathbb{N}_0} E\left[X_t^-\right] < \infty. \tag{2.10}$$

Next, for $a < b$ and $t \in \mathbb{N}$, define $U_t^{a,b}$ as in Lemma 2.41. Since $(U_t^{a,b})_{t \in \mathbb{N}}$ is nondecreasing, we can set

$$U^{a,b} := \lim_{t \to \infty} U_t^{a,b},$$

which denotes the total numbers of upcrossings of X over $a < b$.

We proceed to show that $U^{a,b} < \infty$ P-a.s. for each $a < b$. So fix $a < b$. It follows from monotone convergence, Lemma 2.41, the fact that $(x-a)^- \leq x^- + a^+$ for all $x, a \in \mathbb{R}$, and the hypothesis that

$$E\left[U^{a,b}\right] = \lim_{t\to\infty} E\left[U_t^{a,b}\right] = \sup_{t\in\mathbb{N}_0} E\left[U_t^{a,b}\right] \le \sup_{t\in\mathbb{N}_0} \frac{E\left[(X_t-a)^-\right]}{b-a}$$
$$\le \frac{1}{b-a}\sup_{t\in\mathbb{N}_0} E\left[(X_t)^-\right] + \frac{a^+}{b-a} < \infty.$$

Thus, $U^{a,b} < \infty$ P-a.s.

Next, note that

$$\left\{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\right\} \subset \bigcup_{\substack{a<b,\\ a,b\in\mathbb{Q}}} \{U^{a,b} = \infty\}.$$

Since the countable union of null sets is again a null set, it follows that

$$\liminf_{t\to\infty} X_t = \limsup_{t\to\infty} X_t \quad P\text{-a.s.}$$

Set $X_\infty := \liminf_{t\to\infty} X_t$. Then $(X_t)_{t\in\mathbb{N}_0}$ converges P-a.s. to X_∞. Moreover, X_∞ is $\mathcal{F}_\infty$-measurable because each X_t is $\mathcal{F}_\infty$-measurable. Finally, by Fatou's lemma and (2.10), we obtain

$$E\left[|X_\infty|\right] = E\left[\lim_{t\to\infty} |X_t|\right] \le \liminf_{t\to\infty} E\left[|X_t|\right] \le \sup_{t\in\mathbb{N}_0} E\left[|X_t|\right] < \infty.$$

Thus $X_\infty \in \mathcal{L}^1(P)$. □

Exercise 2.44. ★☆☆ Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space and $(X_t)_{t\in\mathbb{N}_0}$ a nonnegative martingale. Show that $(X_t)_{t\in\mathbb{N}_0}$ converges almost surely to some nonnegative random variable $X_\infty \in \mathcal{L}^1(P)$.

2.8 L^p-martingale convergence theorem

In this section, we establish convergence theorems in L^p for martingales. We begin by proving the result for $p = 1$.

Theorem 2.45 (L^1-martingale convergence theorem). *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space and $(X_t)_{t\in\mathbb{N}_0}$ a martingale. Then the following are equivalent.*

(1) *$(X_t)_{t\in\mathbb{N}_0}$ is uniformly integrable.*

(2) *$(X_t)_{t\in\mathbb{N}_0}$ converges P-a.s. and in L^1 to an $\mathcal{F}_\infty$-measurable random variable $X_\infty \in \mathcal{L}^1(P)$. Moreover,*

$$X_t = E\left[X_\infty \,|\, \mathcal{F}_t\right] \quad P\text{-a.s.}, \qquad t \in \mathbb{N}_0. \tag{2.11}$$

(3) *There exists an integrable random variable Z such that*

$$X_t = E\left[Z \,|\, \mathcal{F}_t\right] \ \ P\text{-}a.s., \quad t \in \mathbb{N}_0.$$

Remark 2.46. (a) The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous paths), i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

(b) Note that Z in (3) may be different from X_∞. However, with some additional effort one can show that $X_\infty = E\left[Z \,|\, \mathcal{F}_\infty\right]$ P-a.s.

(c) The direction "(1) $\Rightarrow$ (2)" (and its proofs) extends to sub- or supermartingales, with the equality in (2.11) being replaced by an inequality ($\leq$ for sub-, $\geq$ for supermartingales). However, the direction "(3) $\Rightarrow$ (1)" is false for sub- or supermartingales.[11]

Proof of Theorem 2.45. "(1) $\Rightarrow$ (2)": If $(X_t)_{t\in\mathbb{N}_0}$ is uniformly integrable, it is bounded in L^1 by Theorem A.87. Hence X_t converges P-a.s. to an $\mathcal{F}_\infty$-measurable random variable X_∞ by Theorem 2.42. The convergence is also in L^1 by Theorem A.89. Finally, fix $t \in \mathbb{N}_0$. We have to show that

$$E\left[X_\infty \,|\, \mathcal{F}_t\right] = X_t \ \ P\text{-a.s.},$$

Clearly X_t is $\mathcal{F}_t$-measurable. To check the averaging property, let $A \in \mathcal{F}_t$. Since $X_{t+n} \to X_\infty$ in L^1 as $n \to \infty$, we also have $X_{t+n}\mathbf{1}_A \to X_\infty \mathbf{1}_A$ in L^1. Hence, by the martingale property of X, we have

$$E\left[X_\infty \mathbf{1}_A\right] = \lim_{n\to\infty} E\left[X_{t+n}\mathbf{1}_A\right] = \lim_{n\to\infty} E\left[X_t \mathbf{1}_A\right] = E\left[X_t \mathbf{1}_A\right].$$

"(2) $\Rightarrow$ (3)": Set $Z = X_\infty$.

"(3) $\Rightarrow$ (1)": Since Z is integrable, it is uniformly integrable. By Theorem A.87, there exists a nondecreasing convex function $H : [0,\infty) \to [0,\infty)$ with $\lim_{x\to\infty} \frac{H(x)}{x} = \infty$ such that

$$E\left[H(|Z|)\right] < \infty.$$

Fix $t \in \mathbb{N}_0$. Applying the conditional version of Jensen's inequality twice and using that H is nondecreasing, we obtain

$$E\left[H(|X_t|)\right] = E\left[H(|E\left[Z \,|\, \mathcal{F}_t\right]|)\right] \leq E\left[H(E\left[|Z| \,|\, \mathcal{F}_t\right])\right] \leq E\left[E\left[H(|Z|) \,|\, \mathcal{F}_t\right]\right] = E\left[H(Z)\right].$$

Thus,

$$\sup_{t\in\mathbb{N}_0} E\left[H(|X_t|)\right] \leq E\left[H(Z)\right] < \infty.$$

It follows from Theorem A.87 that $(X_t)_{t\in\mathbb{N}_0}$ is UI. $\square$

[11] More precisely, in the case of supermartingales, (3) only implies that X^- is uniformly integrable, and in the case of sub-martingales, (3) only implies that X^+ is uniformly integrable.

The following corollary shows that the stopping theorem for martingales extends to arbitrary (as opposed to bounded) stopping times – provided that the martingale is uniformly integrable. The proof is left as an exercise.

Corollary 2.47. *Let $X = (X_t)_{t\in\mathbb{N}_0}$ be a uniformly integrable (sub-/super-)martingale on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ and τ a stropping time.*

(a) *The stopped process $X^\tau = (X_t^\tau)_{t\in\mathbb{N}}$ is again a uniformly integrable (sub-/super-)martingale.*

(b) *X_τ is P-integrable, and if σ is a stopping time with $\sigma \leq \tau$ P-a.s., then*

$$\begin{aligned} E\left[X_\tau \,|\, \mathcal{F}_\sigma\right] &= X_\sigma \quad P\text{-a.s.} && \text{if } X \text{ is a martingale,}\\ E\left[X_\tau \,|\, \mathcal{F}_\sigma\right] &\geq X_\sigma \quad P\text{-a.s.} && \text{if } X \text{ is a submartingale,}\\ E\left[X_\tau \,|\, \mathcal{F}_\sigma\right] &\leq X_\sigma \quad P\text{-a.s.} && \text{if } X \text{ is a supermartingale.} \end{aligned}$$

Remark 2.48. (a) The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous) paths, i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

Next, we turn to the case $p > 1$.

Theorem 2.49 (L^p-martingale convergence theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space and $(X_t)_{t\in\mathbb{N}_0}$ a martingale. Let $p > 1$ and suppose that X is bounded in L^p, i.e., $\sup_{t\in\mathbb{N}_0} E\left[|X_t|^p\right] < \infty$. Then $(X_t)_{t\in\mathbb{N}_0}$ converges P-a.s. and in L^p to an $\mathcal{F}_\infty$-measurable random variable $X_\infty \in \mathcal{L}^1(P)$. Moreover,*

$$X_t = E\left[X_\infty \,|\, \mathcal{F}_t\right] \quad P\text{-a.s.}, \quad t \in \mathbb{N}_0. \tag{2.12}$$

Remark 2.50. (a) The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous paths), i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

(b) The L^p-convergence is *false* for sub- or supermartingales.

Proof of Theorem 2.49. First, since X is bounded in L^p, it is uniformly integrable by Corollary A.88. Hence, by Theorem 2.45, $(X_t)_{t\in\mathbb{N}_0}$ converges almost surely and in L^1 to some random variable X_∞ in $\mathcal{L}^1(P)$ and (2.12) holds.

Next, since $|X_t|^p \leq (X_\infty^*)^p$, it follows from Corollary 2.39 and the dominated convergence theorem that X_t converges to X_∞ in L^p. □

Exercise 2.51. ★★☆ Give an example of a martingale that is bounded in L^1 but does not converge in L^1. (*Hint:* Let $M_t = \prod_{k=1}^t X_k$, where $X_1, X_2, \ldots$ are i.i.d., nonnegative and have mean 1.)

The following corollary extends Doob's inequality to $t = \infty$.

Corollary 2.52. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{N}_0}, P)$ be a filtered probability space and $X = (X_t)_{t\in\mathbb{N}_0}$ a martingale. Let $p > 1$ and suppose that X is bounded in L^p. Then*

$$E\left[(X_\infty^*)^p\right] \leq \left(\frac{p}{p-1}\right)^p E\left[|X_\infty|^p\right]. \tag{2.13}$$

Remark 2.53. The same result holds true if the index set $\mathbb{N}_0$ is replaced by $\mathbb{R}_+$ and X has in addition continuous (or right-continuous) paths, i.e., for each $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is continuous (or right-continuous).

Exercise 2.54. ★☆☆ Prove Corollary 2.52.

3 Pricing and hedging in discrete time

We now apply the definitions and results of the previous section to model financial markets in finite discrete time. We present the concept of *no-arbitrage* and prove the easy direction of the *fundamental theorem of asset pricing.* We then turn to the *pricing* and *hedging* of *derivative contracts* like call or put options, with a specific emphasis on *complete markets* and the *binomial model.*

3.1 Financial market

We consider a financial market with $1 + d$ assets, which are priced at times $t = 0, 1, \ldots, T$ for some finite *time horizon* $T \in \mathbb{N}$. We work on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\{0,\ldots,T\}}, P)$, where the filtration $\mathbb{F}$ describes the flow of information. To simplify the analysis, we assume without further mentioning throughout this chapter that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$. We model the assets as $\mathbb{F}$-adapted stochastic processes $(S^i_t)_{t\in\{0,\ldots,T\}}$, $i \in \{0, \ldots, d\}$. We assume that S^0 is positive and *predictable*, which means that it is *locally riskless*, i.e., one-period investments in S^0 do not bear risk. As S^0 is positive and predictable, we can write it as

$$S^0_t := \prod_{k=1}^{t}(1 + r_k),$$

where $r_k > -1$ P-a.s. is $\mathcal{F}_{k-1}$-measurable and denotes the *riskles rate* in period k, i.e, from $k - 1$ to k.[12] We also refer to S^0 as *bank account.*[13]

To simplify the notation, we write

$$S_t = (S^1_t, \ldots, S^d_t) \quad \text{and} \quad \overline{S}_t = (S^0_t, S_t), \quad t \in \{0, \ldots, T\},$$

and call the $\mathbb{R}^d$-valued stochastic process $S = (S^1_t, \ldots, S^d_t)_{t\in\{0,\ldots,T\}}$ the *risky assets.*

Example 3.1 (Binomial model)**.** Assume that $d = 1$, i.e., there is only one risky asset. Let $r > -1$ and $u > d > -1$. Assume that the bank account is given by

$$S^0_t = (1 + r)^t, \quad t \in \{0, \ldots, T\},$$

i.e., the interest rate the same across all periods. Moreover, assume that the risky asset $S^1 = (S^1_t)_{t\in\{0,\ldots,T\}}$ is given by

$$S^1_t = S^1_0 \prod_{i=1}^{t} Y_i,$$

[12] Note that the value of r_0 does not play a role.

[13] The bank account is a somewhat artificial asset, that is created by the *roll-over* of short term bonds.

where $S^1_0 > 0$ and $Y_1, \ldots, Y_T$ are i.i.d. random variables satisfying

$$P[Y_i = 1 + u] := p_1 \quad \text{and} \quad P[Y_i = 1 + d] := p_2,$$

where $p_1, p_2 > 0$ and $p_1 + p_2 = 1$.

For a small number of T, e.g. $T = 3$, the above model can be nicely illustrated by the following trees, where the numbers beside the branches denote *transition probabilities*. For convenience, we assume that $S^1_0 = 1$.

$$\begin{tikzcd}
S^0: & 1 \arrow[r, "1"] & 1+r \arrow[r, "1"] & (1+r)^2 \arrow[r, "1"] & (1+r)^3 \\
 & & & & (1+u)^3 \\
 & & & (1+u)^2 \arrow[ur, "p_1"] \arrow[dr, "p_2"'] & \\
 & & 1+u \arrow[ur, "p_1"] \arrow[dr, "p_2"'] & & (1+u)^2(1+d) \\
S^1: & 1 \arrow[ur, "p_1"] \arrow[dr, "p_2"'] & & (1+u)(1+d) \arrow[ur, "p_1"'] \arrow[dr, "p_2"] & \\
 & & 1+d \arrow[ur, "p_1"'] \arrow[dr, "p_2"'] & & (1+u)(1+d)^2 \\
 & & & (1+d)^2 \arrow[ur, "p_1"'] \arrow[dr, "p_2"'] & \\
 & & & & (1+d)^3
\end{tikzcd}$$

Let us finally describe how to give a *rigorous description* of the binomial model. This is for instance important for implementing the binomial model on a computer. To this end, we use the setting of Example 2.4: For Ω, we take the *path space*

$$\Omega := \{1,2\}^T = \Big\{\omega = (x_1, \ldots, x_T) : x_1, \ldots, x_T \in \{1,2\}\Big\},$$

i.e., each $\omega = (x_1, \ldots, x_T)$ describes one path in the tree corresponding to the model. We set $\mathcal{F} := 2^\Omega$ and define the random variables $Y_1, \ldots, Y_T$ by

$$Y_t(\omega) = Y_t((x_1, \ldots, x_T)) = \begin{cases} 1 + u =: y_1 & \text{if } x_t = 1, \\ 1 + d =: y_2 & \text{if } x_t = 2. \end{cases}$$

Moreover, the probability measure P is given by

$$P[\{\omega\}] := P[\{(x_1, \ldots, x_T)\}] = \prod_{t=1}^{T} p_{x_t}.$$

Finally, for $t \in \{1, \ldots, T\}$, we set

$$\mathcal{F}_t = \sigma\left(A_{(x_1,\ldots,x_t)} : x_1, \ldots, x_t \in \{1,2\}\right),$$

where for $x_1, \dots, x_t \in \{1, 2\}$,

$$A_{(x_1,\dots,x_t)} := \{\omega = (\widetilde{x}_1, \dots, \widetilde{x}_T) \in \Omega : \widetilde{x}_1 = x_1, \dots, \widetilde{x}_t = x_t\} = \{Y_1 = y_{x_1}, \dots, Y_t = y_{x_t}\},$$

denotes the node in the tree which is reached by the (partial) path $(x_1, \dots, x_t)$. Then $\mathcal{F}_t = \sigma(Y_1, \dots, Y_t) = \sigma(S_1^1, \dots, S_t^1)$ for $t \in \{1, \dots, T\}$, i.e., $\mathbb{F} = (\mathcal{F}_t)_{t\in\{0,\dots,T\}}$ is the smallest filtration to which S is adapted.

Remark 3.2. Note that for the binomial model, the tree for S^1 is *recombining*, so that the number of nodes only grows *linearly* in time. For *non-recombining* trees, the number of nodes grows *exponentially* in time. This difference is very important from a computational/numerical perspective.

We *discount* by S^0 (or take S^0 as *numéraire*) and define the *discounted assets* $X^0, \dots, X^d$ by

$$X_t^i := \frac{S_t^i}{S_t^0}, \quad t \in \{0, \dots, T\}, \quad i \in \{0, \dots, d\}.$$

Then $X^0 \equiv 1$, and $X = (X_t^1, \dots, X_t^d)_{t\in\{0,\dots,T\}}$ expresses the value of the risky assets in units of the numéraire S^0.

Exercise 3.3. ★☆☆ Let $\overline{S} = (S_t^0, S_t^1)_{t\in\{0,\dots,3\}}$ be a binomial model with parameters $r = 0.1$, $u = 0.21$, $d = -0.01$ and $p_1 = 0.7$. Write down the tree for the discounted stock price X^1.

3.2 Self-financing strategies

To describe trading in the multiperiod market $\overline{S} = (S_t^0, S_t)_{t\in\{0,\dots,T\}}$, we have to describe for each stock i and for each trading period t, the number ϑ_t^i of shares held in asset i in period t, i.e., from $t-1$ to t. As we cannot look into the future, ϑ_t^i can only use the information available at the beginning of period t, i.e., at time $t-1$, so it must be $\mathcal{F}_{t-1}$-measurable.

Definition 3.4. Let $\overline{S} = (S_t^0, S_t)_{t\in\{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\{0,\dots,T\}}, P)$. A *trading strategy* is an $\mathbb{R}^{1+d}$-valued stochastic process $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in\{0,\dots,T\}}$ that is predictable.[14]

If $\overline{\vartheta}$ is a trading strategy for $\overline{S}$, then $\vartheta_t^i S_{t-1}^i$ is the amount invested into asset i at time $t-1$ (after trading), and $\vartheta_t^i S_t^i$ is the resulting value at time t (before trading). So for all assets together, the amount invested at time $t-1$ (after trading) is

$$\overline{\vartheta}_t \cdot \overline{S}_{t-1} = \sum_{i=0}^{d} \vartheta_t^i S_{t-1}^i,$$

[14] Note that the value of $\overline{\vartheta}_0$ does not matter.

and the resulting value at time t (before trading) is

$$\overline{\vartheta}_t \cdot \overline{S}_t = \sum_{i=0}^{d} \vartheta_t^i S_t^i,$$

where $\cdot$ denotes the scalar product in $\mathbb{R}^{d+1}$ or $\mathbb{R}^d$, respectively.

So $\overline{\vartheta}_t \cdot \overline{S}_t$ is the *pre-trading value* of $\overline{\vartheta}$ at time t and $\overline{\vartheta}_{t+1} \cdot \overline{S}_t$ is the *post-trading value* of $\overline{\vartheta}$ at time t. If we assume that there are no transaction consts and we neither withdraw nor inject money at time t, we must have the "book-keeping identity"

$$\overline{\vartheta}_t \cdot \overline{S}_t = \overline{\vartheta}_{t+1} \cdot \overline{S}_t \ \ P\text{-a.s.},$$

i.e., the pre-trading and the post-trading value of the strategy coincide. This leads to the notion of a *self-financing strategy*.

Definition 3.5. Let $\overline{S} = (S_t^0, S_t)_{t \in \{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \in \{0,\dots,T\}}, P)$. A trading strategy $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t \in \{0,\dots,T\}}$ is called a *self-financing strategy* if[15]

$$\overline{\vartheta}_t \cdot \overline{S}_t = \overline{\vartheta}_{t+1} \cdot \overline{S}_t \ \ P\text{-a.s.} \quad \text{for } t \in \{1, \dots, T-1\}. \tag{3.1}$$

The self-financing condition (3.1) is extremely natural from an economic perspective. From a mathematical perspective, however, it is a rather inconvenient constraint. For this reason, we seek to find an alternative characterisation of self-financing strategies. It turns out that to this end, we have to look at discounted quantities and stochastic integrals.

We first generalise the definition of the stochastic integral from Definition 2.13 to vector-valued integrands and integrators.

Definition 3.6. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{N}_0})$ be a filtered measurable space, $X = (X_t^1, \dots, X_t^d)_{t \in \mathbb{N}_0}$ an $\mathbb{R}^d$-valued $\mathbb{F}$-adapted process and $H = (H_t^1, \dots, H_t^d)_{t \in \mathbb{N}_0}$ an $\mathbb{R}^d$-valued $\mathbb{F}$-predictable process. Then the real-valued process $H \bullet X = ((H \bullet X)_t)_{t \in \mathbb{N}_0}$ defined by

$$(H \bullet X)_t := \sum_{k=1}^{t} H_k \cdot (X_k - X_{k-1}) = \sum_{k=1}^{t} \sum_{i=1}^{d} H_k^i (X_k^i - X_{k-1}^i)$$

is called the *discrete stochastic integral* of H with respect to X.

In slight abuse of notation, we often write $H \bullet X_t$ for $(H \bullet X)_t$.

We now introduce the (discounted) gains and value processes.

Definition 3.7. Let $\overline{S} = (S_t^0, S_t)_{t \in \{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \{0,\dots,T\}}, P)$. For a trading strategy $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t \in \{0,\dots,T\}}$, define the

[15]Note that we are not interested in the pre-trading value at time 0 or the post-trading value at time T.

- *(discounted) value process* $(V_t(\overline{\vartheta}))_{t\in\{0,\dots,T\}}$ by

$$V_0(\overline{\vartheta}) := \overline{\vartheta}_1 \cdot \overline{X}_0 \quad \text{and} \quad V_t(\overline{\vartheta}) := \overline{\vartheta}_t \cdot \overline{X}_t \quad \text{for } t \in \{1,\dots,T\};$$

- *(discounted) gains process* $(G_t(\vartheta))_{t\in\{0,\dots,T\}}$ by

$$G_t(\vartheta) := \vartheta \bullet X_t, \quad t \in \{0,\dots,T\}.$$

The name "value process" for $V(\overline{\vartheta})$ comes from the fact that $V(\vartheta)$ denotes the (discounted) value of the strategy $\overline{\vartheta}$ at time t (after trading for $t=0$ and before trading for $t \in \{1,\dots,T\}$). The (discounted) gains process is the stochastic integral of the (risky part of the) trading strategy ϑ with respect to the discounted stock price X.[16]

The following result provides an equivalent characterisation of self-financing strategies. The proof is left as an exercise.

Proposition 3.8. *Let $\overline{S} = (S^0_t, S_t)_{t\in\{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\{0,\dots,T\}}, P)$ and $\overline{\vartheta} = (\vartheta^0_t, \vartheta_t)_{t\in\{0,\dots,T\}}$ a trading strategy. Then the following are equivalent:*

(a) $\overline{\vartheta}$ *is self-financing.*

(b) $\overline{\vartheta}_t \cdot \overline{X}_t = \overline{\vartheta}_{t+1} \cdot \overline{X}_t$ *P-a.s. for $t \in \{1,\dots,T-1\}$.*

(c) $V_t(\overline{\vartheta}) = V_0(\overline{\vartheta}) + G_t(\vartheta) = \overline{\vartheta}_1 \cdot \overline{X}_0 + \vartheta \bullet X_t$ *P-a.s. for $t \in \{0,\dots,T\}$.*

Exercise 3.9. ★☆☆ Prove Proposition 3.8.

The equivalence of (a) and (c) in Proposition 3.8 has an important consequence: Any pair (V_0, ϑ), where $V_0 \in \mathbb{R}$ and $\vartheta = (\vartheta_t)_{t\in\{0,\dots,T\}}$ is a $\mathbb{R}^d$-valued predictable process can be *identified* with the self-financing strategy $\overline{\vartheta} = (\vartheta^0_t, \vartheta_t)_{t\in\{0,\dots,T\}}$ whose value process satisfies

$$V_t(\overline{\vartheta}) = V_0 + G_t(\vartheta) = V_0 + \vartheta \bullet X_t, \quad t \in \{0,\dots,T\}. \tag{3.2}$$

More precisely, define the process $(\vartheta^0_t)_{t\in\{0,\dots,T\}}$ by[17]

$$\vartheta^0_t := V_0 + \vartheta \bullet X - \vartheta_t \cdot X_t = V_0 + \vartheta \bullet X_{t-1} - \vartheta_t \cdot X_{t-1}, \quad t \in \{1,\dots,T\}, \tag{3.3}$$

and set $\overline{\vartheta} := (\vartheta^0, \vartheta)$. It follows from the second equality in (3.3) that ϑ^0 and hence also $\overline{\vartheta}$ are

[16] Note that the value process $V(\overline{\vartheta})$ depends on all $1+d$ coordinates of $\overline{\vartheta}$, whereas the gains process $G(\vartheta)$ only depends on the last d coordinates ϑ of $\overline{\vartheta}$.

[17] The value of ϑ^0_0 does not play a role and we may just choose any constant.

predictable. Moreover, by the definition of the value process and (3.3), we obtain

$$V_0(\overline{\vartheta}) = \vartheta_1^0 + \vartheta_1 \cdot X_0 = V_0 + \vartheta \bullet X_0 - \vartheta_1 \cdot X_0 + \vartheta_1 \cdot X_0 = V_0,$$
$$V_t(\overline{\vartheta}) = \vartheta_t^0 + \vartheta_t \cdot X_t = V_0 + \vartheta \bullet X_t - \vartheta_t \cdot X_t + \vartheta_t \cdot X_t = V_0 + \vartheta \bullet X_t, \quad t \in \{1, \ldots, T\}.$$

We will make use of the identification (3.2) throughout the rest of the chapter. To this end, we introduce the shorthand notation

$$\overline{\vartheta} \mathrel{\widehat{=}} (V_0, \vartheta).$$

3.3 No-arbitrage and the Fundamental Theorem of Asset Pricing

We proceed to define the key notion of arbitrage.

Definition 3.10. Let $\overline{S} = (S_t^0, S_t)_{t \in \{0,\ldots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \{0,\ldots,T\}}, P)$. A self-financing strategy $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t \in \{0,\ldots,T\}}$ is called an *arbitrage opportunity* if

$$\overline{\vartheta}_1 \cdot \overline{S}_0 \leq 0, \quad \overline{\vartheta}_T \cdot \overline{S}_T \geq 0 \;\; P\text{-a.s.} \quad \text{and} \quad P[\overline{\vartheta}_T \cdot \overline{S}_T > 0] > 0.$$

The financial market $\overline{S}$ is called *arbitrage-free* if there are no arbitrage opportunities. In this case one also says that $\overline{S}$ satisfies *NA*.

An arbitrage opportunity gives *something* (a positive chance of strictly positive final wealth $P[\overline{\vartheta} \cdot \overline{S}_T > 0] > 0$) *out of nothing* (zero or negative initial wealth $\overline{\vartheta} \cdot \overline{S}_0 \leq 0$) *without risk* (almost sure nonnegative final wealth $\overline{\vartheta} \cdot \overline{S}_T \geq 0$ P-a.s.).

Remark 3.11. One can show that if the market $\overline{S}$ admits arbitrage, there always exists an arbitrage opportunity with $\overline{\vartheta}_1 \cdot \overline{S}_0 = 0$.

The following result provides an equivalent characterisation of NA in terms of stochastic integrals. Its proof is left as an exercise.

Proposition 3.12. *Let $\overline{S} = (S_t^0, S_t)_{t \in \{0,\ldots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \{0,\ldots,T\}}, P)$. The following are equivalent:*

(a) *The market $\overline{S}$ satisfies NA.*

(b) *There does not exist a predictable process $\vartheta = (\vartheta_t^1, \ldots, \vartheta_t^d)_{t \in \{0,\ldots,T\}}$ such that*

$$\vartheta \bullet X_T \geq 0 \; P\text{-a.s.} \quad \text{and} \quad P[\vartheta \bullet X_T > 0] > 0.$$

Exercise 3.13. ★☆☆ Prove Proposition 3.12.

We proceed to introduce the key notion of an *equivalent martingale measure.* To this end, we first recall what it means for two probability measures P and Q to be *equivalent.*

Definition 3.14. Let $(\Omega, \mathcal{F})$ be a measurable space. Two probability measures P and Q on $(\Omega, \mathcal{F})$ are called *equivalent* on $\mathcal{F}$ (notation: $P \approx Q$) if

$$P[A] = 0 \quad \Leftrightarrow \quad Q[A] = 0, \quad A \in \mathcal{F}.$$

Loosely speaking, two probability measures are equivalent if they agree on "which events will not happen", i.e., have probability zero. But they may still assign different probabilities to "events that may happen".

Exercise 3.15. ★☆☆ Let $\Omega = \{\omega_1, \dots, \omega_N\}$ be a finite sample space, $\mathcal{F} := 2^\Omega$ and P a probability measure on $(\Omega, \mathcal{F})$ with $P[\{\omega_n\}] > 0$ for all $n \in \{1, \dots, N\}$. Show that a probability measure Q on $(\Omega, \mathcal{F})$ is equivalent to P if and only if $Q[\{\omega_n\}] > 0$ for all $n \in \{1, \dots, N\}$.

We proceed to define the concept of an equivalent martingale measure.

Definition 3.16. Let $\overline{S} = (S^0_t, S_t)_{t \in \{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \{0,\dots,T\}}, P)$. Denote by $X := S/S_0$ the discounted risky assets. A measure Q on $(\Omega, \mathcal{F})$ is called an *equivalent martingale measure* (EMM) for X if $Q \approx P$ and each X^i is a Q-martingale, i.e., a martingale under the measure Q.

We also need the following generalisation of Theorem 2.15(a), which requires a more advanced proof; see [3, Theorem 5.14].

Theorem 3.17. *Let $M = (M^1_t, \dots M^d_t)_{t \in \{0,\dots,T\}}$ be a d-dimensional martingale (i.e., each M^i is a martingale) and $H = (H^1_t, \dots, H^d_t)_{t \in \{0,\dots,T\}}$ an $\mathbb{R}^d$-valued predictable process on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \{0,\dots,T\}}, P)$. Suppose that $H \bullet M_T \geq -a$ P-a.s. for some $a \geq 0$. Then $H \bullet M$ is a martingale.*

We proceed to state and prove the easy direction of the fundamental theorem of asset pricing.

Theorem 3.18 (Fundamental Theorem of Asset Pricing). *Let $\overline{S} = (S^0_t, S_t)_{t \in \{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \{0,\dots,T\}}, P)$. The following are equivalent:*

(a) *The market $\overline{S}$ satisfies NA.*

(b) *There exists an EMM for the discounted risky assets $X = S/S^0$.*

Proof. " (b) $\Rightarrow$ (a)". Let $Q \approx P$ be an EMM and $\vartheta = (\vartheta_t^1, \ldots, \vartheta_t^d)_{t\in\{0,\ldots,T\}}$ a predictable process with $\vartheta \bullet X \geq 0$ P-a.s. By Proposition 3.12, it suffices to show that $\vartheta \bullet X = 0$ P-a.s. As $Q \approx P$, we have $\vartheta \bullet X \geq 0$ Q-a.s. It suffices to show that $\vartheta \bullet X = 0$ Q-a.s. Theorem 3.17 gives that $\vartheta \bullet X$ is a Q-martingale, which implies in particular that

$$0 = \vartheta \bullet X_0 = E^Q\left[\vartheta \bullet X_T \,|\, \mathcal{F}_0\right] = E^Q\left[\vartheta \bullet X_T\right]. \qquad \square$$

"(a) $\Rightarrow$ (b)". This is very difficult. For a proof, see [3, Section 1.6 and Theorem 5.16]

We proceed to illustrate the FTAP in the case of the Binomial model.

Example 3.19. Consider the Binomial model from Example 3.1. Set $y_1 := 1+u$ and $y_2 := 1+d$ so that each Y_k takes the values y_1 or y_2. Any probability measure Q on $\mathcal{F}$ can be described by

$$Q[\{(x_1, \ldots, x_T)\}] = q_{x_1} q_{x_2|x_1} q_{x_3|(x_1,x_2)} \times \cdots \times q_{x_T|(x_1,\ldots,x_{T-1})}$$

where $q_{x_1} = Q[Y_1 = y_{x_1}]$ and

$$q_{x_t|(x_1,\ldots,x_{t-1})} := Q[Y_t = y_{x_t} \,|\, Y_1 = y_{x_1}, \ldots, Y_{t-1} = y_{x_{t-1}}], \quad t \in \{2, \ldots, T\}.$$

By Example 3.15, $Q \approx P$ if and only if $q_{x_1}, \ldots, q_{x_T|(x_1,\ldots,x_{t-1})} > 0$ for all $x_1, \ldots, x_T \in \{1, 2\}$. Moreover, if $Q \approx P$, then Q is an EMM for X^1 if and only if[18]

$$E^Q\left[X_t^1 \,\middle|\, \mathcal{F}_{t-1}\right] = X_{t-1}^1 \;\; Q\text{-a.s.}, \quad t \in \{1, \ldots, T\}.$$

By the fact that

$$X_t^1 = S_0^1 \prod_{k=1}^t \frac{Y_k}{1+r} = \frac{Y_t}{1+r} X_{t-1}^1,$$

it follows from the pull-out property of conditional expectations, that X^1 is a Q-martingale if and only if

$$E^Q\left[\frac{Y_t}{1+r} \,\middle|\, \mathcal{F}_{t-1}\right] = 1 \;\; Q\text{-a.s.} \quad \Leftrightarrow \quad E^Q\left[Y_t \,|\, \mathcal{F}_{t-1}\right] = 1 + r \;\; Q\text{-a.s.}$$

It follows from $\mathcal{F}_{t-1} = \sigma\left(A_{(x_1,\ldots,x_{t-1})} : x_1, \ldots, x_{t-1} \in \{1, 2\}\right)$ and Example 1.5 that

$$\begin{aligned} E^Q\left[Y_1 \,|\, \mathcal{F}_0\right] &= E^Q\left[Y_1\right] = q_1(1+u) + q_2(1+d), \\ E^Q\left[Y_t \,|\, \mathcal{F}_{t-1}\right] &= \sum_{x_1,\ldots,x_{t-1}\in\{1,2\}} E^Q\left[Y_t \,\middle|\, A_{(x_1,\ldots,x_{t-1})}\right] \mathbf{1}_{A_{(x_1,\ldots,x_{t-1})}} \;\; Q\text{-a.s.} \end{aligned}$$

[18] Note that Q integrability of X^1 is trivially satisfied as X^1 is bounded.

Moreover, for $x_1, \dots, x_{t-1} \in \{1, 2\}$

$$E^Q \left[Y_t \,\middle|\, A_{(x_1,\dots,x_{t-1})} \right] = (1+u) q_{1|(x_1,\dots,x_{t-1})} + (1+d) q_{2|(x_1,\dots,x_{t-1})}$$

So X^1 is a Q-martingale if and only if

$$\begin{aligned} q_1(1+u) + q_2(1+d) &= 1 + r \\ (1+u) q_{1|(x_1,\dots,x_{t-1})} + (1+d) q_{2|(x_1,\dots,x_{t-1})} &= (1+r), \quad x_1, \dots, x_{t-1} \in \{1,2\}, \quad t \in \{2, \dots, T\} \end{aligned}$$

Using that $q_1 + q_2 = 1$ and $q_{1|(x_1,\dots,x_{t-1})} + q_{2|(x_1,\dots,x_{t-1})} = 1$ it follows that X^1 is a Q-martingale if and only if

$$\begin{aligned} q_1 &= \frac{r-d}{u-d} \quad \text{and} \quad q_2 = \frac{u-r}{u-d}, \\ q_{1|(x_1,\dots,x_{t-1})} &= \frac{r-d}{u-d} \quad \text{and} \quad q_{2|(x_1,\dots,x_{t-1})} = \frac{u-r}{u-d}, \quad x_1, \dots, x_{t-1} \in \{1,2\}, \quad t \in \{2, \dots, T\}. \end{aligned}$$

Note that $q_{1|(x_1,\dots,x_{t-1})}$ and $q_{2|(x_1,\dots,x_{t-1})}$ do not depend on $x_1, \dots, x_{t-1}$. This implies that the Y_k are also independent under Q. Clearly, $q_1, q_2, q_{1|(x_1,\dots,x_{t-1})}, q_{2|(x_1,\dots,x_{t-1})} > 0$ if and only if $d < r < u$. So, in conclusion, the Binomial model is arbitrage-free if and only if $d < r < u$, and in this case there exists a unique EMM for X^1 satisfying

$$Q[\{x_1, \dots, x_T\}] = \prod_{t=1}^{T} q_{x_t},$$

where $q_1 := \frac{r-d}{u-d}$ and $q_2 := \frac{u-r}{u-d}$.

3.4 Valuation of contingent claims

In the previous section, we have fully characterised those models of financial markets in finite discrete time that are reasonable in the sense that they are arbitrage-free. We proceed to study what we can say about the *price* or *value* of a *derivative asset* like a European call or put option in an arbitrage-free market. More precisely, we assume that the market without the derivative asset is arbitrage-free and want to price the derivative asset in such a way that there arise no new arbitrage opportunities.

Let us first give a general definition of derivative assets.

Definition 3.20. Let $\overline{S} = (S_t^0, S_t)_{t \in \{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \{0,\dots,T\}}, P)$. A nonnegative $\mathcal{F}_T$-measurable random variable C is called a *European contingent claim* (with maturity T). It is called a *derivative security* if it can be written as a measurable function of $S_t^0, \dots, S_t^d$ for $t \in \{0, \dots, T\}$.

Example 3.21. (a) The owner of a *European call option* on asset $i \in \{1, \dots, d\}$ with *strike*

$K > 0$ and maturity T has the right but not the obligation to buy asset i at time T for price K. So the value of the option at time T is given by the contingent claim

$$C = (S_T^i - K)^+.$$

(b) The owner of a *European put option* on asset $i \in \{1, \dots, d\}$ with *strike* $K > 0$ and maturity T has the right but not the obligation to sell the asset i at time T for the price K. So the value of the option at time T is given by the contingent claim

$$C = (K - S_T^i)^+.$$

(c) Let A denote the event of an extreme weather situation like hail at time T. It is natural to assume that A is $\mathcal{F}_T$-measurable but independent of the market $\overline{S}$. A toy example of a *weather derivative* is a contract that pays one unit of money at time T if the extreme event A happens and zero otherwise. The corresponding contingent claim is given by

$$C = \mathbf{1}_A.$$

This is unlike the call or put option not a derivative security.

Remark 3.22. (a) The (somewhat confusing) qualifier "European" signifies that the contingent claim may exercised only at one date, at maturity. By contrast, so-called *American contingent claims* can be exercised at any time up to and including maturity. Whereas in reality most contingent claims are American, for simplicity, we only consider European ones in the sequel. For an excellent treatment of American contingent claims, we refer to [3, Chapter 7].

(b) The notion of a European contingent claim can be extended to contracts with maturity $t < T$. A contingent claim C with maturity $t < T$ is just a nonnegative $\mathcal{F}_t$-random variable.

We proceed to study the question how we can assign to a contingent claim C a *value* at times $t < T$, in particular at $t = 0$. Assuming that the underlying market $\overline{S}$ satisfies NA, we want to do this in such a way that we do not create any new arbitrage opportunities. Here, the key idea is the notion of *replication*.

Definition 3.23. Let $\overline{S} = (S_t^0, S_t)_{t \in \{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \{0,\dots,T\}}, P)$. A contingent claim C is called *attainable* or *replicable* if there exists a self-financing trading strategy $\overline{\vartheta}$ such that

$$\overline{\vartheta}_T \cdot \overline{S}_T = C \;\; P\text{-a.s.}$$

In this case $\overline{\vartheta}$ is called a *replication strategy* or *(perfect) hedge* for C.

Remark 3.24. A contingent claim C is attainable if and only if the corresponding *discounted* *contingent claim* $H := \frac{C}{S_T^0}$ satisfies $H = \overline{\vartheta}_T \cdot \overline{X}_T = V_T(\overline{\vartheta})$. In this case, we say that the discounted contingent claim H is *attainable* and call $\overline{\vartheta}$ a *replication strategy* for H.

Exercise 3.25. ★☆☆ Let $\overline{S} = (S_t^0, S_t^1)_{t\in\{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\{0,\dots,T\}}, P)$. Assume that $S_t^0 = (1+r)^t$ for some $r > -1$. Denote the payoff of a call and put option on asset 1 with strike K and maturity T by C^C and C^P, respectively. Suppose that the call option is attainable with replication strategy $\overline{\vartheta}^C = (\vartheta_t^{0,C}, \vartheta_t^{1,C})_{t\in\{0,\dots,T\}}$. Show that the put option is also attainable and find a replication strategy $\overline{\vartheta}^P = (\vartheta_t^{0,P}, \vartheta_t^{1,P})_{t\in\{0,\dots,T\}}$. (*Hint*: Denote the discounted payoffs of the call and put option by H^C and H^P, respectively. First show that $H^C - H^P = X_T - K/(1+r)^T$.)

The following result shows that for arbitrage-free markets, the value process of an attainable contingent claim is unique and can be easily computed if one knows at least one EMM.

Theorem 3.26. *Let $\overline{S} = (S_t^0, S_t)_{t\in\{0,\dots,T\}}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\{0,\dots,T\}}, P)$ and H an attainable discounted contingent claim. Assume that $\overline{S}$ satisfies NA and denote by $\mathcal{P}$ the set of all EMMs for $X = S/S^0$. Then*

(a) $E^Q[H] < \infty$ *for all* $Q \in \mathcal{P}$.

(b) $E^{Q_1}[H \mid \mathcal{F}_t] = E^{Q_2}[H \mid \mathcal{F}_t]$ *P-a.s. for all $Q_1, Q_2 \in \mathcal{P}$ and all $t \in \{0,\dots,T\}$.*

(c) *There exists a P-a.s. unique adapted process $(V_t^H)_{t\in\{0,\dots,T\}}$ with $V_T^H = H$ P-a.s. such that the extended $(1+d+1)$-dimensional market*

$$(\overline{S}, S^0 V^H) = (S_t^0, S_t^1, \dots, S_t^d, S_t^0 V_t^H)_{t\in\{0,\dots,T\}}$$

satisfies NA. It is given by

$$V_t^H = E^Q[H \mid \mathcal{F}_t] \quad P\text{-a.s.}, \quad t \in \{0,\dots,T\}, \quad \text{for all } Q \in \mathcal{P}.$$

(d) *If $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in\{0,\dots,T\}}$ is any replication strategy for H, then*

$$V_t^H = V_t(\overline{\vartheta}) = V_0^H + \vartheta \bullet X_t \quad P\text{-a.s.}, \quad t \in \{0,\dots,T\}, \tag{3.4}$$

and $\vartheta \bullet X$ is a $(Q, \mathbb{F})$-martingale for any EMM Q.

Proof. As H is attainable, there exists a self-financing strategy $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in\{1,\dots,T\}}$ such that

$$H = V_T(\overline{\vartheta}) \;\; P\text{-a.s.}$$

Let $Q \in \mathcal{P}$ and note that $\mathcal{P} \neq \emptyset$ by the fundamental theorem of asset pricing (Theorem 3.18). As $H = V_T(\overline{\vartheta}) \geq 0$ P-a.s. and hence also Q-a.s., it follows from Proposition 3.8 and Theorem

3.17 that $V(\overline{\vartheta}) = V_0(\overline{\vartheta}) + \vartheta \bullet X$ and $\vartheta \bullet X$ are Q-martingales. This implies in particular that $H = V_T(\overline{\vartheta})$ is Q-integrable, and so we have (a) (and the last part of (d)).

Next, fix $Q_1 \in \mathcal{P}$ and define the process $(V_t^H)_{t \in \{0,\dots,T\}}$ by

$$V_t^H = E^{Q_1}\left[H \,|\, \mathcal{F}_t\right], \quad t \in \{0, \dots, T\}. \tag{3.5}$$

Then the Q_1-martingale property of $V(\overline{\vartheta}) = V_0(\overline{\vartheta}) + \vartheta \bullet X$ and $\vartheta \bullet X$, the fact that $H = V_T(\overline{\vartheta})$ and the fact that $Q_1 \approx P$ imply that

$$V_t^H = E^{Q_1}\left[H \,|\, \mathcal{F}_t\right] = E^{Q_1}\left[V_T(\overline{\vartheta}) \,\middle|\, \mathcal{F}_t\right] = V_t(\overline{\vartheta}) \;\; P\text{-a.s.}, \quad t \in \{0, \dots, T\}. \tag{3.6}$$

As the left-hand side of (3.6) does not depend on $\overline{\vartheta}$, we have the first equality in (3.4). This implies in particular that $V_0^H = V(\overline{\vartheta}_0)$. Now the second equality in (3.4) follows from $V(\overline{\vartheta}) = V_0(\overline{\vartheta}) + \vartheta \bullet X$. So we have (d).

Now, let $Q \in \mathcal{P}$ be arbitrary. Then by the Q-martingale property of $V(\overline{\vartheta})$, the fact that $H = V_T(\overline{\vartheta})$, (3.6) and the fact that $Q \approx P$,

$$E^{Q}\left[H \,|\, \mathcal{F}_t\right] = E^{Q}\left[V_T(\overline{\vartheta}) \,\middle|\, \mathcal{F}_t\right] = V_t(\overline{\vartheta}) = V_t^H \;\; P\text{-a.s.}, \quad t \in \{0, \dots, T\} \tag{3.7}$$

As the right-hand side of (3.7) does not depend on Q, we have (b).

Finally, if $(V_t)_{t \in \{0,\dots,T\}}$ is an adapted process with $V_T = H$, by the Fundamental Theorem of Asset Pricing, the extended market $(\overline{S}, S^0 V) = (S_t^0, S_t^1, \dots, S_t^d, S_t^0 V_t)_{t \in \{0,\dots,T\}}$ satisfies NA if and only if there exists an EMM for (X, V), i.e., if and only if there is $Q \in \mathcal{P}$ such that V is a Q-martingale. If there exists $Q \in \mathcal{P}$ such that V is a Q-martingale, then by the fact that $H = V_T$ Q-a.s. (this uses that $Q \approx P$), the martingale property and (3.7), we obtain

$$V_t = E^{Q}\left[V_T \,|\, \mathcal{F}_t\right] = E^{Q}\left[H \,|\, \mathcal{F}_t\right] = V_t^H \;\; P\text{-a.s.}, \quad t \in \{0, \dots, T\}.$$

On the other hand V^H is a Q-martingale for any $Q \in \mathcal{P}$ by (3.7) and Example 2.8(b). So we have (c). □

3.5 Complete markets and the predictable representation property

Theorem 3.26 answers the question of valuation for a (discounted) contingent claim H *provided* that it is attainable. However, it does not give any criterion to determine whether a containing claim is attainable or not, nor does it provide any guidance concerning the valuation of non-attainable contingent claims. While we do not pursue here the (very important and challenging) question how to value non-attainable contingent claims, we now consider the somewhat ideal situation that *all* contingent claims are attainable.

Definition 3.27. A financial market $\overline{S} = (S_t^0, S_t)_{t \in \{0,\dots,T\}}$ on a filtered probability space

$(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\{0,\ldots,T\}}, P)$ is called *complete* if each contingent claim C is attainable. Otherwise, it is called *incomplete*.

The notion of a complete market it intimately linked to the so-called *predictable representation property* of a d-dimensional $(P, \mathbb{F})$-martingale $M = (M_t^1, \ldots, M_t^d)$.

Definition 3.28. Let $M^1 = (M_t^1)_{t\in\{0,\ldots,T\}}, \ldots, M^d = (M_t^d)_{t\in\{0,\ldots,T\}}$ be $(P, \mathbb{F})$-martingales on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\{0,\ldots,T\}}, P)$. Then $M = (M^1, \ldots, M^d)$ is said to have the *predictable representation property* for the pair $(P, \mathbb{F})$, if for every $(P, \mathbb{F})$-martingale $N = (N_t)_{t\in\{0,\ldots,T\}}$, there exists an $\mathbb{R}^d$-valued $\mathbb{F}$-predictable process $H = (H_t^1, \ldots, H_t^d)_{t\in\{0,\ldots,T\}}$ such that

$$N_t = N_0 + H \bullet M_t \;\; P\text{-a.s.}, \quad t \in \{0, \ldots, T\}.$$

The following result links the notion of complete markets, the predictable representation property, and the cardinality of the set of EMMs.

Theorem 3.29 (Second Fundamental Theorem of Asset Pricing)**.** *Let $\overline{S} = (S_t^0, S_t)_{t\in\{0,\ldots,T\}}$ be an arbitrage-free financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in\{0,\ldots,T\}}, P)$. Denote by $X = S/S^0$ the discounted risky assets. Then the following are equivalent:*

(a) *The market $\overline{S}$ is complete.*

(b) *X has the predictable representation property for the pair $(Q^*, \mathbb{F})$ for some EMM Q^*.*

(c) *There exists a unique EMM Q^* for X.*

Proof. "(a) $\Rightarrow$ (b)". We prove the (formally stronger) claim that X has the predictable representation property for the pair $(Q, \mathbb{F})$ for *each* EMM Q. So let Q be an EMM (which exists by the FTAP) and $N = (N_t)_{t\in\{0,\ldots,T\}}$ a $(Q, \mathbb{F})$-martingale. Set $H^{(+)} := N_T^+$ and $H^{(-)} := N_T^-$. Then $H^{(+)}, H^{(-)}$ are (discounted) contingent claims for the market $\overline{S}$. By completeness of $\overline{S}$, they are both attainable. Denote by $\overline{\vartheta}^{(+)} = (\vartheta_t^{0,(+)}, \vartheta_t^{(+)})_{t\in\{0,\ldots,T\}}$ and $\overline{\vartheta}^{(-)} = (\vartheta_t^{0,(-)}, \vartheta_t^{(-)})_{t\in\{0,\ldots,T\}}$ corresponding replicating strategies. It follows from Theorem 3.26(d) that

$$H^{(+)} = E^Q\left[H^{(+)}\right] + \vartheta^{(+)} \bullet X_T \;\; Q\text{-a.s.} \quad \text{and} \quad H^{(-)} = E^Q\left[H^{(-)}\right] + \vartheta^{(-)} \bullet X_T \;\; Q\text{-a.s.}$$

Setting $\vartheta := \vartheta^{(+)} - \vartheta^-$, it follows from linearity of the expectation and linearity of the stochastic integral that

$$N_T = E^Q\left[N_T\right] + \vartheta \bullet X \;\; Q\text{-a.s.} \tag{3.8}$$

Using that N and $\vartheta \bullet X$ are Q-martingales (the latter by Theorem 3.26(d)), it follows that

$$N_t = N_0 + \vartheta \bullet X_t \;\; Q\text{-a.s.}, \quad t \in \{0, \ldots, T\}.$$

"(b) $\Rightarrow$ (c)". Let Q^* be as in (b) and Q another EMM. It suffices to show that $Q = Q^*$. Fix $A \in \mathcal{F} = \mathcal{F}_T$. Set $H := \mathbf{1}_A$. Then defining the Q^*-martingale $N = (N_t)_{t \in \{0,\dots,T\}}$ by $N_t := E\left[H \mid \mathcal{F}_t\right]$, it follows from (b) that there exists a predictable process $\vartheta = (\vartheta^1_t, \dots, \vartheta^d_t)_{t \in \{0,\dots,T\}}$ with

$$\mathbf{1}_A = Q^*[A] + \vartheta \bullet X_T$$

Now take Q-expectations. Note that $\vartheta \bullet X$ is a Q-martingale by the fact that Q is an EMM for X and Theorem 3.17 (using $\vartheta \bullet X_T \geq Q^*[A]$). This yields

$$Q[A] = Q^*[A]. \tag{3.9}$$

As $A \in \mathcal{F}$ was arbitrary, it follows that $Q = Q^*$.

"(c) $\Rightarrow$ (a)". This is very difficult. For a proof see, [3, Theorem 5.32(b)]. □

Theorem 3.29 together with the *First Fundamental Theorem of Asset Pricing* (Theorem 3.18) give a very beautiful and conclusive description of financial markets in finite discrete time:

- *Existence* of an EMM is equivalent to the market being *arbitrage-free.*
- *Uniqueness* of an EMM is equivalent to the market being *complete* (and arbitrage free).

Both results can be extended to continuous or infinite discrete time. However, the precise formulations become more subtle and the proofs far more difficult.

Remark 3.30. One can show that if $\overline{S}$ is complete, then necessarily $\mathcal{F} = \mathcal{F}_T$ is finite. More precisely, it may have at most $(1+d)^T$ *atoms*; see [3, Theorem 5.37]. This shows that even though it makes things nice and simple, completeness is a very restrictive assumption.

3.6 Pricing and hedging in the binomial model

We conclude this chapter by applying the preceding theory to the special but important case of a binomial model.

Theorem 3.31. *Let $\overline{S} = (S^0_t, S^1_t)_{t \in \{0,\dots,T\}}$ be the binomial model on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in \{0,\dots,T\}}, P)$ from Example 3.1. Suppose that $u > r > d$. Then the market $\overline{S}$ is arbitrage-free and complete, and the discounted risky asset $X^1 := S^1/S^0$ satisfies the predictable representation property for $(Q, \mathbb{F})$, where Q is the unique EMM from Example 3.19. Moreover, if H is a discounted contingent claim, its unique arbitrage-free value at time $t \in \{0,\dots,T\}$ is given by*

$$V^H_t = E^Q\left[H \mid \mathcal{F}_t\right]. \tag{3.10}$$

The corresponding hedging strategy $\overline{\vartheta}^H \mathrel{\widehat{=}} (V_0^H, \vartheta^{1,H})$ *is unique and can be calculated as*[19]

$$\vartheta_t^{1,H} = \frac{V_t^H - V_{t-1}^H}{X_t^1 - X_{t-1}^1}, \quad t \in \{1, \dots, T\}. \tag{3.11}$$

Proof. Apart from (3.11), all claims follow from Example 3.19 and Theorems 3.18, 3.26 and 3.29. To establish (3.11), let $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t^1)_{t \in \{0,\dots,T\}}$ be any replication strategy for H. Then by Theorem 3.26(d),

$$V_t^H - V_{t-1}^H = \vartheta_t^{1,H}(X_t^1 - X_{t-1}^1), \quad t \in \{1, \dots, T\}.$$

Since $X_t^1 - X_{t-1}^1 \neq 0$, rearranging gives (3.11) and also shows uniqueness as the right hand side of (3.11) is independent of any replication strategy. □

Remark 3.32. The formula for the hedging strategy (3.11) can be interpreted as the "discrete delta hedge", i.e., the "discrete derivative" of the value process V^H with respect to the underlying X^1.

We finish this section by briefly explaining how (3.10) and (3.11) can be calculated in practise; e.g. on a computer.

By the Q-martingale property of V^H, we can calculate (3.10) by the *recursive algorithm*

$$V_T^H := H \quad \text{and} \quad V_{t-1}^H := E^Q\left[V_t^H \,\middle|\, \mathcal{F}_{t-1}\right], \quad t \in \{1, \dots, T\}.$$

Using that $\mathcal{F}_t = \sigma(\{A_{(x_1,\dots,x_t)} : x_1, \dots, x_t \in \{1,2\}\})$ for $t \in \{1, \dots, T\}$, it follows from Examples 1.5 and 3.19 that

$$V_t^H = \sum_{x_1,\dots,x_t \in \{1,2\}} v_t^H(x_1, \dots, x_t) \mathbf{1}_{A_{(x_1,\dots,x_t)}}, \quad t \in \{1, \dots, T\}, \quad V_0^H = v_0^H,$$

where the functions $v_t^H : \{1,2\}^t \to [0, \infty)$, $t \in \{1, \dots, T\}$, and the number v_0^H can be calculated recursively by

$$\begin{aligned} v_T^H(x_1, \dots, x_T) &= H((x_1, \dots, x_T)), \\ v_{t-1}^H(x_1, \dots, x_{t-1}) &= q_1 v_t^H(x_1, \dots, x_{t-1}, 1) + q_2 v_t^H(x_1, \dots, x_{t-1}, 2), \quad t \in \{1, \dots, T-1\}, \\ v_0^H &= q_1 v_1^H(1) + q_2 v_1^H(2), \end{aligned}$$

with $q_1 = \frac{r-d}{u-d}$ and $q_2 = \frac{u-r}{u-d}$ being the risk-neutral transition probabilities.

[19] As always, the value of $\vartheta_0^{1,H}$ does not play a role and is not unique.

To calculate (3.11), using that $\vartheta_t^{H,1}$ is $\mathcal{F}_{t-1}$ measurable, it follows that

$$\begin{aligned}\vartheta_t^{H,1} &= \sum_{x_1,\ldots,x_{t-1}\in\{1,2\}} \zeta_t^H(x_1,\ldots,x_{t-1})\mathbf{1}_{A_{(x_1,\ldots,x_{t-1})}}, \quad t\in\{2,\ldots,T\},\\ \vartheta_1^{H,1} &= \zeta_1^H,\end{aligned}$$

where the functions $\zeta_t^H : \{1,2\}^{t-1}\to\mathbb{R}$, $t\in\{2,\ldots T\}$, and the number ζ_1^H in $\mathbb{R}$ can be calculated as

$$\zeta_t^H(x_1,\ldots,x_{t-1}) = \frac{v_t^H(x_1,\ldots,x_{t-1},1)-v_{t-1}^H(x_1,\ldots,x_{t-1})}{\xi_t^1(x_1,\ldots,x_{t-1},1)-\xi_{t-1}^1(x_1,\ldots,x_{t-1})} \tag{3.12}$$

$$= \frac{v_t^H(x_1,\ldots,x_{t-1},2)-v_{t-1}^H(x_1,\ldots,x_{t-1})}{\xi_t^1(x_2,\ldots,x_{t-1},2)-\xi_{t-1}^1(x_1,\ldots,x_{t-1})}, \tag{3.13}$$

$$\zeta_1^H = \frac{v_1^H(1)-v_0^H}{\xi_1(1)-\xi_0} = \frac{v_1^H(2)-v_0^H}{\xi_1(2)-\xi_0} \tag{3.14}$$

where $\xi_t^1(x_1,\ldots,x_t) := S_0^1\prod_{k=1}^t \frac{y_{x_k}}{1+r}$, $t\in\{1,\ldots,T\}$, and $\xi_0 := S_0^1$.[20]

Remark 3.33. (a) If we multiply (3.12) and (3.13) by the respective denominators, then subtract the first from the second term on both sides and rearrange (and do similarly in (3.14)), we obtain

$$\zeta_t^H(x_1,\ldots,x_{t-1}) = \frac{v_t^H(x_1,\ldots,x_{t-1},2)-v_t^H(x_1,\ldots,x_{t-1},1)}{\xi_t^1(x_1,\ldots,x_{t-1},2)-\xi_t^1(x_1,\ldots,x_{t-1},1)}, \tag{3.15}$$

$$\zeta_1^H = \frac{v_1^H(2)-v_1^H(1)}{\xi_1(2)-\xi_1(1)}. \tag{3.16}$$

This is the "delta hedging formula" that can be found in most textbooks. Of course computationally, there is no real difference between using (3.12)/(3.14) or (3.15)/(3.16), and it is really a matter of taste. We have opted for the first (and slightly non-standard) version as this allows to write the "delta hedge" in the abstract sense of (3.11).[21]

(b) Note that the value process V^H and the (risky part of the) hedging strategy $\vartheta^{H,1}$ can be calculated *in parallel* while working backwards through the tree for V^H – and this is true both for (3.12)/(3.14) or (3.15)/(3.16).[22]

(c) If the contingent claim H is not *path-dependent*, i.e., $H((x_1,\ldots,x_T))$ is invariant under permutations of $x_1,\ldots,x_T$, the above algorithm can be made substantially faster since the trees for V^H and $\vartheta^{H,1}$ are recombining. Therefore, the complexity of the algorithm grows only linearly (rather then exponentially) with the number of periods T.

Exercise 3.34. ★☆☆ Suppose that $u = 0.21, d = -0.01, r = 0.1$ and $S_0^1 = 100$. Let $T = 2$

[20]Recall that $y_1 = (1+u)$ and $y_2 = (1+d)$.

[21]We also believe that this is the "correct" formula from a stochastic analysis perspective.

[22]Strictly speaking only (3.15)/(3.16) is fully parallel and this might be a small advantage over (3.12)/(3.14).

and calculate the discounted value process V^C as well as the corresponding hedging strategy $\overline{\vartheta}^C \mathrel{\widehat{=}} (V^C, \vartheta^{C,1})$ for a call option on S^1 with maturity 2 and (undiscounted) strike price $K = 121$. Why is the value of $p_1 \in (0,1)$ irrelevant to answer this question?

4 Brownian motion and continuous local martingales

In this chapter we study *Brownian motion*, which is perhaps *the* central object of modern probability theory. We then look more generally at so-called *continuous local martingales*.

4.1 Definition and existence of Brownian motion

Definition 4.1. A real-valued stochastic process $W = (W_t)_{t\geq 0}$ on some probability space $(\Omega, \mathcal{F}, P)$ is called a *Brownian motion* (with respect to P) if the following three properties are satisfied:

(1) $P[W_0 = 0] = 1$.

(2) For any $n \in \mathbb{N}$ and $0 \leq t_0 < t_1 < \cdots < t_n < \infty$, the *increments* $W_{t_i} - W_{t_{i-1}}$, $i \in \{1, \ldots, n\}$, are independent and normally distributed with mean 0 and variance $t_i - t_{i-1}$.

(3) For P-almost every ω, the path $t \mapsto W_t(\omega)$ is continuous.

Remark 4.2. (a) The first rigorous proof for existence of Brownian motion was given by the mathematician N. Wiener (1894-1964) in 1923. For this reason, Brownian motion is often denoted by the letter W.

(b) Definition 4.1 does not mention any filtration. If we are given a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ on $(\Omega, \mathcal{F})$, we say that process $W = (W_t)_{t\geq 0}$ is *Brownian motion with respect to* $\mathbb{F}$ (and P), if it is adapted to $\mathbb{F}$, satisfies properties (1) and (3) of Definition 4.1 as well as

(2') For $0 \leq s < t < \infty$, the increments $W_t - W_s$ is independent of $\mathcal{F}_s$ and normally distributed with mean 0 and variance $t - s$.

(c) One can show that $W = (W_t)_{t\geq 0}$ is a Brownian motion in the sense of Definition 4.1 if and only if W is also a Brownian motion with respect to its *natural filtration* $\mathbb{F}^W := (\mathcal{F}^W_t)_{t\geq 0}$, where

$$\mathcal{F}^W_t := \sigma(W_s : s \leq t) = \sigma\left(\bigcup_{s=0}^{t} \sigma(W_s)\right), \quad t \geq 0.$$

(d) An $\mathbb{R}^d$-valued stochastic process $W = (W^1_t, \ldots, W^d_t)_{t\geq 0}$ is called a d-dimensional Brownian motion if $W^1, \ldots, W^d$ are independent one-dimensional Brownian motions.[23]

Exercise 4.3. ★☆☆ Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on some probability space $(\Omega, \mathcal{F}, P)$. Compute the function $t \mapsto E\left[W_t^2\right]$.

[23] If W is a d-dimensional Brownian motion with respect to a filtration $\mathbb{F}$, then for each $0 \leq s < t < \infty$, $W^1_t - W^1_s \ldots, W^d_t - W^d_s$ and $\mathcal{F}_s$ are independent.

We now turn to the question of *existence* of Brownian motion. We are using the construction via the so-called *functional central limit theorem.* To motivate this construction, let $X_1, X_2, \dots$ be i.i.d. random variables with $P[X_i = 1] = P[X_i = -1] = 1/2$ and $S_t := \sum_{k=1}^{t} X_k$, i.e., $S = (S_t)_{t \in \mathbb{N}_0}$ is the symmetric random walk. We now scale S in time and space. To wit, for $n \in \mathbb{N}$, define the continuous-time process $(W_t^n)_{t \geq 0}$ by

$$W_t^n = \frac{1}{\sqrt{n}} S_{\lfloor tn \rfloor},$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Note that the spacing is $1/n$ in time and $1/\sqrt{n}$ in space.

By the central limit theorem, it is not difficult to check that for each fixed t, W_t^n converges in distribution to a normal random variable with mean 0 and variance t as $n \to \infty$. Similarly, it is not difficult to check that for any $m, n \in \mathbb{N}$ and $0 \leq t_0 < t_1 < \ldots < t_m$, the increments $W_{t_i}^n - W_{t_{i-1}}^n$, $i \in \{1, \dots, m\}$, are independent and by the multidimensional central limit theorem, the random vector $(W_{t_1}^n - W_{t_0}^n, \dots, W_{t_m}^n - W_{t_{m-1}}^n)$ converges in distribution to a multidimensional random vector that is multivariate normal distributed with mean vector 0 and a diagonal covariance matrix $\mathrm{diag}(t_1 - t_0, \dots, t_m - t_{m-1})$.[24] We now linearly interpolate each W^n by setting

$$\tilde{W}_t^n = W_t^n + \frac{1}{\sqrt{n}} X_{\lfloor tn \rfloor + 1}(tn - \lfloor tn \rfloor), \quad t \geq 0.$$

Then each process $\tilde{W}^n = (\tilde{W}_t^n)_{t \geq 0}$ has continuous paths and hence satisfies properties (1) and (3) of a Brownian motion exactly. For large $n \in \mathbb{N}$, $\tilde{W}^n$ also satisfies property (2) of a BM approximately. Therefore, we might hope that the limit $\lim_{n \to \infty} \tilde{W}^n$ is a Brownian motion. The following theorem shows that this is indeed the case (even under more general assumptions on the X_i), if we take the limit in distribution. The proof of this so-called *functional central limit theorem* is rather involved. We refer to [9, Theorem 21.43 and Chapter 21] for details.

Theorem 4.4 (Functional central limit theorem)**.** *Let $X_1, X_2, \dots$ be i.i.d. random variables with $E[X_i] = 0$ and $\mathrm{Var}[X_i] = 1$. Define the discrete time process $S = (S_t)_{t \in \mathbb{N}_0}$ by $S_t := \sum_{k=1}^{t} X_k$. For each $n \in \mathbb{N}$, define the continuous-time scaled and interpolated process $(\tilde{W}_t^n)_{t \geq 0}$ by*

$$\tilde{W}_t^n = \frac{1}{\sqrt{n}} S_{\lfloor tn \rfloor} + \frac{1}{\sqrt{n}} X_{\lfloor tn \rfloor + 1}(nt - \lfloor tn \rfloor). \tag{4.1}$$

Then the processes $(\tilde{W}^n)_{n \in \mathbb{N}}$ converge weakly to a Brownian motion W as $n \to \infty$.[25]

[24] Here $\mathrm{diag}(t_1 - t_0, \dots, t_m - t_{m-1})$ stands for the $m \times m$ matrix that has the entries $t_1 - t_0, \dots, t_m - t_{m-1}$ on the diagonal and 0 otherwise.

[25] The correct state space for this convergence is the paths space $(C[0, \infty), \mathcal{B}_{C[0,\infty)})$, where $C[0, \infty)$ denotes all continuous function on $[0, \infty)$, endowed with the topology of uniform convergence on compact sets; see [9, Section 21.6] for details.

4.2 Fundamental properties of Brownian motion

In this section, we present some fundamental properties of Brownian motion,

We first study some elementary martingale properties of Brownian motion.

Proposition 4.5. *Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$. The the following processes are martingales:*

(a) *W itself.*

(b) $(W_t^2 - t)_{t\geq 0}$.

(c) $\exp(aW_t - \frac{1}{2}a^2 t)_{t\geq 0}$ *for $a \in \mathbb{R}$.*

Proof. We only establish parts (a) and (b); part (c) is left as an exercise.

(a) Adaptedness of W holds by definition and integrability follows from the fact that $W_t = W_t - W_0$ P-a.s. is normal distributed with mean 0 and variance $t - 0 = t$ by (2').[26] To check the martingale property, we use (2'). So fix $s \leq t$. Then $W_t - W_s$ is independent of $\mathcal{F}_s$ by (2') and normal distributed with mean 0 and variance $t - s$. Hence, by the independence property of conditional expectations, we have

$$E\left[W_t \,|\, \mathcal{F}_s\right] = E\left[W_s + (W_t - W_s) \,|\, \mathcal{F}_s\right] = W_s + E\left[W_t - W_s\right] = W_s \;\; P\text{-a.s.}$$

(b) Adaptedness of $(W_t^2 - t)_{t\geq 0}$ follows from the fact that for each $t \geq 0$, $W_t^2 - t$ is $\mathcal{F}_t$-measurable because W_t is and t is a constant. Integrability follows as above from the fact that $W_t \sim \mathcal{N}(0, t)$. To check the martingale property, we use (2'). So fix $s \leq t$. Using that W_s is measurable, W is a martingale by (a), and $W_t - W_s$ is independent of $\mathcal{F}_s$ by (2') and normal distributed with mean 0 and variance $t - s$, we obtain by linearity, the pull-out property and the independence property of conditional expectations,

$$\begin{aligned}
E\left[W_t^2 - t \,\middle|\, \mathcal{F}_s\right] &= E\left[(W_s + (W_t - W_s))^2 \,\middle|\, \mathcal{F}_s\right] - t \\
&= E\left[W_s^2 + 2W_s(W_t - W_s) + (W_t - W_s)^2 \,\middle|\, \mathcal{F}_s\right] - t \\
&= W_s^2 + 2W_s E\left[(W_t - W_s) \,|\, \mathcal{F}_s\right] + E\left[(W_t - W_s)^2 \,\middle|\, \mathcal{F}_s\right] - t \\
&= W_s^2 + 2W_s \times 0 + E\left[(W_t - W_s)^2\right] - t \\
&= W_s^2 + (t - s) - t = W_s^2 - s \;\; P\text{-a.s.}
\end{aligned}$$

□

The next result list some transformation properties of Brownian motion.

Proposition 4.6. *Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$. Then for $s \geq 0$ and $c > 0$ the following processes are again Brownian motions:*

(a) $(-W_t)_{t\geq 0}$.

[26] Recall that two random variables that are P-a.s. equal have the same distribution.

(b) $(W_{t+s} - W_s)_{t \geq 0}$.

(c) $(cW_{c^{-2}t})_{t \geq 0}$.

Exercise 4.7. ★☆☆ Prove Proposition 4.6.

We now turn to some important paths properties of Brownian motion; for a proof see [9, Theorems 21.17 and 22.1].[27]

Theorem 4.8. *Let $W = (W_t)_{t \geq 0}$ be a Brownian motion on some probability space $(\Omega, \mathcal{F}, P)$.*

(a) *For P-almost every ω, the trajectory $t \mapsto W_t(\omega)$ is nowhere differentiable.*

(b) *We have the* strong law of large numbers*:*

$$\lim_{t \to \infty} \frac{W_t}{t} = 0 \quad P\text{-a.s.}$$

(c) *We have the* law of iterated logarithm*:*

$$\limsup_{t \to \infty} \frac{W_t}{\sqrt{2t \log(\log(t))}} = 1 \quad P\text{-a.s.}$$

Our final result in this section shows that Proposition 4.6(b) generalises to stopping times, i.e., $W_{t+\tau} - W_\tau$ is again a Brownian motion, when τ is a stopping time. This is know as the *strong Markov property* of Brownian motion; for a proof, see [7, Theorem 13.11].

Theorem 4.9. *Let $W = (W_t)_{t \geq 0}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ and let τ be an $\mathbb{F}$-stopping time with $\tau < \infty$ P-a.s. For $t \geq 0$, set $\mathcal{G}_t = \mathcal{F}_{\tau+t}$. Then the process $\widetilde{W} = (\widetilde{W}_t)_{t \geq 0}$ defined by*

$$\widetilde{W}_t := (W_{\tau+t} - W_\tau).$$

is independent of $\mathcal{F}_\tau$ and a Brownian motion with respect to $\mathbb{G} = (\mathcal{G}_t)_{t \geq 0}$.

4.3 The variation of Brownian motion

Theorem 4.8(a) shows that the paths of Brownian motion are very "rough". To understand this better, we next look at the variation and quadratic variation of Brownian motion. To this end, we first need to introduce some notation.

Definition 4.10. Let $t > 0$. A *partition* of $[0, t]$ is a finite set $\Pi = \{t_0, t_1, \ldots, t_n\}$ with $0 = t_0 < t_1 < \cdots < t_n = t$. The *mesh size* of a partition Π is denoted by $|\Pi|$ and defined as

$$|\Pi| := \sup\{t_1 - t_0, t_2 - t_1, \ldots, t_n - t_{n-1}\}.$$

[27]Note that part (b) in Theorem 4.8 follows directly from part (c).

Example 4.11. For $n \in \mathbb{N}$ set $\Pi_n := \{k2^{-n} : k = 0, \dots, 2^n\}$. Then Π_n is a partition of $[0,1]$ with mesh size $|\Pi| = 2^{-n}$.

Definition 4.12. Let $F : [0,\infty) \to \mathbb{R}$ be a continuous function. Then for $t \geq 0$, the *total variation of* F at time t is defined by $|F|(0) := 0$ and

$$|F|(t) := \sup\left\{ \sum_{t_i \in \Pi \setminus \{t\}} \big|F(t_{i+1}) - F(t_i)\big| : \Pi \text{ partition of } [0,t] \right\}, \quad t > 0.$$

If $|F|(t)$ is finite, we say that F is *of finite variation* on $[0,t]$, and if $|F|(t) = \infty$, we say that F is of *infinite variation* on $[0,t]$. If $|F|(t) < \infty$ for all $t \geq 0$, we say that F is *of finite variation*.

Remark 4.13. One can show that if a continuous function $F : [0,\infty) \to \mathbb{R}$ is of finite variation, then the function $t \mapsto |F|(t)$ is again continuous and nondecreasing.

Example 4.14. Let $F : [0,\infty) \to \mathbb{R}$ be a continuously differentiable function. Then F is of finite variation and

$$|F|(t) = \int_0^t |F'(s| \,\mathrm{d}s, \quad t \geq 0.$$

Indeed, fix $t > 0$ and let $\Pi = \{t_0, \dots, t_n\}$ be a partition of $[0,t]$. Then by the mean value theorem,

$$\sum_{i=0}^{n-1} \big|F(t_{i+1}) - F(t_i)\big| = \sum_{i=0}^{n-1} \big|F'(\vartheta_i)(t_{i+1} - t_i)\big| = \sum_{i=0}^{n-1} \big|F'(\vartheta_i)\big|(t_{i+1} - t_i)$$

where $\vartheta_i \in (t_i, t_{i+1})$. Now the claim follows from the definition of the Riemann integral.[28]

Exercise 4.15. ★☆☆ $F : [0,\infty) \to \mathbb{R}$ be given by $F(t) = (t-1)^2$. Compute $|F|$.

Given that the the paths of Brownian motion are almost surely nowhere differentiable by Theorem 4.8(a), we may guess that the paths of Brownian motion are almost surely of infinite variation. This is the content of the following result; its proof is left as an exercise.

Theorem 4.16. *Let $W = (W_t)_{t \geq 0}$ be a Brownian motion on some probability space $(\Omega, \mathcal{F}, P)$. Then $|W|_t = +\infty$ P-a.s. for each $t > 0$.*

Given that the total variation of Brownian motion is almost surely infinite, we might next look at the question if there is some higher variation of Brownian that is finite. It turns out that the quadratic variation of Brownian motion is finite.

[28] Note that for Riemann integrable functions, the Riemann integral coincides with the Lebesgue-integral; see [9, Chapter 4.3].

Theorem 4.17. *Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on some probability space $(\Omega, \mathcal{F}, P)$. Fix $t > 0$. Then for any sequence of partitions $(\Pi_n)_{n\in\mathbb{N}}$ of $[0,t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$,*

$$\sum_{t_i \in \Pi_n \setminus \{t\}} (W_{t_{i+1}} - W_{t_i})^2 \xrightarrow{P} t.$$

Proof. Let $(\Pi_n)_{n\in\mathbb{N}}$ be any sequence of partitions of $[0,t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$. Set

$$Y_n := \sum_{t_i \in \Pi_n \setminus \{t\}} (W_{t_{i+1}} - W_{t_i})^2, \quad n \in \mathbb{N}.$$

Then for each $t_i \in \Pi_n \setminus \{t\}$, we have

$$E\left[(W_{t_{i+1}} - W_{t_i})^2\right] = t_{i+1} - t_i, \tag{4.2}$$

$$\begin{aligned}\mathrm{Var}[(W_{t_{i+1}} - W_{t_i})^2] &= E\left[(W_{t_{i+1}} - W_{t_i})^4\right] - E\left[(W_{t_{i+1}} - W_{t_i})^2\right]^2 \\ &= 3(t_{i+1} - t_i)^2 - (t_{i+1} - t_i)^2 = 2(t_{i+1} - t_i)^2. \end{aligned} \tag{4.3}$$

Summing (4.2) over $t_i \in \Pi_n \setminus \{t\}$ yields

$$E[Y_n] = \sum_{t_i \in \Pi_n \setminus \{t\}} E\left[(W_{t_{i+1}} - W_{t_i})^2\right] = \sum_{t_i \in \Pi_n \setminus \{t\}} t_{i+1} - t_i = t.$$

Moreover, using that the $W_{t_{i+1}} - W_{t_i}$ are independent, the Bienaymé formula (Proposition A.69) and (4.3) give

$$\begin{aligned}\mathrm{Var}[Y_n] &= \sum_{t_i \in \Pi_n \setminus \{t\}} \mathrm{Var}[(W_{t_{i+1}} - W_{t_i})^2] = \sum_{t_i \in \Pi_n \setminus \{t\}} 2(t_{i+1} - t_i)^2 \\ &\leq 2|\Pi_n| \left(\sum_{t_i \in \Pi_n \setminus \{t\}} t_{i+1} - t_i \right) = 2|\Pi_n| t.\end{aligned}$$

We may conclude that

$$\lim_{n\to\infty} E\left[(Y_n - t)^2\right] = \lim_{n\to\infty} \mathrm{Var}[Y_n] \leq \lim_{n\to\infty} 2|\Pi_n| t = 0.$$

Hence, $(Y_n)_{n\in\mathbb{N}}$ converges to t in L^2 and hence in probability. □

Theorem 4.17 justifies the following definition.

Definition 4.18. Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$. Then the *quadratic variation* of W is the process $\langle W \rangle = (\langle W \rangle_t)_{t\geq 0}$ defined by

$$\langle W \rangle_t := t, \quad t \geq 0.$$

4.4 Square-integrable martingales and their quadratic variation

Brownian motion is the key example of a continuous martingale that is square integrable. In this section, we study square-integrable martingales with continuous paths and define their quadratic variation. To make the notation more precise, for a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$, we denote by $\mathcal{M}^{2,c}$ the set of all continuous square-integrable $(P, \mathbb{F})$-martingales such that M_0 is a constant,[29] and set $\mathcal{M}_0^{2,c} := \{M \in \mathcal{M}^{2,c} : M_0 = 0\}$. We denote the subset of all L^2-bounded continuous martingales by $\mathcal{H}^{2,c}$ and set $\mathcal{H}_0^{2,c} := \{M \in \mathcal{H}^{2,c} : M_0 = 0\}$.

We proceed to introduce the concept of a continuous finite variation process.

Definition 4.19. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability. A continuous $\mathbb{F}$-adapted process $A = (A_t)_{t\geq 0}$ is said to be of finite variation if A_0 is a constant[30] and $|A|_t < \infty$ for all $t \geq 0$; cf. Definition 4.12. In this case we write $A \in \mathrm{FV}^c$, and $A \in \mathrm{FV}_0^c$ if in addition $A_0 = 0$.

The following result shows that a continuous square integrable martingale with finite variation is constant.

Theorem 4.20. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space. If $M \in \mathcal{M}^{2,c} \cap \mathrm{FV}^c$, then the paths of M are P-a.s.-constant.*

Proof. We prove the result under the additional assumption that the variation of M is square-integrable, i.e., $E\left[|M|_t^2\right] < \infty$ for all $t \geq 0$.

It suffices to show that $M_t = M_0$ P-a.s. for all $t \geq 0$. So fix $t > 0$. First, for any partition Π of $[0, t]$, the martingale property of M gives

$$E\left[(M_t - M_0)^2\right] = E\left[\left(\sum_{t_i \in \Pi\setminus\{t\}} M_{t_{i+1}} - M_{t_i}\right)^2\right] = E\left[\sum_{t_i \in \Pi\setminus\{t\}} \left(M_{t_{i+1}} - M_{t_i}\right)^2\right] \tag{4.4}$$

because the mixed terms disappear. Indeed, for $i < j$, using that $M_{t_{i+1}} - M_{t_i}$ is $\mathcal{F}_{t_j}$-measurable, we obtain by the martingale property of M,

$$\begin{aligned} E\left[(M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})\right] &= E\left[E\left[(M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j}) \,\middle|\, \mathcal{F}_{t_j}\right]\right] \\ &= E\left[(M_{t_{i+1}} - M_{t_i}) E\left[M_{t_{j+1}} - M_{t_j} \,\middle|\, \mathcal{F}_{t_j}\right]\right] \\ &= E\left[(M_{t_{i+1}} - M_{t_i}) \times 0\right] = 0. \end{aligned}$$

Now, let $(\Pi_n)_{n\in\mathbb{N}}$ be any sequence of partitions of $[0, t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$. For each

[29] This condition can be slightly generalised.

[30] This assumption is only made for consistency reasons with our definition of local martingales.

$n \in \mathbb{N}$, we have

$$\begin{aligned}\sum_{t_i \in \Pi_m \setminus \{t\}} \left(M_{t_{i+1}} - M_{t_i}\right)^2 &\leq \sup_{t_i \in \Pi_m \setminus \{t\}} \left|M_{t_{i+1}} - M_{t_i}\right| \sum_{t_i \in \Pi_m \setminus \{t\}} \left|M_{t_{i+1}} - M_{t_i}\right| \\ &\leq \sup_{t_i \in \Pi_m \setminus \{t\}} \left|M_{t_{i+1}} - M_{t_i}\right| |M|_t \leq |M|_t^2. \end{aligned} \tag{4.5}$$

As M is continuous, it is uniformly continuous on $[0, t]$. Thus, $\lim_{n\to\infty} |\Pi_n| = 0$ yields

$$\lim_{n\to\infty} \sup_{t_i \in \Pi_n \setminus \{t\}} \left|M_{t_{i+1}} - M_{t_i}\right| = 0 \;\; P\text{-a.s.}$$

This together with (4.5) and dominated convergence (using that $E\left[|M_t|^2\right] < \infty$) imply that $\sum_{t_i \in \Pi_m \setminus \{t\}} \left(M_{t_{i+1}} - M_{t_i}\right)^2$ converges to 0 P-a.s. and in L^1 as $n \to \infty$.

Combining this with (4.4) shows that $E\left[(M_t - M_0)^2\right] = 0$, and thus $M_t = M_0$ P-a.s. □

The following corollary gives an alternative characterisation of the quadratic variation of Brownian motion.

Corollary 4.21. *Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$. Then $\langle W \rangle = (\langle W \rangle_t)_{t\geq 0} = (t)_{t\geq 0}$ is the P-a.s. unique process in FV_0^c such that $(W_t^2 - \langle W \rangle_t)_{t\geq 0}$ is a continuous martingale.*

Proof. Existence follows from Proposition 4.5(b).

We prove uniqueness only in the class of all $A \in \mathrm{FV}_0^c$ that are also square-integrable. So let $A \in \mathrm{FV}_0^c$ be square integrable and such that $(W_t^2 - A_t)_{t\geq 0}$ is a continuous martingale. Then the process $M = (M_t)_{t\geq 0}$, defined by

$$M_t := (W_t^2 - t) - (W_t^2 - A_t) = A_t - t, \quad t \geq 0,$$

is a continuous martingale, square-integrable and of finite variation.[31] Hence, it must be constant by Theorem 4.20, and this constant is 0 because $M_0 = A_0 - 0 = 0$. We may conclude that $A_t = t$ for all t P-a.s. □

We now aim to define for each square-integrable continuous martingale M, its quadratic variation $\langle M \rangle$. The idea is to proceed as in Theorem 4.17 and to define $\langle M \rangle_t$ for each t as the limit in probability of $\sum_{t_i \in \Pi_n \setminus \{t\}} (M_{t_{i+1}} - M_{t_i})^2$, where $(\Pi_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} |\Pi_n| = 0$ is a sequence of partitions of $[0, t]$. However, unlike in the case of Brownian motion, it is a priori not clear if this limit exists, is independent of the choice of partitions, and if the path $t \mapsto \langle M \rangle_t$ is continuous and increasing. For technical reasons, we shall therefore henceforth assume that the filtration satisfies the so-called *usual conditions* of right-continuity and completeness; see

[31] Note that $|A + B|_t \leq |A|_t + |B|_t$ for all $A, B \in \mathrm{FV}_0^c$.

Appendix B for details. Among others, this allows us to assume that paths of processes are continuous for *every* ω.

The following result, which is one of the cornerstones of stochastic analysis, shows that quadratic variation $\langle M \rangle$ exists for each $M \in \mathcal{M}^{2,c}$ and can be characterised in an abstract fashion as in Corollary 4.21. The rather complicated proof can be found in [11, Theorems IV.(1.3) and IV.(1.8)].

Theorem 4.22. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions. For $M \in \mathcal{M}^{2,c}$, there exists a P-a.s. unique process $\langle M \rangle = (\langle M \rangle_t)_{t \geq 0}$ in* FV_0^c, *called the* quadratic variation of M, *such that*

$$(M_t^2 - \langle M \rangle_t)_{t \geq 0} \text{ is a continuous martingale.} \tag{4.6}$$

It is nondecreasing and for any $t > 0$ and any sequence of partitions $(\Pi_n)_{n \in \mathbb{N}}$ of $[0, t]$ with $\lim_{n \to \infty} |\Pi_n| = 0$,

$$\sum_{t_i \in \Pi_n \setminus \{t\}} (M_{t_{i+1}} - M_{t_i})^2 \xrightarrow{P} \langle M \rangle_t. \tag{4.7}$$

Exercise 4.23. ★☆☆ Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{M}^{2,c}$. Using the defining property (4.6) of the quadratic variation, show that $\langle \lambda M \rangle = \lambda^2 \langle M \rangle$ for $\lambda \in \mathbb{R}$.

The following corollary introduces the *covariation* of two continuous square-integrable martingales M and N.

Corollary 4.24. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions. For $M, N \in \mathcal{M}^{2,c}$, there exists a P-a.s. unique process $\langle M, N \rangle = (\langle M, N \rangle_t)_{t \geq 0}$ in* FV_0^c, *called the* covariation of M and N, *such that*

$$(M_t N_t - \langle M, N \rangle_t)_{t \geq 0} \text{ is a continuous martingale.} \tag{4.8}$$

Moreover, for any $t > 0$ and any sequence of partitions $(\Pi_n)_{n \in \mathbb{N}}$ of $[0, t]$ with $\lim_{n \to \infty} |\Pi_n| = 0$,

$$\sum_{t_i \in \Pi_n \setminus \{t\}} (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}) \xrightarrow{P} \langle M, N \rangle_t. \tag{4.9}$$

Proof. We only establish existence of $\langle M, N \rangle$. Uniqueness of $\langle M, N \rangle$ and (4.9) are left as an exercise.

By the defining property (4.8) of the covariation, we have to find a process $A \in \mathrm{FV}_0^c$ such that $MN - A$ is a continuous martingale. To this end, we use the method of *polarisation*. Noting that $MN = \frac{1}{4}(M + N)^2 - \frac{1}{4}(M - N)^2$, we set

$$A := \langle M, N \rangle := \frac{1}{4} \langle M + N \rangle - \frac{1}{4} \langle M - N \rangle. \tag{4.10}$$

Then $A \in \mathrm{FV}_0^c$ because $\langle M+N \rangle$ and $\langle M-N \rangle$ are. Moreover, $(M+N)^2 - \langle M+N \rangle$ and $(M-N)^2 - \langle M-N \rangle$ are continuous martingales by defining property of the quadratic variation. Since continuous martingales form a vector space, it follows that also

$$\frac{1}{4}\big((M+N)^2 - \langle M+N \rangle\big) - \frac{1}{4}\big((M-N)^2 - \langle M-N \rangle\big) = MN - A$$

is a continuous martingale. $\square$

Exercise 4.25. ★☆☆ Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M, N \in \mathcal{M}^{2,c}$. Using the defining properties of the quadratic variation (4.6) and the covariation (4.8) show that $\langle M+N \rangle = \langle M \rangle + 2\langle M, N \rangle + \langle N \rangle$.

The following result shows that the covariation of two independent continuous square-integrable martingales vanishes. Its proof is left as an exercise.

Proposition 4.26. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M, N \in \mathcal{M}^{2,c}$ be independent. Then*

$$\langle M, N \rangle = 0 \ \ P\text{-a.s.}$$

Exercise 4.27. ★☆☆ Let $B = (B_t)_{t\geq 0}$ and $W = (W_t)_{t\geq 0}$ be independent Brownian motions on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Compute $\langle B + W \rangle$.

4.5 Continuous local martingales

While the class of square-integrable continuous martingales has good properties and we can define the concept of quadratic variation and co-variation, it is not large enough to develop a powerful theory of stochastic integration in full generality. For this reason, we need to extend this class.[32]

Recall that we are working the under usual *usual conditions* of right-continuity and completeness; see Appendix B for details. Among others, this allows us to make use of the very deep *début theorem*; see [7, Theorem 7.7] for a proof.

Theorem 4.28 (Début theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $X = (X_t)_{t\geq 0}$ an $\mathbb{F}$-adapted real-valued process with continuous paths. Let $A \in \mathcal{B}_{\mathbb{R}}$ be a Borel set. Then the first hitting time of A, given by*

$$\tau_A(\omega) := \inf\{t \geq 0 : X_t(\omega) \in A\},$$

is an $\mathbb{F}$-stopping time.

[32] Also note that from a technical perspective, the proofs of Theorems 4.20 and Corollary 4.21 were somewhat incomplete.

We proceed to define the notion of *localisation*, which is key to extending the class of square-integrable martingales.

Definition 4.29. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $X = (X_t)_{t \geq 0}$ an $\mathbb{F}$-adapted process with continuous paths. Let $\mathfrak{E}$ be a property that X may or may not have. Then X is said to satisfy $\mathfrak{E}$ *locally* if there exists a nondecreasing sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \tau_n = \infty$ P-a.s. such that for each n, the stopped process $X^{\tau_n} = (X_t^{\tau_n})_{t \geq 0}$ has property $\mathfrak{E}$.[33] Moreover, $(\tau_n)_{n \in \mathbb{N}}$ is called a *localising sequence* for X (for the property $\mathfrak{E}$).

We illustrate this concept of localisation by showing that every real-valued continuous process is *locally bounded*.

Proposition 4.30. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $X = (X_t)_{t \geq 0}$ an $\mathbb{F}$-adapted real-valued process with continuous paths and X_0 a constant. Then X is locally bounded.*

Proof. We seek to define a nondecreasing sequence of $\mathbb{F}$-stopping times $(\tau_n)_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} \tau_n = \infty$ P-a.s. such that for each n, the stopped process $X^{\tau_n} = (X_t^{\tau_n})_{t \geq 0}$ is bounded, i.e., there exist a constant C such that $|X_t^{\tau_n}(\omega)| \leq C$ for all $t \in [0, \infty), \omega \in \Omega$. For $n \in \mathbb{N}$, set

$$\tau_n := \{\inf t \geq 0 : |X_t| \geq n\}.$$

Then each τ_n is an $\mathbb{F}$-stopping time by the début theorem (Theorem 4.28) because it is the first hitting time of the Borel set $(-\infty, -n] \cup [n, \infty)$. Moreover, $(\tau_n)_{n \in \mathbb{N}}$ is nondecreasing and satisfies $\lim_{n \to \infty} \tau_n = \infty$ P-a.s. because X is continuous.[34] By continuity of the paths of X,

$$|X_{\tau_n}| = \begin{cases} |X_0| & \text{if } |X_0| > n, \\ n & \text{if } |X_0| \leq n. \end{cases}$$

This together with the definition of τ_n implies that $|X_t^{\tau_n}(\omega)| \leq |X_0| \vee n$ for all $t \in [0, \infty)$, $\omega \in \Omega$. Thus, each stopped process X^{τ_n} is bounded. $\square$

We proceed to define the key notion of a *continuous local martingale*.

Definition 4.31. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions. A continuous $\mathbb{F}$-adapted real-valued stochastic process $M = (M_t)_{t \geq 0}$ is called a *continuous local $(P, \mathbb{F})$-martingale* if M_0 is a constant[35] and there is a nondecreasing sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \tau_n = \infty$ P-a.s. such that for each n, the

[33] Note that each X^{τ_n} is an $\mathbb{F}$-adapted continuous process by Remark 2.28.

[34] Recall that each continuous function attains is finite maximum on each compact set. This implies that for each $t \geq 0$ and $\omega \in \Omega$, there is $n \in \mathbb{N}$ with $n > \max_{s \in [0,T]} |X_s|$ and hence $\tau_n(\omega) \geq t$.

[35] This condition can be slightly generalised.

stopped process $M^{\tau_n} = (M_t^{\tau_n})_{t\geq 0}$ is a continuous $(P,\mathbb{F})$-martingale. In this case, we write $M \in \mathcal{M}^c_{\mathrm{loc}}$, and $M \in \mathcal{M}^c_{0,\mathrm{loc}}$ if in addition $M_0 = 0$.

Remark 4.32. (a) If $M \in \mathcal{M}^c_{\mathrm{loc}}$, we can always chose a localising sequence $(\tau_n)_{n\in\mathbb{N}}$ such that the stopped process $M^{\tau_n} = (M_t^{\tau_n})_{t\geq 0}$ is bounded (and hence, in particular, in $\mathcal{H}^{2,c}$ and in $\mathcal{M}^{2,c}$). Indeed, for $n \in \mathbb{N}$, set

$$\sigma_n := \inf\{t \geq 0 : |M_t| \geq n\}.$$

By Proposition 4.30, $(\sigma_n)_{n\in\mathbb{N}}$ is nondecreasing, $\lim_{n\to\infty} \sigma_n = \infty$ P-a.s. and for each n, the stopped processes M^{σ_n} is bounded by $|M_0| \vee n$. Now if $(\tau_n)_{n\in\mathbb{N}}$ is a localising sequence for M, then so is $(\sigma_n \wedge \tau_n)_{n\in\mathbb{N}}$ by the (continuous-time version of the) stopping theorem, and each stopped process $M^{\sigma_n\wedge\tau_n}$ is bounded.

(b) If $M, N \in \mathcal{M}^c_{\mathrm{loc}}$ and $a, b \in \mathbb{R}$, then $aM + bN \in \mathcal{M}^c_{\mathrm{loc}}$. Indeed, if $(\sigma_n)_{n\in\mathbb{N}}$ is a localising sequence for M and $(\tau_n)_{n\in\mathbb{N}}$ is a localising sequence for N, then $(\sigma_n \wedge \tau_n)_{n\in\mathbb{N}}$ is a common localising sequence for M and N and therefore also for $aM + bN$.

We proceed to generalise Theorem 4.20 to continuous local martingales

Theorem 4.33. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions. If $M \in \mathcal{M}^c_{\mathrm{loc}} \cap \mathrm{FV}^c$, then the paths of M are P-a.s.-constant.*

Proof. Let $(\tau_n)_{n\in\mathbb{N}}$ be a localising sequence for M such that both M^{τ_n} and its total variation $|M^{\tau_n}| = |M|^{\tau_n}$ are bounded processes.[36] Since a process is P-a.s. constant if and only if it is locally P-a.s. constant, it suffices to check that each stopped process M^{τ_n} is constant. As $M^{\tau_n} \in \mathcal{M}^{2,c}$, this follows from Theorem 4.20.[37] □

By a similar argument we can extend Theorem 4.22 and Corollary 4.24 to continuous local martingales.

Theorem 4.34. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions. For $M \in \mathcal{M}^c_{\mathrm{loc}}$, there exists a P-a.s. unique process $\langle M \rangle = (\langle M \rangle_t)_{t\geq 0}$ in FV^c_0, called the* quadratic variation of M, *such that*

$$(M_t^2 - \langle M \rangle_t)_{t\geq 0} \text{ is in } \mathcal{M}^c_{\mathrm{loc}}. \tag{4.11}$$

It is nondecreasing and for any $t > 0$ and any sequence of partitions $(\Pi_n)_{n\in\mathbb{N}}$ of $[0,t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$,

$$\sum_{t_i \in \Pi_n \setminus \{t\}} (M_{t_{i+1}} - M_{t_i})^2 \xrightarrow{P} \langle M \rangle_t.$$

[36] This is possible by an extension of Remark 4.32 by setting $\sigma_n := \inf\{t \geq 0 : |M_t| \geq n \text{ or } |M|_t \geq n\}$ and using that $|M|$ has continuous paths by Remark 4.13.

[37] Note that the variation $|M^{\tau_n}|$ is square-integrable so that we use Theorem 4.20 only in the form that we proved.

Corollary 4.35. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions. For $M, N \in \mathcal{M}^c_{\text{loc}}$, there exists a P-a.s. unique process $\langle M, N\rangle = (\langle M, N\rangle_t)_{t\geq 0}$ in FV^c_0, called the* covariation of M and N, *such that*

$$(M_t N_t - \langle M, N\rangle_t)_{t\geq 0} \text{ is in } \mathcal{M}^c_{\text{loc}}. \tag{4.12}$$

Moreover, for any $t > 0$ and any sequence of partitions $(\Pi_n)_{n\in\mathbb{N}}$ of $[0, t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$,

$$\sum_{t_i \in \Pi_n \setminus \{t\}} (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}) \xrightarrow{P} \langle M, N\rangle_t.$$

Remark 4.36. It follows from the defining property (4.11) of the covariation that $\langle \cdot, \cdot \rangle$

- is *symmetric*, i.e., $\langle M, N\rangle = \langle N, M\rangle$ for $M, N \in \mathcal{M}^c_{\text{loc}}$.
- is *bilinear*, i.e., $\langle aM^{(1)} + bM^{(2)}, N\rangle = a\langle M^{(1)}, N\rangle + b\langle M^{(2)}, N\rangle$ and $\langle M, aN^{(1)} + bN^{(2)}\rangle = a\langle M, N^{(1)}\rangle + b\langle M, N^{(2)}\rangle$ for $M, M^{(1)}, M^{(2)}, N, N^{(1)}, N^{(2)} \in \mathcal{M}^c_{\text{loc}}$ and $a, b \in \mathbb{R}$.
- *commutes with stopping*, i.e., $\langle M, N\rangle^\tau = \langle M, N^\tau\rangle = \langle M^\tau, N\rangle = \langle M^\tau, N^\tau\rangle$ for any stopping time τ and $M, N \in \mathcal{M}^c_{\text{loc}}$.

Exercise 4.37. Prove that the covariation $\langle \cdot, \cdot \rangle$

(a) ☆☆☆ is symmetric.

(b) ☆☆☆ is bilinear.

(c) ★☆☆ satisfies $\langle M, N\rangle^\tau = \langle M^\tau, N^\tau\rangle$.

(d) ★★☆ satisfies $\langle M^\tau, N\rangle = \langle M^\tau, N^\tau\rangle$. (Hint: Using a localising argument, show that $M^\tau(N - N^\tau)$ is a continuous local martingale.)

The following result is an extension of Proposition 4.26 to continuous local martingales. Its proof is left as an exercise.

Proposition 4.38. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M = (M_t)_{t\geq 0}$ and $N = (N_t)_{t\geq 0}$ independent continuous local martingales. Then*

$$\langle M, N\rangle = 0 \;\; P\text{-a.s.}$$

Exercise 4.39. ☆☆☆ Prove Proposition 4.38. (*Hint:* Use Proposition 4.26 and the fact that the covariation commutes with stopping.)

We end this section by providing a very useful criterion to decide when a continuous local martingale is in fact a square-integrable martingale or an L^2-bounded martingale. Its proof is left as an exercise.

Proposition 4.40. *Let* $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ *be a filtered probability space satisfying the usual conditions and* $M \in \mathcal{M}^c_{\mathrm{loc}}$*. Set* $\langle M\rangle_\infty := \lim_{t\to\infty}\langle M\rangle_t$*, which is well defined and valued in* $[0,\infty]$*. Then:*

(a) $M \in \mathcal{M}^{2,c}$ *if and only if* $E[\langle M\rangle_t] < \infty$ *for all* $t \geq 0$*. Moreover, in this case, we have the identity*

$$E\left[M_t^2\right] = M_0^2 + E[\langle M\rangle_t], \quad t \geq 0.$$

(b) $M \in \mathcal{H}^{2,c}$ *if and only if* $E[\langle M\rangle_\infty] < \infty$*. Moreover, in this case, we have the identity*

$$E\left[M_\infty^2\right] = M_0^2 + E[\langle M\rangle_\infty].$$

Warning: If $M \in \mathcal{M}^c_{\mathrm{loc}}$ and $E\left[M_t^2\right] < \infty$ for all $t \geq 0$, it does *not* follow that $M \in \mathcal{M}^{2,c}$ because M may fail to be a martingale.

Exercise 4.41. ★☆☆ Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Show that W is not in $\mathcal{H}^{2,c}$.

5 Theory of stochastic integration

In this (rather technical) chapter we develop the theory of stochastic integration of predictable processes with respect to continuous semimartingales, which is a continuous-time analogue to the discrete stochastic integral from Section 2.3.

5.1 Predictable processes in continuous time

We first introduce a continuous-time notion of *predictable processes*, which will serve as integrands. To this end, we first have to take a step back and revisit our notion of a stochastic process.

So far, we have looked at a real-valued stochastic process $X = (X_t)_{t\geq 0}$ on some measurable space $(\Omega, \mathcal{F})$ either as a family of random variables (each X_t is a real-valued random variable) or as a random function (for each ω, $t \mapsto X_t(\omega)$ is a real-valued function). There is another, third way, to look at stochastic processes: We may also understand it as a measurable map $X : \Omega \times [0,\infty) \to \mathbb{R}$, $(\omega, t) \mapsto X_t(\omega)$ on the product space $\overline{\Omega} := \Omega \times [0,\infty)$. This understanding lies at the heart of modern stochastic analysis.

The product space $\overline{\Omega} := \Omega \times [0,\infty)$ can be endowed with different σ-algebras. A first simple choice is to use the product σ-algebra.

Definition 5.1. Let $(\Omega, \mathcal{F})$ be a measurable space. A stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F})$ is called *product measurable* if X (understood as a map from $\overline{\Omega}$ to $\mathbb{R}$) is $\mathcal{F} \otimes \mathcal{B}_{[0,\infty)}$-measurable.

While the notion of product measurability is straightforward it is of very limited use since it does not take into account the flow of information.

Exercise 5.2. ★☆☆ Give an example of a stochastic process $X = (X_t)_{t\geq 0}$ on a filtered measurable space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0})$ that is product measurable but not $\mathbb{F}$-adapted.

We proceed to define predictable processes in continuous time by defining another σ-algebra on $\overline{\Omega}$, the so-called *predictable σ-algebra $\mathcal{P}$*.

Definition 5.3. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0})$ be a filtered measurable space.

(a) An stochastic process $H = (H_t)_{t\geq 0}$ on $(\Omega, \mathcal{F})$ is called *bounded elementary* for $\mathbb{F}$, if it can be written as

$$H_t(\omega) = \sum_{i=1}^{n} h_{t_{i-1}}(\omega) \mathbf{1}_{(t_{i-1}, t_i]}(t), \quad (\omega, t) \in \overline{\Omega},$$

where $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \ldots < t_n < \infty$ and each $h_{t_{i-1}}$ is $\mathcal{F}_{t_{i-1}}$ measurable and bounded. We denote the collection of all bounded elementary processes for $\mathbb{F}$ by $\mathbf{b}\mathcal{E}$.

(b) The *predictable σ-algebra* on $\overline{\Omega}$ is defined by

$$\mathcal{P} := \sigma(\mathbf{b}\mathcal{E}),$$

i.e., it is the smallest σ-algebra on $\overline{\Omega}$ for which all bounded elementary processes for $\mathbb{F}$ are measurable.

(c) A stochastic process $H = (H_t)_{t\geq 0}$ on $(\Omega, \mathcal{F})$ is called *predictable* for $\mathbb{F}$ if H (understood as a map from $\overline{\Omega}$ to $\mathbb{R}$) is $\mathcal{P}$-measurable.

Exercise 5.4. ★☆☆ Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0})$ be a filtered measurable space and $H \in \mathbf{b}\mathcal{E}$. Show that H is $\mathbb{F}$-adapted, product measurable and has left-continuous paths on $(0, \infty)$.

Remark 5.5. Using the result from Exercise 5.4 one can show that every predictable process is $\mathbb{F}$-adapted and product measurable. The converses are false.

A bounded elementary process is a natural generalisation of a bounded predictable process in discrete time. General predictable processes can be thought of as processes that can be "approximated" by bounded elementary processes. The following result shows that the class of predictable processes is quite large.[38]

Proposition 5.6. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0})$ be a filtered measurable space. Any $\mathbb{F}$-adapted process $H = (H_t)_{t\geq 0}$ with left-continuous paths on $(0, \infty)$ and H_0 a constant is $\mathbb{F}$-predictable.*

Proof. We need to show that H is $\mathcal{P}$-measurable. Since constant processes are trivially $\mathcal{P}$-measurable, we may assume without loss of generality that $H_0 = 0$.

For $n \in \mathbb{N}$, define the process $H^{(n)} = (H^{(n)}_t)_{t\geq 0}$ by

$$H^{(n)}_t(\omega) = \sum_{i=1}^{n2^n} k^{(n)}(H_{(i-1)2^{-n}}(\omega))\mathbf{1}_{((i-1)2^{-n}, i2^{-n}]}(t), \quad (\omega, t) \in \overline{\Omega},$$

where the cut-off function $k^{(n)} : \mathbb{R} \to [-n, n]$ is defined by

$$k^{(n)}(x) = \begin{cases} n & \text{if } x > n, \\ x & \text{if } x \in [-n, n] \\ -n & \text{if } x < -n. \end{cases}$$

Then each $H^{(n)}$ is bounded elementary and hence $\mathcal{P}$-measurable. It follows from left-continuity of H, that $H = \lim_{n\to\infty} H^{(n)}$. Since the limit of $\mathcal{P}$-measurable functions is $\mathcal{P}$-measurable, it follows that H is $\mathcal{P}$-measurable. □

[38] In fact, the class of predictable processes is substantially larger than the class of adapted left-continuous processes because it also contains limits of left-continuous adapted processes (which need not be left-continuous), limits of those processes, etc.

Remark 5.7. (a) A stochastic process $H = (H_t)_{t\geq 0}$ on a filtered measurable space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F})_{t\geq 0})$ is called *progressively measurable* for $\mathbb{F}$ if for each $t > 0$, H restricted to $[0, t]$ (understood as a map from $\Omega \times [0, t] \to \mathbb{R}$) is $\mathcal{F}_t \otimes \mathcal{B}_{[0,t]}$-measurable.

(b) It follows from the definition of the product σ-algebra that every progressively measurable process is $\mathbb{F}$-adapted. The converse is false. Moreover, every progressively measurable process is product measurable. Again the converse is false.

(c) One can show that every predictable process is progressively measurable. The converse is false.

(d) In several textbooks, stochastic integration is defined with progressively measurable (as opposed to predictable) processes as integrands. However, there is no gain in doing this, and unlike for predictable processes, there is no economic underpinning of the concept and the definition does not generalise to the case of discontinuous integrators.[39]

5.2 Integral with respect to a continuous finite variation process

We begin our discussion of the stochastic integral by considering as integrators continuous processes of finite variation. It turns out that in this case the integral can be defined *pathwise*, i.e., by fixing ω. To this end, we first need to recall some further properties of finite variation functions and recall the definition of the Lebesgue-Stieltjes integral.

The following result shows that any continuous finite variation function can be decomposed into two nondecreasing continuous functions. The proof is left as an exercise.

Proposition 5.8. *A continuous function $F : [0, \infty) \to \mathbb{R}$ is of finite variation if and only if it can be written as the difference of two nondecreasing continuous functions. Moreover, in this case, define the functions $F^\uparrow, F^\downarrow : [0, \infty) \to [0, \infty)$ by*

$$F^\uparrow(t) = \frac{|F|(t) + (F(t) - F(0))}{2},$$
$$F^\downarrow(t) = \frac{|F|(t) - (F(t) - F(0))}{2}.$$

Then $F^\uparrow$ and $F^\downarrow$ are nondecreasing continuous function, null at zero (i.e., $F^\uparrow(0) = 0$ and $F^\downarrow(0) = 0$) and satisfy

$$F(t) = F(0) + F^\uparrow(t) - F^\downarrow(t), \quad t \geq 0,$$
$$|F|(t) = F^\uparrow(t) + F^\downarrow(t), \quad t \geq 0.$$

We proceed to define the *Lebesgue–Stieltjes integral.*

[39] In the following sections, the word "predictable" can be replaced by "progressively measurable" in every single instant with no change in the mathematics whatsoever.

Definition 5.9. Let $F : [0, \infty) \to \mathbb{R}$ be a continuous function of finite variation and $H : [0, \infty) \to \mathbb{R}$ a measurable function. Fix $t > 0$. Define the finite measures $\mu_t^{\uparrow}$, $\mu_t^{\downarrow}$ and $|\mu|_t$ on $([0, t], \mathcal{B}_{[0,t]})$ by

$$\begin{aligned}
\mu_t^{\uparrow}([0, s]) &:= F^{\uparrow}(s), \quad s \leq t, \\
\mu_t^{\downarrow}([0, s]) &:= F^{\downarrow}(s), \quad s \leq t, \\
|\mu|_t([0, s]) &:= |F|(s), \quad s \leq t.
\end{aligned}$$

Then the function H is said to be *F-integrable on* $[0, t]$ if

$$\int_0^t |H(u)| \, \mathrm{d}|F|(u) := \int_{[0,t]} |H(u)| \, |\mu|_t(\mathrm{d}u) < \infty.$$

In this case, the *Lebesgue–Stieltjes integral* of H with respect to F on $[0, t]$ is defined by

$$\begin{aligned}
\int_0^t H(u) \, \mathrm{d}F(u) &:= \int_0^t H(u) \, \mathrm{d}F^{\uparrow}(u) - \int_0^t H(u) \, \mathrm{d}F^{\downarrow}(u) \\
&:= \int_{[0,t]} H(u) \, \mu_t^{\uparrow}(\mathrm{d}u) - \int_{[0,t]} H(u) \mu_t^{\downarrow}(\mathrm{d}u).
\end{aligned}$$

Remark 5.10. (a) If $F : [0, \infty) \to \mathbb{R}$ is a continuous function of finite variation and $H : [0, \infty) \to \mathbb{R}$ a measurable function that is F-integrable for all $t > 0$, then H is said to be F-integrable. Moreover, in this case, one can show that the function $t \mapsto \int_0^t H(u) \, \mathrm{d}F(u)$ is continuous, of finite variation, and null at 0.

(b) If $F : [0, \infty) \to \mathbb{R}$ is continuously differentiable and $H : [0, \infty) \to \mathbb{R}$ a measurable function, then it follows from Example 4.14 that H is F-integrable on $[0, t]$ if and only if

$$\int_0^t |H(u)||F'(u)| \, \mathrm{d}u < \infty.$$

Moreover, in this case,

$$\int_0^t H(u) \, \mathrm{d}F(u) = \int_0^t H(u) F'(u) \, \mathrm{d}u.$$

Exercise 5.11. ★☆☆ Let $F : [0, \infty) \to \mathbb{R}$ be given by $F(x) = x^2$. Show that F is F-integrable and calculate

$$\int_0^t F(u) \, \mathrm{d}F(u), \quad t \geq 0.$$

We proceed to define the integral of a predictable process H with respect to a continuous finite variation process. This can be done *pathwise*, i.e., by fixing ω.

Definition 5.12. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions, $A \in \mathrm{FV}^c$ and $H = (H_t)_{t \geq 0}$ an $\mathbb{F}$-predictable process.[40] Then H is said to be

[40] By Remark 5.5, this implies that H is product measurable. In particular for each $\omega \in \Omega$, the map

A-integrable if there is a P-null set $N \in \mathcal{F}$ such that for all $\omega \in \Omega \setminus N$

$$\int_0^t |H_u(\omega)| \,\mathrm{d}|A|_u(\omega) < \infty, \quad \text{for all } t > 0.$$

In this case, we define the integral $H \bullet A = ((H \bullet A)_t)_{t \geq 0}$ by

$$(H \bullet A)_t(\omega) := \begin{cases} \int_0^t H_u(\omega) \,\mathrm{d}A_u(\omega) & \text{if } \omega \in \Omega \setminus N, \\ 0 & \text{if } \omega \in N. \end{cases}$$

We denote the collection of all A-integrable predictable processes by $L(A)$.

Remark 5.13. (a) We also write $\int H \,\mathrm{d}A = (\int_0^t H_u \,\mathrm{d}A_u)_{t \geq 0}$ for $H \bullet A$.

(b) If $H = \sum_{i=1}^n h_{t_{i-1}} \mathbf{1}_{(t_{i-1}, t_i]} \in \mathbf{b}\mathcal{E}$, using the definition of the Lebesgue–Stieltjes integral, it is not difficult to check that $H \in L(A)$ and

$$(H \bullet A)_t = \sum_{i=1}^n h_{t_{i-1}} (A_{t_i \wedge t} - A_{t_{i-1} \wedge t}), \quad t \geq 0.$$

This shows that Definition 5.12 is a natural generalisation of Definition 2.13. Moreover, it is straightforward to check that $H \bullet A \in \mathrm{FV}_0^c$ in this case.

(c) Using part (b), one can show that $H \bullet A \in \mathrm{FV}_0^c$ also for general $H \in L(A)$. Here, the difficult part is to check that $H \bullet A$ is indeed $\mathbb{F}$-adapted.

One key example of a continuous finite variation process is the covariation $\langle M, N \rangle$ of two continuous local martingales $M, N \in \mathcal{M}_{\mathrm{loc}}^c$. In this respect, the following inequality, which is an advanced version of the Cauchy–Schwarz inequality, is very useful. For a proof, see [11, Proposition IV.(1.15)].

Proposition 5.14 (Kunita-Watanabe-inequality)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}), P)$ be a filtered probability space satisfying the usual conditions, $M, N \in \mathcal{M}_{\mathrm{loc}}^c$ and G, H be predictable processes with $G^2 \in L(\langle M \rangle)$ and $H^2 \in L(\langle N \rangle)$. Then $GH \in L(\langle M, N \rangle)$. Moreover, P-a.s. for all $t \geq 0$, we have the inequality*

$$\int_0^t |G_s||H_s| \mathrm{d}|\langle M, N \rangle|_s \leq \sqrt{\int_0^t G_s^2 \,\mathrm{d}\langle M \rangle_s} \sqrt{\int_0^t H_s^2 \,\mathrm{d}\langle N \rangle_s}.$$

Exercise 5.15. ★☆☆ Let $M, N \in \mathcal{M}_{\mathrm{loc}}^c$. Show that for each $t \geq 0$,

$$E\left[|\langle M, N \rangle_t|\right]^2 \leq E\left[\langle M \rangle_t\right] E\left[\langle N \rangle_t\right]$$

(*Hint:* Combine the Kunita-Watanabe-inequality with the Cauchy–Schwarz inequality for

---

$t \mapsto H_t(\omega)$ is Borel-measurable.

expectations.)

We conclude this section by stating that the integral with respect to a finite variation process is linear (both in the integrand and in the integrator) and associative. This follows from the respective properties of the Lebesgue-Stieltjes integral.

Theorem 5.16. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}), P)$ be filtered probability space satisfying the usual conditions and $A, B \in \mathrm{FV}^c$.*

(a) *If $G, H \in L(A)$ and $a, b \in \mathbb{R}$, then $aG + bH \in L(A)$ and*

$$(aG + bH) \bullet A = a(G \bullet A) + b(H \bullet A).$$

(b) *If $G \in L(A)$, $G \in L(B)$ and $a, b \in \mathbb{R}$, then $G \in L(aA + bB)$ and*

$$G \bullet (aA + bB) = a(G \bullet A) + b(G \bullet B).$$

(c) *If $H \in L(A)$ then $G \in L(H \bullet A)$ if and only if $GH \in L(A)$. Moreover, in this case*

$$G \bullet (H \bullet A) = (GH) \bullet A.$$

5.3 Stochastic integral with respect to an L^2-bounded martingale

Our goal is now to define a stochastic integral of a predictable process with respect to a continuous local martingale. Unlike in the case of finite variation processes this *cannot* be done pathwise for general predictable processes H.[41] We proceed in four steps. For the first two steps, we will assume that the inegrator is an L^2-bounded martingale.

First, we define the stochastic integral of a bounded elementary process with respect to an L^2-bounded continuous martingale. This can be done pathwise and is straightforward.

Definition 5.17. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions. Then for $M \in \mathcal{H}^{2,c}$ and $H = \sum_{i=1}^n h_{t_{i-1}} \mathbf{1}_{(t_{i-1}, t_i]} \in \mathbf{b}\mathcal{E}$, the *stochastic integral of H with respect to M* is the process $H \bullet M = ((H \bullet M)_t)_{t\geq 0}$ defined by

$$(H \bullet M)_t := \sum_{i=1}^n h_{t_{i-1}} (M_{t_i \wedge t} - M_{t_{i-1} \wedge t}), \quad t \geq 0. \tag{5.1}$$

The following result list some key properties of the stochastic integral defined above.

Theorem 5.18. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions, $M \in \mathcal{H}^{2,c}$, and $H = \sum_{i=1}^n h_{t_{i-1}} \mathbf{1}_{(t_{i-1}, t_i]} \in \mathbf{b}\mathcal{E}$. Then:*

[41] See [10, Section I.8] for a more detailed explanation and rigorous proof of this fact.

(a) $H \bullet M \in \mathcal{H}_0^{2,c}$.

(b) $\langle H \bullet M \rangle = \int_0^\cdot H_u^2 \, \mathrm{d}\langle M \rangle_u$ *P-a.s. and*

$$E\left[(H \bullet M)_t^2\right] = E\left[\int_0^t H_u^2 \, \mathrm{d}\langle M \rangle_u\right], \quad t \in [0, \infty].$$

(c) *For each* $N \in \mathcal{H}_0^{2,c}$,

$$\langle H \bullet M, N \rangle = \int_0^\cdot H_u \, \mathrm{d}\langle M, N \rangle_u \ \ P\text{-a.s.}$$

Proof. (a) This is left as an exercise.

(b) Let $N = H \bullet M$, which is in $\mathcal{H}_0^{2,c}$ by part (a). Then by part (c), symmetry of the covariation process and associativity of the Lebesgue-Stieltjes integral (Theorem 5.16(c)),

$$\begin{aligned} \langle H \bullet M, N \rangle &= \int_0^\cdot H_u \, \mathrm{d}\langle M, N \rangle_u = \int_0^\cdot H_u \, \mathrm{d}\langle N, M \rangle_u \\ &= \int_0^\cdot H_u \, \mathrm{d}\langle H \bullet M, M \rangle_u = \int_0^\cdot H_u \, \mathrm{d}\left(\int_0^\cdot H_s \, \mathrm{d}\langle M \rangle_s\right)_u \\ &= \int_0^\cdot H_u^2 \, \mathrm{d}\langle M \rangle_s. \end{aligned} \tag{5.2}$$

The additional claim now follows from Proposition 4.40, noting that $H \bullet M \in \mathcal{H}_0^{2,c}$.

(c) Let $N \in \mathcal{H}_0^{2,c}$. For convenience set $L := H \bullet M$. By the defining property of the covariation 4.8, it suffices to show that $LN - \int_0^\cdot H_u \, \mathrm{d}\langle M, N \rangle_u$ is a martingale.

Adaptedness follows from adaptedness of M, N, and L (this uses part (a)) and the fact that $H \bullet \langle M, N \rangle \in FV_0^c$ by Remark 5.13(b).

To check integrability, fix $t \geq 0$. Then $N_t, L_t \in \mathcal{L}^2(P)$ by the fact that L, N are square-integrable and hence $N_t L_t \in \mathcal{L}^1(P)$ by Hölder's inequality (Theorem A.62). Moreover, since H is bounded, there exists a constant $C > 0$ such that P-a.s. $|H_t| \leq C$ for all $t \geq 0$. Hence, by monotonicity and definition of the Lebesgue-Stieltjes integral, Exercise 5.15 and the fact that $E\left[\langle M \rangle_t\right] < \infty$ and $E\left[\langle N \rangle_t\right] < \infty$ by the fact that M, N are square-integrable,

$$\begin{aligned} E\left[\left|\int_0^t H_u \, \mathrm{d}\langle M, N \rangle_u\right|\right] &\leq E\left[\int_0^t |H_u| \, \mathrm{d}|\langle M, N \rangle|_u\right] \leq CE\left[|\langle M, N \rangle|_u\right] \\ &\leq CE\left[\langle M \rangle_t\right]^{\frac{1}{2}} E\left[\langle M \rangle_t\right]^{\frac{1}{2}} < \infty. \end{aligned}$$

Since the sum of integrable random variables is integrable, it follows that $N_t L_t - \int_0^t H_u \, \mathrm{d}\langle M, N \rangle_u$ is integrable.

For the martingale property, let $s \leq t$. To simplify the presentation, we only consider the case $s = t_{i-1}$ and $t = t_i$ for $i \in \{1, \dots, n\}$. Using that M and N are martingales, we obtain

$$\begin{aligned} E\left[(M_t - M_s)(N_t - N_s) \,|\, \mathcal{F}_s\right] &= E\left[M_t N_t - M_t N_s \,|\, \mathcal{F}_s\right] - E\left[M_s(N_t - N_s) \,|\, \mathcal{F}_s\right] \\ &= E\left[M_t N_t - M_s N_s \,|\, \mathcal{F}_s\right] \quad P\text{-a.s.} \end{aligned} \tag{5.3}$$

Next, using that $MN - \langle M, N\rangle$ is a martingale, we obtain after rearrangement,

$$E\left[M_t N_t - M_s N_s \,|\, \mathcal{F}_s\right] = E\left[\langle M, N\rangle_t - \langle M, N\rangle_s \,|\, \mathcal{F}_s\right] \quad P\text{-a.s.} \tag{5.4}$$

Now using (5.3) both for the pair L and N and the pair M and N, (5.4), the facts that $L_t - L_s = h_s(M_t - M_s)$, h_s is $\mathcal{F}_s$-measurable and $h_s\left(\langle M, N\rangle_t - \langle M, N\rangle_s\right) = \int_s^t H_u \,\mathrm{d}\langle M, N\rangle_u$, we obtain

$$\begin{aligned} E\left[L_t N_t - L_s N_s \,|\, \mathcal{F}_s\right] &= E\left[(L_t - L_s)(N_t - N_s) \,|\, \mathcal{F}_s\right] = E\left[h_s(M_t - M_s)(N_t - N_s) \,|\, \mathcal{F}_s\right] \\ &= h_s E\left[(M_t - M_s)(N_t - N_s) \,|\, \mathcal{F}_s\right] = h_s E\left[\langle M, N\rangle_t - \langle M, N\rangle_s \,|\, \mathcal{F}_s\right] \\ &= E\left[h_s\left(\langle M, N\rangle_t - \langle M, N\rangle_s\right) \,|\, \mathcal{F}_s\right] = E\left[\int_s^t H_u \,\mathrm{d}\langle M, N\rangle_u \,\middle|\, \mathcal{F}_s\right] \\ &= E\left[\int_0^t H_u \,\mathrm{d}\langle M, N\rangle_u - \int_0^s H_u \,\mathrm{d}\langle M, N\rangle_u \,\middle|\, \mathcal{F}_s\right] \end{aligned}$$

Rearranging establishes the claim. □

Exercise 5.19. ★☆☆ Prove Theorem 5.18(a). (*Hint:* Argue as in the proof of Theorem 2.15(a).)

Next, we want to extend the stochastic integral with respect to an L^2-bounded martingale to a larger class of predictable processes. As already indicated above, this *cannot* be done "naively" by approximating a predictable processes by a sequence of bounded elementary processes (which is not always possible), then fixing ω, and then taking limits in (5.1).[42] Instead, we take the properties of the stochastic integral for bounded elementary integrands in Theorem 5.18 as a starting point and try to define the stochastic integral via its *properties*.

Definition 5.20. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{H}^{2,c}$. Denote by $L^2(M)$ the collection of all predictable processes in H such that

$$E\left[\int_0^\infty H_u^2 \,\mathrm{d}\langle M\rangle_u\right] < \infty.$$

Note that $H \in L^2(M)$ for any $H \in \mathbf{b}\mathcal{E}$ and $M \in \mathcal{H}^{2,c}$.

[42]Note, however, that a Riemann-type approximation is possible for continuous integrands if the limit is taken *in probability*; cf. Proposition 6.1 below.

Theorem 5.21. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{H}^{2,c}$. Then for each $H \in L^2(M)$, there exists a P-a.s. unique $L \in \mathcal{H}_0^{2,c}$ such that*

$$\langle L, N\rangle = \int_0^\cdot H_u \, \mathrm{d}\langle M, N\rangle_u \ \ P\text{-a.s.}, \qquad \text{for each } N \in \mathcal{H}^{2,c}. \tag{5.5}$$

We call L the stochastic integral of H with respect to M *and write $L = H \bullet M$ or $L = \int H \, \mathrm{d}M$. Moreover,*

$$\langle H \bullet M\rangle = \int_0^\cdot H_u^2 \, \mathrm{d}\langle M\rangle_u, \tag{5.6}$$

$$E\left[\int_0^t H_u^2 \, \mathrm{d}\langle M\rangle_u\right] = E\left[(H \bullet M)_t^2\right], \quad t \in [0, \infty]. \tag{5.7}$$

Remark 5.22. Property (5.7) is often referred to as *Itô isometry* of the stochastic integral.

Proof of Theorem 5.21. We only argue uniqueness of L and show (5.6) and (5.7); a proof of existence can be found in [11, Theorem IV.(2.2)].

Suppose there are $L^{(1)}, L^{(2)} \in \mathcal{H}_0^{2,c}$ satisfying (5.5). Then

$$\langle L^{(1)} - L^{(2)}, N\rangle = 0 \ \ P\text{-a.s.}, \qquad \text{for each } N \in \mathcal{H}^{2,c}.$$

Considering $N = L^{(1)} - L^{(2)}$, we obtain

$$\langle L^{(1)} - L^{(2)}\rangle = 0 \ \ P\text{-a.s.}$$

By the defining property of quadratic variation, this implies that $(L^{(1)} - L^{(2)})^2$ is a martingale. Since $L^{(1)} - L^{(2)}$ is also an (L^2-bounded) martingale, it follows that

$$\begin{aligned} E\left[(L_t^{(1)} - L_t^{(2)})^2\right] &= E\left[(L_0^{(1)} - L_0^{(2)})^2\right] = 0 \\ &= E\left[L_0^{(1)} - L_0^{(2)}\right]^2 = E\left[L_t^{(1)} - L_t^{(2)}\right]^2, \quad t \geq 0 \end{aligned} \tag{5.8}$$

A rearrangement shows that $\mathrm{Var}[L_t^{(1)} - L_t^{(2)}] = 0$ for all $t \geq 0$. This implies that $L_t^{(1)} - L^{(2)}$ is P-a.s. constant, and it follows from (5.8) that this constant is 0. Since the paths of $L^{(1)}$ and $L^{(2)}$ are continuous, this implies that $L^{(1)} = L^{(2)}$ P-a.s.

Next, (5.6), follows by the same calculation as in (5.2). Finally, (5.7) follows from (5.6), the fact that $H \bullet M \in \mathcal{H}_0^{2,c}$ and Proposition 4.40. □

5.4 Stochastic integral with respect to a continuous local martingale

We now come to final two steps in our construction, where we extend the stochastic integral to integrators that are continuous local martingales and also to a wider class of predictable processes.

The third step is surprisingly simple. We only need to extend the definition of $L^2(M)$ from $M \in \mathcal{H}^{2,c}$ to $\mathcal{M}^c_{\mathrm{loc}}$.

Definition 5.23. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{M}^c_{\mathrm{loc}}$. Denote by $L^2(M)$ the collection of all predictable processes in H such that

$$E\left[\int_0^\infty H_u^2 \,\mathrm{d}\langle M\rangle_u\right] < \infty.$$

One can show that Theorem 5.21 carries over *verbatim* to the case $M \in \mathcal{M}^c_{\mathrm{loc}}$. Therefore, we will not repeat it.

Exercise 5.24. ★☆☆ Formulate Theorem 5.21 for the case $M \in \mathcal{M}^c_{\mathrm{loc}}$ and check that the proof carries over.

In preparation for the final step, we proceed to study some properties of the stochastic integral defined so far. To this end, we define for two stopping time $\sigma \leq \tau$, the following *stochastic intervals*:[43]

$$\begin{aligned}
[\![\sigma, \tau]\!] &:= \{(\omega, t) \in \Omega \times [0, \infty) : \sigma(\omega) \leq t \leq \tau(\omega)\}\,, \\
]\!]\sigma, \tau]\!] &:= \{(\omega, t) \in \Omega \times [0, \infty) : \sigma(\omega) < t \leq \tau(\omega)\}\,, \\
[\![\sigma, \tau[\![&:= \{(\omega, t) \in \Omega \times [0, \infty) : \sigma(\omega) \leq t < \tau(\omega)\}\,, \\
]\!]\sigma, \tau[\![&:= \{(\omega, t) \in \Omega \times [0, \infty) : \sigma(\omega) < t < \tau(\omega)\}\,.
\end{aligned}$$

Exercise 5.25. ★☆☆ Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions, $A \in \mathrm{FV}^c$, $H \in L(A)$ and τ a stopping time. Show that $H \in L(A^\tau)$ and $H\mathbf{1}_{[\![0,\tau]\!]} \in L(A)$ as well as

$$(H \bullet A)^\tau = H \bullet (A^\tau) = (H\mathbf{1}_{[\![0,\tau]\!]}) \bullet A \;\; P\text{-a.s.}$$

The next result shows that the stochastic integral is linear (both in the integrand and the integrator), associative and commutes with stopping. Moreover, it states how the quadratic covariation of stochastic integrals can be calculated. This follows from the definition of the stochastic integral, the properties the covariation and the Lebesgue-Stieltjes integral. The details of the proof are left as an exercise.

Theorem 5.26. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M, N \in \mathcal{M}^c_{\mathrm{loc}}$.*

(a) *If $G, H \in L^2(M)$ and $a, b \in \mathbb{R}$, then $aG + bH \in L^2(M)$ and*

$$(aG + bH) \bullet M = a(G \bullet M) + b(H \bullet M) \;\; P\text{-}a.s.$$

[43] Note that all the stochastic intervals are subsets of $\Omega \times [0, \infty)$: even if $\tau(\omega) = \infty$ for some ω, (ω, ∞) is *not* an element of $[\![\sigma, \tau]\!]$.

(b) *If* $G \in L^2(M)$, $G \in L^2(N)$ *and* $a, b \in \mathbb{R}$, *then* $G \in L^2(aM + bN)$ *and*

$$G \bullet (aM + bN) = a(G \bullet M) + b(G \bullet N) \;\; P\text{-}a.s.$$

(c) *If* $H \in L^2(M)$, *then* $G \in L^2(H \bullet M)$ *if and only if* $GH \in L^2(M)$. *Moreover, in this case*

$$G \bullet (H \bullet M) = (GH) \bullet M \;\; P\text{-}a.s.$$

(d) *If* $H \in L^2(M)$ *and* τ *is a stopping time, then* $H \in L^2(M^\tau)$ *and* $H\mathbf{1}_{[\![0,\tau]\!]} \in L^2(M)$, *and*

$$(H \bullet M)^\tau = H \bullet (M^\tau) = (H\mathbf{1}_{[\![0,\tau]\!]}) \bullet M \;\; P\text{-}a.s.$$

(e) *If* $G \in L^2(M)$ *and* $H \in L^2(N)$, *then*

$$\langle G \bullet M, H \bullet N\rangle = (GH) \bullet \langle M, N\rangle \;\; P\text{-}a.s.$$

Exercise 5.27. Prove Theorem 5.26.

(a) ★☆☆ (*Hint*: To show that $aG+bH \in L^2(M)$, use the elementary inequality $(\alpha+\beta)^2 \leq 2\alpha^2 + 2\beta^2$ for $\alpha, \beta \in \mathbb{R}$. Also use Theorem 5.16(a) and the defining property (5.5) of the stochastic integral.)

(b) ★★☆ (*Hint*: To show that $G \in L^2(aM + bN)$, use the Kunita-Watanabe inequality (Proposition 5.14). Also use Theorem 5.16(b) and the defining property (5.5) of the stochastic integral.)

(c) ★☆☆ (*Hint*: Use Theorem 5.16(c) and the defining property (5.5) of the stochastic integral.)

(d) ★☆☆ (*Hint*: Use Remark 4.36, Exercise 5.25, and the defining property (5.5) of the stochastic integral.)

(e) ☆☆☆ (*Hint*: Use the defining property (5.5) of the stochastic integral.)

Finally, we extend the stochastic integral to the case that H is only locally in $L^2(M)$. This crucially uses Theorem 5.26(d).

Definition 5.28. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{M}^c_{\text{loc}}$. Denote by $L^2_{\text{loc}}(M)$ the collection of all predictable processes H for which there is a nondecreasing sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \tau_n = \infty$ P-a.s. such that $H\mathbf{1}_{[\![0,\tau_n]\!]} \in L^2(M)$ for each $n \in \mathbb{N}$, i.e.,

$$E\left[\int_0^{\tau_n} H_u^2 \,\mathrm{d}\langle M\rangle_u\right] < \infty, \quad n \in \mathbb{N}. \tag{5.9}$$

The following result extends Theorem 5.21 (in the version for $M \in \mathcal{M}^c_{\text{loc}}$) from $H \in L^2(M)$ to $H \in L^2_{\text{loc}}(M)$. The proof is left as an exercise.

Theorem 5.29. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{M}^c_{\text{loc}}$. For each $H \in L^2_{\text{loc}}(M)$, there exists a P-a.s. unique $L \in \mathcal{M}^c_{0,\text{loc}}$ such that*

$$\langle L, N \rangle = \int_0^{\cdot} H_u \,\mathrm{d}\langle M, N \rangle_u \;\; P\text{-a.s.}, \qquad \text{for each } N \in \mathcal{M}^c_{0,\text{loc}}$$

We call L the stochastic integral of H with respect to M *and write $L = H \bullet M$ or $L = \int H \,\mathrm{d}M$.*

Remark 5.30. One can show that Theorem 5.26 extends to the case that $L^2(M)$ is replaced by $L^2_{\text{loc}}(M)$, etc.

We proceed to give an equivalent simpler characterisation of the set $L^2_{\text{loc}}(M)$ that we will use in the sequel.

Proposition 5.31. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{M}^c_{\text{loc}}$. Then $H \in L^2_{\text{loc}}(M)$ if and only if*

$$\int_0^t H_u^2 \,\mathrm{d}\langle M \rangle_u < \infty \quad P\text{-a.s. for each } t > 0. \tag{5.10}$$

Proof. Let H be a predictable process. First assume that there exists a nondecreasing sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \tau_n = \infty$ P-a.s. such that (5.9) is satisfied. Then $\int_0^{\tau_n} H_u^2 \,\mathrm{d}\langle M \rangle_u < \infty$ P-a.s. for each $n \in \mathbb{N}$.[44] Fix $t \geq 0$. Then for P-a.e. ω, there exists $n \in \mathbb{N}$ such that $\tau_n(\omega) \geq t$ and $\int_0^{\tau_n} H_u^2(\omega) \,\mathrm{d}\langle M \rangle_u(\omega) < \infty$. Hence $\int_0^t H_u^2(\omega) \,\mathrm{d}\langle M \rangle_u(\omega) < \infty$. Since $t \geq 0$ was arbitrary (and the function $\int_0^t H_u^2 \,\mathrm{d}\langle M \rangle_u$ is continuous), we have (5.10). Conversely, suppose (5.10) is satisfied. For $n \in \mathbb{N}$, define the stopping time

$$\tau_n := \inf\left\{ t \geq 0 : \int_0^t H_u^2 \,\mathrm{d}\langle M \rangle_u \geq n \right\}.$$

Then $(\tau_n)_{n \in \mathbb{N}}$ is nondecreasing and satisfies $\lim_{n \to \infty} \tau_n = \infty$ P-a.s. by (5.10). Moreover, by continuity of $\int_0^{\cdot} H_u^2 \,\mathrm{d}\langle M \rangle_u$, it follows that $\int_0^{\tau_n} H_u^2 \,\mathrm{d}\langle M \rangle_u \leq n$, whence we have (5.9) $\square$

Exercise 5.32. ★☆☆ Show that if $W = (W_t)_{t \geq 0}$ is a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ satisfying the usual conditions, then $H \in L^2_{\text{loc}}(W)$ if and only if

$$\int_0^t H_u^2 \,\mathrm{d}u < \infty \;\; P\text{-a.s. for each } t > 0.$$

If $H \in L^2_{\text{loc}}(M)$, the stochastic integral $H \bullet M$ is in general only a local martingale. One is often interested in the case that it is a (true) martingale. The following result provides an important sufficient criterion.

[44] Recall that for a nonnegative random variable X, $E[X] < \infty$ implies that $X < \infty$ P-a.s.

Proposition 5.33. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions, $M \in \mathcal{M}^c_{\text{loc}}$, and $H \in L^2_{\text{loc}}(M)$. Then $H \bullet M$ is a (square-integrable) martingale if*

$$E\left[\int_0^t H_u^2 \,\mathrm{d}\langle M\rangle_u\right] < \infty \quad \textit{for all } t > 0.$$

Proof. By definition, $H \bullet M \in \mathcal{M}^c_{\text{loc}}$ and $\langle H \bullet M\rangle = \int H^2 \,\mathrm{d}\langle M\rangle$. Now the claim follow from Proposition 4.40. □

5.5 Stochastic integral with respect to a continuous semimartingale

In this section, we extend the stochastic integral to integrators that are continuous semimartingales.

Definition 5.34. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions. An $\mathbb{F}$-adapted continuous process $X = (X_t)_{t\geq 0}$ is called a *continuous* $(P, \mathbb{F})$-*semimartingale* if it can be decomposed as

$$X_t = X_0 + M_t + A_t, \quad t \geq 0, \tag{5.11}$$

where X_0 is a constant[45], $M = (M_t)_{t\geq 0} \in \mathcal{M}^c_{0,\text{loc}}$ and $A = (A_t)_{t\geq 0} \in FV^c_0$.

If there is no danger of confusion, we often drop the qualifier "$(P, \mathbb{F})$".

Exercise 5.35. ★☆☆ Show that the decomposition (5.11) is P-a.s. unique. (*Hint:* Use Theorem 4.33.)

We proceed to define the quadratic variation of a continuous semimartingale X and the covariation of two continuous semimartingales X and Y.

Definition 5.36. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions, $X = (X_t)_{t\geq 0}$ a continuous semimartingale with decomposition $X = X_0 + M + A$ and $Y = (X_t)_{t\geq 0}$ a continuous semimartingale with decomposition $Y = Y_0 + N + B$, where $M, N \in \mathcal{M}^c_{0,\text{loc}}$ and $A, B \in FV^c_0$.

(a) The *quadratic variation* of X is the process $\langle X\rangle = (\langle X\rangle_t)_{t\geq 0}$ defined by

$$\langle X\rangle_t := \langle M\rangle_t, \quad t \geq 0.$$

(b) The *covariation* of X and Y is the process $\langle X, Y\rangle = (\langle X, Y\rangle_t)_{t\geq 0}$ defined by

$$\langle X, Y\rangle_t := \langle M, N\rangle_t, \quad t \geq 0.$$

[45]This assumption is only made for consistency reasons with our definition of local martingales.

Remark 5.37. (a) Definition 5.36 might seem a bit ad hoc. However, it is justified by the fact that one can show that for any $t > 0$ and any sequence of partitions $(\Pi_n)_{n\in\mathbb{N}}$ of $[0,t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$,

$$\sum_{t_i\in\Pi_n\setminus\{t\}} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) \xrightarrow{P} \langle X, Y\rangle_t.$$

(b) If X, Y are semimartingales and X is of finite variation, it follows from Definition 5.36 that $\langle X\rangle = 0$ and $\langle X, Y\rangle = 0$.

(c) It follows from Definition 5.36 and Remark 4.36 that $\langle\cdot,\cdot\rangle$

- is *symmetric*, i.e., $\langle X, Y\rangle = \langle Y, X\rangle$ for continuous semimartingales X, Y.
- is *bilinear*, i.e., $\langle aX^{(1)} + bX^{(2)}, Y\rangle = a\langle X^{(1)}, Y\rangle + b\langle X^{(2)}, Y\rangle$ and $\langle X, aY^{(1)} + bY^{(2)}\rangle = a\langle X, Y^{(1)}\rangle + b\langle X, Y^{(2)}\rangle$ for continuous semimartingales $X, X^{(1)}, X^{(2)}, Y, Y^{(1)}, Y^{(2)}$ and $a, b \in \mathbb{R}$.
- *commutes with stopping*, i.e., $\langle X, Y\rangle^\tau = \langle X, Y^\tau\rangle = \langle X^\tau, Y\rangle = \langle X^\tau, Y^\tau\rangle$ for any stopping time τ and continuous semimartingales X, Y.

Exercise 5.38. ★☆☆ Let $B = (B_t)_{t\geq 0}$ and $W = (W_t)_{t\geq 0}$ be independent Brownian motions on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Define the processes $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ by $X_t = a_X B_t + b_X W_t + c_X t$, $Y_t = a_Y B_t + b_Y W_t + c_Y t$, where $a_X, a_Y, b_X, b_Y, c_X, c_Y \in \mathbb{R}$. Compute $\langle X, Y\rangle$.

We proceed to define the stochastic integral of a predictable process with respect to a continuous semimartingale.

Definition 5.39. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and X a continuous semimartingale with decomposition $X = X_0 + M + A$ for $M \in \mathcal{M}^c_{0,\mathrm{loc}}$ and $A \in FV^c_0$. Denote by $L(X)$ the collection of all predictable processes H that are both in $L^2_{\mathrm{loc}}(M)$ and $L(A)$. Moreover, for $H \in L(X)$, the *stochastic integral of H with respect to X* is denoted by $H \bullet X$ or $\int H \,\mathrm{d}X$ and defined by

$$(H \bullet X)_t := (H \bullet M)_t + (H \bullet A)_t, \quad t \geq 0.$$

Remark 5.40. It follows immediately from the definition and the properties of the (stochastic) integral with respect to a continuous local martingale and a continuous finite variation process that the stochastic integral $H \bullet X$ is again a semimartingale and has the decomposition

$$H \bullet X = H \bullet M + H \bullet A.$$

The next result shows that the stochastic integral with respect to a continuous semimartingale is linear (both in the integrand and in the integrator), associative, and commutes with stopping. Moreover, it states how the quadratic covariation of stochastic integrals can be calculated. It can be proved by combining (the localised version of) Theorem 5.26, Theorem 5.16 and the defining property of the stochastic integral with respect to a continuous local martingale. The details of the proof are left as an exercise.

Theorem 5.41. *Let X and Y be continuous semimartingales on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$ satisfying the usual conditions.*

(a) *If $G, H \in L(X)$ and $a, b \in \mathbb{R}$, then $aG + bH \in L(X)$ and*

$$(aG + bH) \bullet X = a(G \bullet X) + b(H \bullet X) \quad P\text{-a.s.}$$

(b) *If $G \in L(X)$, $G \in L(Y)$ and $a, b \in \mathbb{R}$, then $G \in L(aX + bY)$ and*

$$G \bullet (aX + bY) = a(G \bullet X) + b(G \bullet Y) \quad P\text{-a.s.}$$

(c) *If $H \in L(X)$, then $G \in L(H \bullet X)$ if and only if $GH \in L(X)$. Moreover, in this case*

$$G \bullet (H \bullet X) = (GH) \bullet X \quad P\text{-a.s.}$$

(d) *If $H \in L(X)$ and τ is a stopping time, then $H \in L(X^\tau)$ and $H\mathbf{1}_{[\![0,\tau]\!]} \in L(X)$, and*

$$(H \bullet X)^\tau = H \bullet (X^\tau) = (H\mathbf{1}_{[\![0,\tau]\!]}) \bullet X \quad P\text{-a.s.}$$

(e) *If $G \in L(X)$ and $H \in L(Y)$, then*

$$\langle G \bullet X, H \bullet Y \rangle = (GH) \bullet \langle X, Y \rangle \quad P\text{-a.s.}$$

Exercise 5.42. Prove Theorem 5.41.

(a) ✰☆☆ (*Hint*: Combine Theorem 5.26(a), Remark 5.30 and Theorem 5.16(a).)

(b) ✰☆☆ (*Hint*: Combine Theorem 5.26(b), Remark 5.30 and Theorem 5.16(b).)

(c) ★✰☆ (*Hint*: Combine Theorem 5.26(c), Remark 5.30 and Theorem 5.16(c).)

(d) ★✰☆ (*Hint*: Combine Theorem 5.26(d), Remark 5.30 and Exercise 5.25.)

(e) ✰☆☆ (*Hint*: Combine Theorem 5.26(e), Remark 5.30 and Definition 5.36(b).)

6 Stochastic Calculus

In this chapter, we consider how to "calculate" stochastic integrals in practice using the celebrated Itô's formula. We also study the behaviour of semimartingales under an equivalent change of probability measures and show that Brownian motion has the predictable representation property for its natural (augmented) filtration.

6.1 Itô's formula in one dimension

In this section, we consider how to "calculate" stochastic integrals in practice. In particular, we study the "chain rule" for stochastic integrals. This is known as "Itô's formula" and constitutes the cornerstone for all calculations of stochastic integrals.

We start our discussion by an approximation result of stochastic integrals for continuous adapted integrands; for a proof see [11, Proposition IV.(2.13)].

Proposition 6.1. *Let $X = (X_t)_{t\geq 0}$ be a continuous semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions and $H = (H_t)_{t\geq 0}$ a continuous adapted process with H_0 a constant.*[46] *Then for any $t > 0$ and any sequence of partitions $(\Pi_n)_{n\in\mathbb{N}}$ of $[0,t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$,*

$$\sum_{t_i \in \Pi_n \setminus \{t\}} H_{t_i}(X_{t_{i+1}} - X_{t_i}) \xrightarrow{P} \int_0^t H_u \,\mathrm{d}X_u.$$

We illustrate the above result by an example.

Example 6.2. Let $X = (X_t)_{t\geq 0}$ be a continuous semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Then

$$\int_0^t X_u \,\mathrm{d}X_u = \frac{1}{2}(X_t^2 - X_0^2 - \langle X\rangle_t) \;\; P\text{-a.s. for each } t \geq 0.$$

Indeed, let $t > 0$ and $(\Pi_n)_{n\in\mathbb{N}}$ a sequence of partitions of $[0,t]$ with $\lim_{n\to\infty} |\Pi_n| = 0$. Using the elementary identity $a(b-a) = \frac{1}{2}b^2 - \frac{1}{2}a^2 - \frac{1}{2}(b-a)^2$ for $a, b \in \mathbb{R}$, we obtain

$$\begin{aligned}\sum_{t_i \in \Pi_n \setminus \{t\}} X_{t_i}(X_{t_{i+1}} - X_{t_i}) &= \sum_{t_i \in \Pi_n \setminus \{t\}} \left(\frac{1}{2}X_{t_{i+1}}^2 - \frac{1}{2}X_{t_i}^2 - \frac{1}{2}(X_{t_{i+1}} - X_{t_i})^2\right) \\ &= \frac{1}{2}X_t^2 - \frac{1}{2}X_0^2 - \frac{1}{2}\sum_{t_i \in \Pi_n \setminus \{t\}} (X_{t_{i+1}} - X_{t_i})^2, \quad n \in \mathbb{N}\end{aligned}$$

Now the result follows by taking the limit in probability as $n \to \infty$ and using that the last term converges to $\langle X\rangle_t$ by Remark 5.37(a).

[46] Note that by Remark 5.5(b), H is predictable.

The above example shows that for a semimartingale X and $f(x) := x^2$, we have the formula

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}\langle X\rangle_s \ \ P\text{-a.s. for each } t \geq 0.$$

The following result shows that this formula is true for any twice continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$. This is known as Itô's formula; for a proof, see [11, Theorem IV.(3.3)].

Theorem 6.3 (Itô's formula in one dimension)**.** *Let $X = (X_t)_{t\geq 0}$ be a continuous semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions and $f : \mathbb{R} \to \mathbb{R}$ twice continuously differentiable. Then $f(X) = (f(X_t))_{t\geq 0}$ is again a continuous semimartingale and we have* Itô's formula*:*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}\langle X\rangle_s \ \ P\text{-a.s. for each } t \geq 0.$$

Remark 6.4. (a) Itô's formula is often written more compactly in differential notation as:

$$\mathrm{d}f(X_t) = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)\,\mathrm{d}\langle X\rangle_t.$$

(b) Itô's formula remains true if f is only defined on an open set $U \subset \mathbb{R}$ and X is P-a.s. U-valued.

(c) If X is of finite variation, Itô's formula remains true if f is only once continuously differentiable. Obviously, there is no $\mathrm{d}\langle X\rangle$-term in this case.

Exercise 6.5. ★☆☆ Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Using Itô's formula, calculate the dynamics of $(W_t^3)_{t\geq 0}$.

As an application of Itô's formula, we compute the stochastic exponential of a continuous semimartingale X.

Proposition 6.6. *Let $X = (X_t)_{t\geq 0}$ be a continuous semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Then the stochastic differential equation (SDE)*

$$\mathrm{d}Z_t = Z_t\,\mathrm{d}X_t, \quad Z_0 = 1,$$

has a unique solution Z given by

$$Z_t := \exp\left(X_t - X_0 - \frac{1}{2}\langle X\rangle_t\right), \quad t \geq 0.$$

This solution is called the stochastic exponential of X *and denoted by $\mathcal{E}(X) = (\mathcal{E}(X)_t)_{t\geq 0}$.*

Proof. We only show existence. The proof of uniqueness is left as an exercise. Set $Y :=$

$X - X_0 - \frac{1}{2}\langle X \rangle$. Then Y is again a semimartingale and satisfies

$$\begin{aligned}\langle Y \rangle &= \left\langle X - X_0 - \frac{1}{2}\langle X \rangle \right\rangle = \langle X - X_0 \rangle + 2 \times \left(-\frac{1}{2}\right) \langle X - X_0, \langle X \rangle \rangle + \frac{1}{4}\langle \langle X \rangle \rangle \\ &= \langle X - X_0 \rangle = \langle X \rangle - 2\langle X, X_0 \rangle + \langle X_0, X_0 \rangle = \langle X \rangle.\end{aligned}$$

by bilinearity of the covariation and the fact that $\langle X \rangle$ and the constant process X_0 are of finite variation. Since $Z = \exp(Y)$, Itô's formula gives

$$\begin{aligned}\mathrm{d}Z_t = d\exp(Y_t) &= \exp(Y_t)\,\mathrm{d}Y_t + \frac{1}{2}\exp(Y_t)\,\mathrm{d}\langle Y \rangle_t \\ &= Z_t \left(\mathrm{d}X_t - \frac{1}{2}\,\mathrm{d}\langle X \rangle_t \right) + \frac{1}{2} Z\,\mathrm{d}\langle X \rangle_t = Z_t\,\mathrm{d}X_t.\end{aligned}$$

Moreover, $Z_0 = \exp(Y_0) = \exp(X_0 - X_0 - \langle X \rangle_0) = \exp(0) = 1$. □

Exercise 6.7. ★☆☆ Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. For $\sigma > 0$ and $\mu \in \mathbb{R}$, define the semimartingale $X = (X_t)_{t\geq 0}$ by

$$X_t := \mu t + \sigma W_t, \quad t \geq 0.$$

X is called *arithmetic Brownian motion* with drift μ and volatility σ. Compute the stochastic exponential $\mathcal{E}(X)$. This is called a *geometric Brownian motion* with drift μ and volatility σ. It is at the heart of the Black–Scholes–Samuelson model.

As another application of Itô's formula, we prove one part of the celebrated *Burkholder–Davis–Gundy (BDG) inequalities.* To this end we extend Definition 2.36 to continuous time and set $X^*_t := \sup_{s\in[0,T]} |X_s|$ for a continuous adapted process X.

Theorem 6.8 (Burkholder–Davis–Gundy (BDG) inequalities)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $M \in \mathcal{M}^c_{0,\mathrm{loc}}$. Let $p \geq 1$. Then there exists constants $c_p, C_p \in (0,\infty)$ only depending on p such that for each $t \geq 0$,*

$$c_p E\left[(\langle M \rangle_t)^{\frac{p}{2}}\right] \leq E\left[(M^*_t)^p\right] \leq C_p E\left[(\langle M \rangle_t)^{\frac{p}{2}}\right]. \tag{6.1}$$

Remark 6.9. It is insightful to compare the BDG inquealities with (the continuous times version of) Doob's inequality (Theorem 2.38). While Doob's inequality is very useful, it requires M to be a *true* martingale. It is *false* if M is only a *local* martingale. By contrast, the BDG inequalities only require M to be a *local* martingale and are therefore more powerful.

Proof of Theorem 6.8. We only establish the right inequality in (6.1) for $p > 2$; for the left-inequality and the case that $p \in [1,2]$, see [11, Section IV.4]. So let $p > 2$.

We first assume that both M and $\langle M\rangle$ are bounded. Noting that the function $x \mapsto |x|^p$ is twice continuously differentiable, Itô's formula yields[47]

$$|M_t|^p = p\int_0^t |M_s|^{p-1}\operatorname{sign}(M_s)\,\mathrm{d}M_s + \frac{1}{2}p(p-1)\int_0^t |M_s|^{p-2}\,\mathrm{d}\langle M\rangle_s. \tag{6.2}$$

Since M and $\langle M\rangle$ are bounded, there exist a constant $K \geq 0$ such that $|M_t|(\omega) \leq K$ and $\langle M\rangle_t(\omega) \leq K$ for all $\omega \in \Omega$, $t \geq 0$. Hence,

$$\begin{aligned} E\left[\int_0^t (|M_s|^{p-1}\operatorname{sign}(M_s))^2\,\mathrm{d}\langle M\rangle_s\right] &= E\left[\int_0^t |M_s|^{2p-2}\,\mathrm{d}\langle M\rangle_s\right] \leq K^{2p-2}E\left[\langle M\rangle_t\right] \\ &\leq K^{2p-1} < \infty. \end{aligned}$$

By Proposition 5.33, this implies that $\int |M|^{p-1}\operatorname{sign}(M)\,\mathrm{d}M$ is a true martingale. Thus, taking expectations in (6.2) yields

$$E\left[|M_t|^p\right] = \frac{1}{2}p(p-1)E\left[\int_0^t |M_s|^{p-2}\,\mathrm{d}\langle M\rangle s\right]. \tag{6.3}$$

Now combining (6.3) with (the continuous-time version of) Doob's L^p-inequality (Theorem 2.38) and using monotonicity of the Lebesgue-Stieltjes-integral, we obtain

$$\begin{aligned} E\left[(M_t^*)^p\right] &\leq \left(\frac{p}{p-1}\right)^p E\left[|M_t|^p\right] = \frac{1}{2}p(p-1)\left(\frac{p}{p-1}\right)^p E\left[\int_0^t |M_s|^{p-2}\,\mathrm{d}\langle M\rangle s\right] \\ &\leq \frac{1}{2}p(p-1)\left(\frac{p}{p-1}\right)^p E\left[\int_0^t (M_t^*)^{p-2}\,\mathrm{d}\langle M\rangle_s\right] \\ &= \frac{1}{2}p(p-1)\left(\frac{p}{p-1}\right)^p E\left[(M_t^*)^{p-2}\langle M\rangle_t\right], \end{aligned} \tag{6.4}$$

Now using that $p-2 = p(1-\frac{2}{p})$ and applying Hölder's inequality (Theorem A.62) with $r = \frac{1}{1-\frac{2}{p}}$ and $s = \frac{p}{2}$, we obtain

$$E\left[(M_t^*)^{p-2}\langle M\rangle_t\right] \leq E\left[(M_t^*)^p\right]^{1-\frac{2}{p}} E\left[(\langle M\rangle_t)^{\frac{p}{2}}\right]^{\frac{2}{p}}$$

Combining this with (6.4) and dividing by $E\left[(M_t^*)^p\right]^{1-\frac{2}{p}}$ yields

$$E\left[(M_t^*)^p\right]^{\frac{2}{p}} \leq \frac{1}{2}p(p-1)\left(\frac{p}{p-1}\right)^p E\left[(\langle M\rangle_t)^{\frac{p}{2}}\right]^{\frac{2}{p}}$$

Now the claim follows by taking $\frac{p}{2}$ powers and setting $C_p := (\frac{1}{2}p(p-1)(\frac{p}{p-1})^p)^{\frac{p}{2}}$.

Now let $M \in \mathcal{M}^c_{0,\mathrm{loc}}$ be general. Let $(\tau_n)_{n\in\mathbb{N}}$ be a localising sequence such that both M^{τ_n}

[47] Here the sign function is defined as $\operatorname{sign}(x) = 1$ if $x \geq 0$ and $\operatorname{sign}(x) = -1$ if $x < 0$.

and $\langle M^{\tau_n}\rangle$ are bounded for each n. Then by the above for each, $n \in \mathbb{N}$

$$E\left[\left(\sup_{s\in[0,t\wedge\tau_n]} |M_s|\right)^p\right] \leq C_p E\left[(\langle M\rangle_{t\wedge\tau_n})^{\frac{p}{2}}\right]$$

The assertion follows by monotone convergence when letting $n \to \infty$. □

Exercise 6.10. ★☆☆ Prove the right BDG inequality in (6.1) for $p = 2$ and show that $C_2 = 4$. (*Hint:* Note that the proof simplifies compared to the case $p > 2$.)

6.2 Itô's formula in multiple dimension and the product formula

We proceed to state the multidimensional version of Itô's formula; for a proof, see [11, Theorem IV.(3.3)]. To this end note that an $\mathbb{R}^d$-*valued continuous semimartingale* is an $\mathbb{R}^d$-valued stochastic process $X = (X^1_t, \dots, X^d_t)_{t\geq 0}$ such that for each $i \in \{1, \dots, d\}$, X^i is a real-valued continuous semimartingale.

Theorem 6.11 (Itô's formula in several dimensions)**.** *Let $X = (X^1_t, \dots, X^d_t)_{t\geq 0}$ be an $\mathbb{R}^d$-valued continuous semimartingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions and $f : \mathbb{R}^d \to \mathbb{R}$ twice continuously differentiable. Then $f(X) = (f(X_t))_{t\geq 0}$ is again a continuous semimartingale and we have* Itô's formula*:*

$$\begin{aligned} f(X_t) = f(X_0) &+ \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s)\,\mathrm{d}X^i_s \\ &+ \frac{1}{2}\sum_{j,k=1}^d \int_0^t \frac{\partial^2 f}{\partial x_j \partial x_k}(X_s)\,\mathrm{d}\langle X^j, X^k\rangle_s \quad P\text{-a.s. for each } t \geq 0. \end{aligned}$$

Remark 6.12. (a) As in the one-dimensional case, Itô's formula is often written more compactly in differential notation as:

$$\mathrm{d}f(X_t) = \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_t)\,\mathrm{d}X^i_t + \frac{1}{2}\sum_{j,k=1}^d \frac{\partial^2 f}{\partial x_j \partial x_k}(X_t)\,\mathrm{d}\langle X^j, X^k\rangle_t.$$

(b) Itô's formula remains true if f is only defined on an open set $U \subset \mathbb{R}^d$ and X is P-a.s. U-valued.

(b) If for some $i \in \{1, \dots, d\}$, X^i is of finite variation, we only need to assume that f is once continuously differentiable with respect to the i-th coordinate. An important special case is if $X^i_t = t$.

Exercise 6.13. ★☆☆ Let $W = (W_t)_{t\geq 0}$ be a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Using Itô's formula in two dimensions, show that the process $(W^3_t - 3tW_t)_{t\geq 0}$ is a continuous local martingale.

An important special case of the multidimensional Itô formula is the *product formula*, also called *integration by parts*.

Theorem 6.14 (Product formula/Integration by parts). *Let X and Y be (real-valued) continuous semimartingales on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Then XY is again an semimartingale, and we have the* product formula*:*

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_u \,\mathrm{d}X_u + \int_0^t X_u \,\mathrm{d}Y_u + \langle X, Y\rangle_t \quad P\text{-a.s. for each } t \geq 0.$$

Proof. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) := xy$. Then Itô's formula gives

$$\begin{aligned}
\mathrm{d}\,(X_t Y_t) = \mathrm{d}f(X_t, Y_t) &= \frac{\partial f}{\partial x}(X_t, Y_t)\,\mathrm{d}X_t + \frac{\partial f}{\partial y}(X_t, Y_t)\,\mathrm{d}Y_t \\
&\quad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_t, Y_t)\,\mathrm{d}\langle X\rangle_t + 2 \times \frac{1}{2}\frac{\partial^2 f}{\partial x \partial y}(X_t, Y_t)\,\mathrm{d}\langle X, Y\rangle_t + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(X_t, Y_t)\,\mathrm{d}\langle Y\rangle_t \\
&= Y_t\,\mathrm{d}X_t + X_t\,\mathrm{d}Y_t + 0 + \mathrm{d}\langle X, Y\rangle_t + 0 \\
&= Y_t\,\mathrm{d}X_t + X_t\,\mathrm{d}Y_t + \mathrm{d}\langle X, Y\rangle_t.
\end{aligned}$$

□

Exercise 6.15. ★☆☆ Let $W = (W_t^1, W_t^2)_{t\geq 0}$ be two-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Using the product formula (together with the one-dimensional Itô formula) compute the dynamics of the process $X = (X_t)_{t\geq 0}$ given by $X_t = ((W_t^1)^2 - t)W_t^2$. Is X a continuous local martingale?

Another application of the-two dimensional Itô's formula yields the celebrated *Lévy characterisation of Brownian motion.* Its proof is left as an exercise.

Theorem 6.16 (Lévy's characterisation of Brownian motion). *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be filtered probability space satisfying the usual conditions and $M \in \mathcal{M}_{0,\mathrm{loc}}^{c}$. Suppose that*

$$\langle M\rangle_t = t \quad P\text{-a.s. for each } t \geq 0.$$

Then M is a Brownian motion.

6.3 Bayes' theorem

Our next goal is to understand how a continuous P-semimartingale behaves under an equivalent change of measure. To this end, we first need to understand how conditional expectations behave under an equivalent change of measure.

First, we recall the famous Radon-Nikodým theorem which shows that there is a one-to-one correspondence between equivalent probability measures Q and positive (normalised) P-integrable random variables Z; for a proof, we refer to [5, Chapter 28].

Theorem 6.17 (Radon-Nikodým). *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A probability measure Q on $(\Omega, \mathcal{F})$ is equivalent to P if and only if there exists an $\mathcal{F}$-measurable P-integrable random variable $Z > 0$ P-a.s. such that*

$$Q[A] = E^P[Z\mathbf{1}_A] \quad \text{for all } A \in \mathcal{F}.$$

Moreover, if it exists, Z is P-a.s. unique.

If $Q \approx P$ and Z is as in Theorem 6.17, we often write $Z = \frac{\mathrm{d}Q}{\mathrm{d}P}$ and call Z or $\frac{\mathrm{d}Q}{\mathrm{d}P}$ the *Radon-Nikodým derivative* of Q with respect to P. Note that $E^P[Z] = Q[\Omega] = 1$.

We note the following important corollary, whose proof is left as an exercise.

Corollary 6.18. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $Q \approx P$ on $\mathcal{F}$ an equivalent probability measure and $Z = \frac{\mathrm{d}Q}{\mathrm{d}P}$ the corresponding Radon-Nikodým derivative. Then for any nonnegative random variable X,*

$$E^Q[X] = E^P[ZX]. \tag{6.5}$$

Moreover, a random variable X is Q-integrable if and only if ZX is P-integrable, in which case, we also have (6.5).

Exercise 6.19. ★☆☆ Prove Corollary 6.18. (*Hint:* Use measure-theoretic induction.)

Suppose now that we are given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions, where $T > 0$ denotes a finite time horizon,[48] and $Q \approx P$ on $\mathcal{F}$. It follows from the definition of equivalent measures that then also $Q \approx P$ on $\mathcal{F}_t$ for each $t \in [0, T]$. A natural question is how the corresponding Radon-Nikodým derivates look like.

Definition 6.20. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtered probability space satisfying the usual conditions and $Q \approx P$ on $\mathcal{F}$. Define the right-continuous martingale $Z = (Z_t)_{t\in[0,T]}$ by[49]

$$Z_t := E^P\left[\frac{\mathrm{d}Q}{\mathrm{d}P} \,\middle|\, \mathcal{F}_t\right], \quad t \in [0, T]. \tag{6.6}$$

Then Z is called the *density process* of Q with respect to P.

Remark 6.21. Using the monotonicity of conditional expectations and the fact that $\frac{\mathrm{d}Q}{\mathrm{d}P} > 0$ P-a.s. it is not difficult to show that $Z_t > 0$ P-a.s. for each $t \in [0, T]$. Moreover, with more work one can show the stronger result that $P[\inf_{t\in[0,T]} Z_t > 0] = 1$.

The following result justifies the name "density process" for Z.

[48] The case of an infinite time horizon is similar. However, the case of a finite time horizon is more useful for applications in Mathematical Finance.

[49] More precisely, one can show that there exists a right-continuous martingale Z such that for each $t \in [0, T]$, (6.6) holds almost surely. One can choose Z even in such a way that it admits left-limits for all $t \in (0, T]$, in which case Z is called *càdlàg* (continue à gauche, limites à droite) or RCLL (right continuous with left limits); see [11, Theorem II.(2.9)] for details.

Proposition 6.22. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtered probability space satisfying the usual conditions, $Q \approx P$ on $\mathcal{F}$, and $Z = (Z_t)_{t\in[0,T]}$ the corresponding density process. Then for any stopping time $\tau \leq T$, $Q \approx P$ on $\mathcal{F}_\tau$ with corresponding Radon-Nikodým derivative*

$$\left.\frac{\mathrm{d}Q}{\mathrm{d}P}\right|_{\mathcal{F}_\tau} = Z_\tau \quad P\text{-a.s.}$$

Proof. Let $\tau \leq T$ be a stopping time and $A \in \mathcal{F}_\tau \subset \mathcal{F}_T \subset \mathcal{F}$. By the definition of the Radon-Nikodým derivative, the averaging property of conditional expectations and the (continuous time version of the) stopping theorem (Theorem 2.31), we obtain

$$Q[A] = E^P\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\mathbf{1}_A\right] = E^P\left[Z_T\mathbf{1}_A\right] = E^P\left[Z_\tau\mathbf{1}_A\right].$$

Since $Z_\tau > 0$ P-a.s. by Remark 6.21, and $E\left[Z_\tau\right] = E\left[Z_0\right] = E\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\right] = 1$, the claim follows from Theorem 6.17. □

We next study how conditional expectations for equivalent measures Q and P are related.

Theorem 6.23 (Bayes)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtered probability space satisfying the usual conditions, $Q \approx P$ on $\mathcal{F}$, and $Z = (Z_t)_{t\in[0,T]}$ the corresponding density process.*

(a) *Let $\sigma \leq \tau \leq T$ be stopping times and U an $\mathcal{F}_\tau$-measurable random variable satisfying $U \geq 0$ or $U \in \mathcal{L}^1(Q)$. Then*

$$E^Q\left[U \,|\, \mathcal{F}_\sigma\right] = \frac{1}{Z_\sigma}E^P\left[Z_\tau U \,|\, \mathcal{F}_\sigma\right] \quad Q\text{-a.s.}$$

(b) *An $\mathbb{F}$-adapted right-continuous process $M = (M_t)_{t\in[0,T]}$ is a (local) Q-martingale if and only if ZM is a (local) P-martingale.*[50]

Proof. (a) The measurability property of conditional expectations is trivially satisfied. To check the averaging property, fix $A \in \mathcal{F}_\sigma$. Then Proposition 6.22, the definition of the Radon-Nikodým derivative and the tower property of conditional expectations give

$$\begin{aligned} E^Q\left[\frac{1}{Z_\sigma}E^P\left[Z_\tau U \,|\, \mathcal{F}_\sigma\right]\mathbf{1}_A\right] &= E^P\left[E^P\left[Z_\tau U \,|\, \mathcal{F}_\sigma\right]\mathbf{1}_A\right] = E^P\left[E^P\left[Z_\tau U\mathbf{1}_A \,|\, \mathcal{F}_\sigma\right]\right] \\ &= E^P\left[Z_\tau U\mathbf{1}_A\right] = E^Q\left[U\mathbf{1}_A\right]. \end{aligned}$$

[50] A right-continuous local martingale is defined in the same way as a continuous local martingale, just replacing "continuous" by "right-continuous" in the definition. In particular, every right-continuous (local) martingale with continuous paths is a continuous (local) martingale. Also note that the localisation procedure naturally extends from the index set $[0,\infty)$ to the index set $[0,T]$. We still require that $\lim_{t\to\infty}\tau_n = +\infty$ P-a.s., i.e., the probability that we do not stop gets smaller and smaller as $n \to \infty$.

(b) We only consider the martingale case; the local martingale case follows by a localisation argument. First, assume that M is a Q-martingale. Then ZM is right-continuous and $\mathbb{F}$-adapted (because Z and M are) as well as P-integrable by Corollary 6.18. To check the martingale property, fix $0 \leq s \leq t \leq T$. Then by part (a) and the Q-martingale property of M,

$$E^P\left[Z_t M_t \,|\, \mathcal{F}_s\right] = Z_s \left(\frac{1}{Z_s} E^P\left[Z_t M_t \,|\, \mathcal{F}_s\right]\right) = Z_s \left(E^Q\left[M_t \,|\, \mathcal{F}_s\right]\right) = Z_s M_s \;\; P\text{-a.s.}$$

Conversely, assume that ZM is a P-martingale. Adaptedness of M follows by hypothesis and Q-integrability by Corollary 6.18. To check the martingale property, fix $0 \leq s \leq t \leq T$. Then by part (a) and the P-martingale property of ZM,

$$E^Q\left[M_t \,|\, \mathcal{F}_s\right] = \frac{1}{Z_s} E^P\left[Z_t M_t \,|\, \mathcal{F}_s\right] = \frac{1}{Z_s}(Z_s M_s) = M_s \;\; Q\text{-a.s.} \qquad \square$$

6.4 Girsanov's theorem

After these preparations, we now come back to the question how a continuous P-semimartingale behaves under an equivalent change of measure. Since being of finite variation is not changed by an equivalent change of measure, it suffices to consider this question for local martingales.

Theorem 6.24 (Girsanov)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtered probability space satisfying the usual conditions. Let $Q \approx P$ on $\mathcal{F}$ and assume that the corresponding density process $Z = (Z_t)_{t\in[0,T]}$ is continuous. Define the continuous local P-martingale $L = (L_t)_{t\in[0,T]}$ by*

$$L_t := \int_0^t \frac{1}{Z_u}\,\mathrm{d}Z_u, \quad t \in [0,T].$$

Then $Z = \mathcal{E}(L)$ and for any continuous local P-martingale $M = (M_t)_{t\in[0,T]}$, the process $\tilde{M} = (\tilde{M}_t)_{t\in[0,T]}$ defined by

$$\tilde{M}_t = M_t - \langle L, M\rangle_t, \quad t \in [0,T],$$

is a continuous local Q-martingale. Moreover, M is a Q-semimartingale with Q-decomposition $M = M_0 + (\tilde{M} - M_0) + \tilde{A}$, where $\tilde{A} = \langle L, M\rangle$.

Proof. First, it follows from the definition of L and associativity of the stochastic integral that

$$\mathrm{d}Z_t = Z_t \frac{1}{Z_t}\,\mathrm{d}Z_t = Z_t\,\mathrm{d}L_t.$$

Since $Z_0 = 1$, it follows from Proposition 6.6 that $Z = \mathcal{E}(L)$.[51]

[51] L is sometimes called the *stochastic logarithm* of Z.

Next, in order to prove that $\tilde{M}$ is a continuous local Q-martingale, by Bayes' theorem (Theorem 6.23(b)), it suffices to check that $Z\tilde{M}$ is a continuous local P-martingale. By the product formula (Theorem 6.14), the definition of $\tilde{M}$, the fact that $\langle L, M\rangle$ is of finite variation and the fact that $Z = \mathcal{E}(L)$, we obtain

$$\begin{aligned}
\mathrm{d}(Z_t\tilde{M}_t) &= Z_t\,\mathrm{d}\tilde{M}_t + \tilde{M}_t\,\mathrm{d}Z_t + \mathrm{d}\langle Z, \tilde{M}\rangle_t \\
&= Z_t\,\mathrm{d}M_t - Z_t\,\mathrm{d}\langle L, M\rangle_t + \tilde{M}_t\,\mathrm{d}Z_t + \mathrm{d}\langle Z, M\rangle_t \\
&= Z_t\,\mathrm{d}M_t - Z_t\,\mathrm{d}\langle L, M\rangle_t + \tilde{M}_t\,\mathrm{d}Z_t + Z_t\,\mathrm{d}\langle L, M\rangle_t \\
&= Z_t\,\mathrm{d}M_t + \tilde{M}_t\,\mathrm{d}Z_t.
\end{aligned}$$

Since both $Z \bullet M$ and $\tilde{M} \bullet Z$ are continuous local P-martingales, it follows that $Z\tilde{M}$ is a continuous local P-martingale.

The final claim is obvious. □

Exercise 6.25. ★☆☆ Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a filtered probability space satisfying the usual conditions. Let $Q \approx P$ on $\mathcal{F}$ and assume that the corresponding density process $Z = (Z_t)_{t\in[0,T]}$ is continuous. Let $X = (X_t)_{t\in[0,T]}$ be a continuous P-semimartingale with decomposition $X = X_0 + M + A$, where $M \in \mathcal{M}^c_{0,\mathrm{loc}}(P)$ and $A \in FV^c_0$. Find the Q-decomposition of X, i.e., write $X = X_0 + \tilde{M} + \tilde{A}$, where $\tilde{M} \in \mathcal{M}^c_{0,\mathrm{loc}}(Q)$ and $\tilde{A} \in FV^c_0$.

Remark 6.26. (a) One can show that Girsanov's theorem can be extended to the case that the density process Z is only right-continuous. In particular, it remains true that M is a continuous Q-semimartingale. This implies firstly that any continuous P-semimartingale X is also a continuous Q-semimartingale, and secondly that a predictable process H is integrable with respect to X under P if and only if it is integrable with respect to X under Q.

(b) In Section 6.5, we will see a condition on the filtration $\mathbb{F}$ that ensures that the density process Z is automatically continuous.

Because of its importance in applications, we study how to apply Girsanov's theorem to "remove the drift" of a Brownian motion by a change of measure.

Theorem 6.27 (Removal of drift for Brownian motion). *Let $W = (W_t)_{t\in[0,T]}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Let $b = (b_t)_{t\in[0,T]}$ be a predictable process satisfying*

$$\int_0^T b_t^2\,\mathrm{d}t < \infty \quad P\text{-a.s.}$$

Define the process $\tilde{W} = (\tilde{W}_t)_{t\in[0,T]}$ by

$$\tilde{W}_t := W_t + \int_0^t b_s\,\mathrm{d}s, \quad t \in [0, T].$$

Suppose that the stochastic exponential $\mathcal{E}(-\int_0^{\cdot} b_u \,\mathrm{d}W_u) = (\mathcal{E}(-\int_0^{\cdot} b_u \,\mathrm{d}W_u)_t)_{t\in[0,T]}$ *is a (true) P-martingale and define the probability measure $Q \approx P$ on $\mathcal{F}$ by* $\frac{\mathrm{d}Q}{\mathrm{d}P} := \mathcal{E}(-\int_0^{\cdot} b_u \,\mathrm{d}W_u)_T$. *Then $\tilde{W}$ is a Q-Brownian motion.*

Proof. It follows from Girsanov's theorem with $L := -\int b\,\mathrm{d}W$ and the defining property of the stochastic integral that

$$W - \langle L, W\rangle = W + \langle b \bullet W, W\rangle = W + b \bullet \langle W\rangle = W + \int_0^{\cdot} b_s \,\mathrm{d}s = \tilde{W}$$

is a continuous local Q-martingale, null at zero. Since

$$\langle \tilde{W}\rangle_t = \langle W\rangle_t = t \;\; Q\text{-a.s. for each } t \geq 0,$$

Lévy's characterisation of Brownian motion (Theorem 6.16) shows that $\tilde{W}$ is a Q-Brownian motion.[52] □

In order to apply Theorem 6.27, we need to check that the stochastic exponential of the local martingale $L = -b \bullet W$ is a (true) martingale. This is in general difficult to check. The following result yields a useful sufficient criterion; for a proof see [11, Proposition IIIV.(1.15)]

Theorem 6.28 (Novikov's condition)**.** *Let $L = (L_t)_{t\in[0,T]}$ be a continuous local martingale on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Suppose that*

$$E\left[\exp\left(\frac{1}{2}\langle L\rangle_T\right)\right] < \infty$$

Then the stochastic exponential $\mathcal{E}(L) = (\mathcal{E}(L)_t)_{t\in[0,T]}$ is a true martingale.

Exercise 6.29. ★☆☆ Let $W = (W_t)_{t\in[0,T]}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Let $X = (X_t)_{t\in[0,T]}$ be an arithmetic Brownian motion with drift $\mu \in \mathbb{R}$ and volatility $\sigma > 0$, i.e., $X_t = \mu t + \sigma W_t$. Show that there exists $Q \approx P$ on $\mathcal{F}$ such that X is a true Q-martingale. (*Hint:* Combine Theorems 6.27 and 6.28.)

6.5 Itô's representation theorem

We end this chapter by stating the very important theoretical result that Brownian motion has the *predictable representation property* with respect to its natural augmented filtration. We formulate the result for a finite time horizon because we are going to use the result in this context.[53] For a proof, see [11, Section V.3].

[52] Here, we have also used that the quadratic variation does not change under an equivalent change of measure. This follows from the fact that convergence in P-probability is equivalent to convergence in Q-probability for equivalent measures.

[53] A similar result holds for the index set $[0, \infty)$.

Theorem 6.30 (Itô's representation). *Let $W = (W_t)_{t\in[0,T]}$ be a Brownian motion on some complete probability space $(\Omega, \mathcal{F}, P)$. Let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be the filtration generated by W and augmented by the P-null sets in $\mathcal{F}$; see Appendix B for details.*[54] *Then every right-continuous local $(P, \mathbb{F})$-martingale $M = (M_t)_{t\in[0,T]}$ has P-a.s. continuous paths. Moreover, there exists a predictable process $H = (H_t)_{t\in[0,T]}$ in $L^2_{\mathrm{loc}}(W)$ such that*

$$M = M_0 + H \bullet W \quad P\text{-a.s.}$$

We note the following corollary which is very relevant in the context of Girsanov's theorem.

Corollary 6.31. *Let $W = (W_t)_{t\in[0,T]}$ be a Brownian motion on some complete probability space $(\Omega, \mathcal{F}, P)$. Let $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ be the filtration generated by W and augmented by the P-null sets in $\mathcal{F}$. Let $Q \approx P$ on $\mathcal{F}$. Denote the corresponding density process by $Z = (Z_t)_{t\in[0,T]}$. Then there exists $\nu \in L^2_{\mathrm{loc}}(W)$ such that*

$$Z = \mathcal{E}(\nu \bullet W) \quad P\text{-a.s.}$$

Proof. The density process Z is a continuous $(P, \mathbb{F})$-martingale by Itô's representation theorem. Define the continuous local martingale $L = (L_t)_{t\in[0,T]}$ by

$$L_t := \int_0^t \frac{1}{Z_u}\,\mathrm{d}Z_u, \quad t \in [0,T].$$

It follows as in the proof of Theorem 6.24 that $Z = \mathcal{E}(L)$. Moreover, using also that $L_0 = 0$, Itô's representation theorem gives $\nu \in L^2_{\mathrm{loc}}(W)$ such that

$$L = \nu \bullet W \quad P\text{-a.s.} \qquad \square$$

[54] Note that $\mathbb{F} = (\mathbb{F}^W)^P$ then satisfies the usual condition by Remark B.2.

7 Stochastic differential equations

The goal of this section is to present some basic notions and results on *stochastic differential equations* (SDEs), i.e., equations (for an unknown process $X = (X_t)_{t\geq 0}$) of the form

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0 = x,$$

where $W = (W_t)_{t\geq 0}$ is a Brownian motion.

7.1 Strong solutions

We first discuss the concept of *strong solutions.*

Definition 7.1. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $W = (W_t)_{t\geq 0}$ a $(P, \mathbb{F})$-Brownian motion. Let $x \in \mathbb{R}$ and $\mu, \sigma : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be Borel measurable functions. Then the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0 = x,$$

is said to have a *strong solution* $X = (X_t)_{t\geq 0}$ if X is continuous, $\mathbb{F}$-adapted and satisfies

$$\int_0^t \left(|\mu(s, X_s)| + |\sigma(s, X_s)|^2\right) \mathrm{d}s < \infty \;\; P\text{-a.s. for each } t \geq 0, \tag{7.1}$$

as well as

$$X_t = x + \int_0^t \mu(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s \;\; P\text{-a.s., for all } t \geq 0. \tag{7.2}$$

Remark 7.2. The integrability condition (7.1) ensures that the ds- and dW-integrals in (7.2) are well defined. Note that if μ and σ are continuous functions, this is automatically satisfied.

Exercise 7.3. ★☆☆ Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $W = (W_t)_{t\geq 0}$ a $(P, \mathbb{F})$-Brownian motion. Let $\mu \in \mathbb{R}$, $\sigma > 0$ and $s > 0$. Show that the SDE

$$\mathrm{d}S_t = S_t(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t), \quad S_0 = s, \tag{7.3}$$

has a strong solution. (*Hint:* Don't forget to check (7.1))

The following result gives existence and uniqueness of a strong solution if both μ and σ are continuous and satisfy a Lipschitz and linear growth condition in x, *uniformly* in t.

Theorem 7.4. *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $W = (W_t)_{t\geq 0}$ a $(P, \mathbb{F})$-Brownian motion. Let $x \in \mathbb{R}$ and $\mu, \sigma : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be continuous functions that satisfy the following two assumptions:*

Assumption 1: *μ and σ are* Lipschitz continuous *in x, uniformly in t, i.e., there is $K_1 > 0$ such that*

$$|f(t,x) - f(t,y)| \le K_1|x-y| \quad \textit{for all } t \in [0,\infty), x,y \in \mathbb{R}, \textit{ and } f \in \{\mu,\sigma\}.$$

Assumption 2: *μ and σ have* linear growth *in x, uniformly in t, i.e., there is $K_2 > 0$ such that*

$$|f(t,x)| \le K_2(1+|x|) \quad \textit{for all } t \in [0,\infty), x \in \mathbb{R}, \textit{ and } f \in \{\mu,\sigma\}.$$

Then the SDE

$$\mathrm{d}X_t = \mu(t,X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t, \quad X_0 = x, \tag{7.4}$$

has a unique strong solution $X = (X_t)_{t\ge0}$.

Proof. It suffices to show that for some $T > 0$ sufficiently small, there exists a strong solution to the SDE (7.4) on $[0,T]$. Indeed, if for some $T > 0$, we can construct a unique strong solution $X^{(1)} = (X^{(1)}_t)_{t\in[0,T]}$ on $[0,T]$, we can then consider the SDE (7.4) on $[T,2T]$ with the initial condition $X_T = X^{(1)}_T$ and construct a unique strong solution $X^{(2)} = (X^{(2)}_t)_{t\in[T,2T]}$ on $[T,2T]$. Proceeding by induction, given a unique strong solution $X^{(n)} = (X^{(n)}_t)_{t\in[(n-1)T,nT]}$ on $[(n-1)T,nT]$, we can consider the SDE (7.4) on $[nT,(n+1)T]$ with the initial condition $X_{nT} = X^{(n)}_T$ and construct a unique strong solution $X^{(n+1)} = (X^{(n+1)}_t)_{t\in[nT,(n+1)T]}$ on $[nT,(n+1)T]$. Finally, we can define $X = (X_t)_{t\ge0}$ by

$$X_t := X^{(n)}_t \quad \text{if } t \in [(n-1)T, nT),$$

and check that X is a unique strong solution on $[0,\infty)$.[55]

So we are left with the problem to construct a unique strong solution to the SDE (7.4) for some $T > 0$ sufficiently small. Denote by $\mathcal{R}^{2,c}_T$ the set of all continuous $\mathbb{F}$-adapted processes $X = (X_t)_{t\in[0,T]}$ that satisfy

$$\|X\|_{\mathcal{R}^{2,c}_T} := E\left[(X^*_T)^2\right]^{\frac{1}{2}} < \infty.$$

Once can show that $\mathcal{R}^{2,c}_T$ endowed with the norm $\|\cdot\|_{\mathcal{R}^{2,c}_T}$ is a complete normed vector space, i.e., a *Banach space*.

We are going to show the slightly weaker claim that for $T > 0$ sufficiently small, the SDE (7.4) has a unique strong solution $X = (X_t)_{t\in[0,T]}$ in $\mathcal{R}^{2,c}_T$.[56] Let $T > 0$, which will be chosen later. For each $X \in \mathcal{R}^{2,c}_T$, we define the map Φ from $\mathcal{R}^{2,c}_T$ to the set of continuous adapted

[55] Making the above process fully precise, requires some work (and a small extension of the argument on $[0,T]$ to a random initial condition).

[56] Note that the uniqueness statement is slightly weaker in that we prove uniqueness only in $\mathcal{R}^{2,c}_T$ rather than in the class of all continuous adapted processes.

processes $[0,T]$ by[57]

$$\Phi(X)_t := x + \int_0^t \mu(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s, \quad t \in [0,T]. \tag{7.5}$$

It suffices to show that Φ maps $\mathcal{R}_T^{2,c}$ to itself and has a unique fixed-point if T is small enough.

We first show that $\Phi(0) \in \mathcal{R}_T^{2,c}$. First, taking absolute values in (7.5), using the triangle inequality and taking the supremum over $[0,T]$, we obtain

$$\sup_{t\in[0,T]} |\Phi(0)_t| \leq |x| + \sup_{t\in[0,T]} \left| \int_0^t \mu(s,0)\,\mathrm{d}s \right| + \sup_{t\in[0,T]} \left| \int_0^t \sigma(s,0)\,\mathrm{d}W_s \right|.$$

Then taking the absolute value into the ds-integral and using Assumption 2 for μ, we obtain

$$\sup_{t\in[0,T]} |\Phi(0)_t| \leq |x| + K_2 T + \sup_{t\in[0,T]} \left| \int_0^t \sigma(s,0)\,\mathrm{d}W_s \right|.$$

Now taking squares, using the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$, taking expectations, applying the BDG inequality (Theorem 6.8) for $p = 2$ with constant $C_2 = 4$ and using Assumption 2 for σ yields

$$\begin{aligned} \|\Phi(0)\|^2_{\mathcal{R}_T^{2,c}} &= E\left[\Big(\sup_{t\in[0,T]} |\Phi(0)_t| \Big)^2 \right] \leq 2(|x| + K_2T)^2 + 2E\left[\left(\sup_{t\in[0,T]} \left| \int_0^t \sigma(s,0)\,\mathrm{d}W_s \right| \right)^2 \right] \\ &\leq 2(|x| + K_2T)^2 + 8E\left[\int_0^T \sigma(s,0)^2 ds \right] \leq 2(|x| + K_2T)^2 + 8K_2^2 T < \infty. \end{aligned}$$

It follows that $\Phi(0) \in \mathcal{R}_T^{2,c}$.

Next, let $X, X' \in \mathcal{R}_T^{2,c}$. Taking absolute values of the difference $\Phi(X)_t - \Phi(X')_t$, using the triangle inequality, taking the supremum over $[0,T]$, taking the absolute value into the ds-integral and using Assumption 1 for μ, we obtain

$$\begin{aligned} &\sup_{t\in[0,T]} |\Phi(X)_t - \Phi(X')| \\ &\quad \leq \sup_{t\in[0,T]} \int_0^t \left| \mu(s, X_s) - \mu(s, X'_s) \right| \,\mathrm{d}s + \sup_{t\in[0,T]} \left| \int_0^t \left(\sigma(s, X_s) - \sigma(s, X'_s) \right) \,\mathrm{d}W_s \right| \\ &\quad \leq K_1 T \sup_{t\in[0,T]} |X_s - X'_s| + \sup_{t\in[0,T]} \left| \int_0^t \left(\sigma(s, X_s) - \sigma(s, X'_s) \right) \,\mathrm{d}W_s \right|. \end{aligned}$$

Now taking squares, using the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$, taking expectations, applying the BDG inequality (Theorem 6.8) for $p = 2$ with constant $C_2 = 4$ and

[57] Note that Φ is well-defined, i.e., both the dt- and the dW-integral exist because μ and σ are continuous functions and X has continuous paths.

using Assumption 1 for σ yields

$$\begin{aligned}
\|\Phi(X)-\Phi(X')\|^2_{\mathcal{R}^{2,c}_T} &= E\left[\left(\sup_{t\in[0,T]}|\Phi(X)_t-\Phi(X')|\right)^2\right]\\
&\leq 2K_1^2T^2E\left[\left(\sup_{t\in[0,T]}|X_s-X'_s|\right)^2\right]+8E\left[\int_0^T\left(\sigma(s,X_s)-\sigma(s,X'_s)\right)^2\,\mathrm{d}s\right]\\
&\leq 2K_1^2T^2E\left[\left(\sup_{t\in[0,T]}|X_s-X'_s|\right)^2\right]+8K_1^2TE\left[\left(\sup_{t\in[0,T]}|X_s-X'_s|\right)^2\right]\\
&\leq K_1^2(2T^2+8T)\|X-X'\|^2_{\mathcal{R}^{2,c}_T}
\end{aligned}$$

Taking square roots yields

$$\|\Phi(X)-\Phi(X')\|_{\mathcal{R}^{2,c}_T}\leq K_1\sqrt{2T^2+8T}\|X-X'\|_{\mathcal{R}^{2,c}_T}. \tag{7.6}$$

Now if we consider (7.6) for $X'=0$, it follows from triangle inequality for $\|\cdot\|_{\mathcal{R}^{2,c}_T}$ and the fact that $\Phi(0)\in\mathcal{R}^{2,c}_T$,

$$\begin{aligned}
\|\Phi(X)\|_{\mathcal{R}^{2,c}_T} &\leq \|\Phi(0)\|_{\mathcal{R}^{2,c}_T}+\|\Phi(X)-\Phi(0)\|_{\mathcal{R}^{2,c}_T}\\
&\leq \|\Phi(0)\|_{\mathcal{R}^{2,c}_T}+K_1\sqrt{2T^2+8T}\|X\|_{\mathcal{R}^{2,c}_T}<\infty,
\end{aligned}$$

which show that Φ maps $\mathcal{R}^{2,c}_T$ to itself.

Finally, if we now choose T small enough that $K_1\sqrt{2T^2+8T}<1$, the map Φ is a contraction and therefore has a unique fixed point by Banach's fixed point theorem (cf. e.g. [1, Theorem 3.48]). □

We note an important corollary to Theorem 7.4 on the Markovian nature of (strong) solutions to SDEs. While it is beyond the scope of this module to discuss the Markovian nature of SDEs in detail, the following result is very useful when computing conditional expectations. It shows that calculating conditional expectations can be reduced to calculate ordinary expectations.

Corollary 7.5. *Let $W=(W_t)_{t\in[0,T]}$ be a Brownian motion on a filtered probability space $(\Omega,\mathcal{F},\mathbb{F}=(\mathcal{F}_t)_{t\in[0,T]},P)$ satisfying the usual conditions. Let $\mu,\sigma\in C([0,T]\times\mathbb{R})$ satisfy Assumptions 1 and 2 of Theorem 7.4.*[58] *Then for each pair $(t,x)\in[0,T)\times\mathbb{R}$, the SDE*

$$\mathrm{d}X_s=\mu(s,X_s)\,\mathrm{d}s+\sigma(t,X_s)\,\mathrm{d}W_s,\quad X_t=x \tag{7.7}$$

has a unique strong solution on $[t,T]$, which we denote by $X^{t,x}=(X^{t,x}_s)_{s\in[t,T]}$. Moreover, let

[58] Obviously, we only consider $t\in[0,T]$ here.

$h : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $h(x) \leq K_3(1 + x^2)$ for some $K_3 > 0$. Then $E[h(X_T^{t,x})] < \infty$ for all $(t, x) \in [0, T) \times \mathbb{R}$ and

$$E\left[h(X_T^{0,x}) \,\middle|\, \mathcal{F}_t\right] = v(t, X_t^{0,x}) \;\; P\text{-a.s.}, \tag{7.8}$$

where

$$v(t, x) = E\left[h(X_T^{t,x})\right]. \tag{7.9}$$

While Theorem 7.4 is a very important result, it does not cover important examples of SDEs that are relevant in Mathematical Finance. We therefore state without proof a much more powerful result (which needs much more advanced methods); for a proof, see [8, Theorem 5.2.13]

Theorem 7.6 (Yamada and Watanabe)**.** *Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual conditions and $W = (W_t)_{t\geq 0}$ a $(P, \mathbb{F})$-Brownian motion. Let $x \in \mathbb{R}$ and $\mu, \sigma : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be continuous functions that satisfy the following two assumptions:*
Assumption 1a: *μ is* Lipschitz continuous in *x, uniformly in t, i.e., there is $K_1 \geq 0$ such that*

$$|\mu(t, x) - f(t, y)| \leq K_1|x - y| \quad \text{for all } t \in [0, \infty), x, y \in \mathbb{R}.$$

Assumption 1b: *σ is* uniformly continuous in *x and* Hölder continuous of order $\frac{1}{2}$ in x near the origin, *uniformly in t, i.e., there is an increasing function $h : [0, \infty) \to [0, \infty)$ with $h(z) \leq K_2\sqrt{z}$ for $z \in [0, 1]$ for some $K_2 > 0$ such that*

$$|\sigma(t, x) - \sigma(t, y)| \leq h(|x - y|) \quad \text{for all } t \in [0, \infty), x, y \in \mathbb{R}.$$

Then the SDE

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0 = x,$$

has a unique strong solution $X = (X_t)_{t\geq 0}$.

Exercise 7.7. ★☆☆ Let $\delta \geq 0$ and $x \geq 0$. Using Theorem 7.6 show that the SDE

$$\mathrm{d}X_t = \delta\,\mathrm{d}t + 2\sqrt{|X_t|}\,\mathrm{d}W_t, \quad X_0 = x, \tag{7.10}$$

has a unique strong solution. It is called a *squared Bessel process of dimension δ*, started at x.[59]

[59] One can show that the solution to (7.10) is P-a.s. nonnegative, whence the absolute value in the square root may be omitted.

7.2 Weak solutions

Not every SDE has a strong solution. For example, one can show that the *Tanaka equation*

$$\mathrm{d}X_t = \operatorname{sign}(X_t)\,\mathrm{d}W_t, \quad X_0 = 0, \tag{7.11}$$

does not have a strong solution if $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ is a complete filtered probability space and $\mathbb{F}$ is the filtration generated by W and augmented by the P-null sets in $\mathcal{F}$; see [9, Example 26.15] for details.

For this reason, one would like to develop a solution concept for SDEs that allows to construct solutions for a wider class of equations, including the one by Tanaka. This leads to the notion of *weak solutions*.

Definition 7.8. Let $\mu, \sigma : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be Borel measurable functions and $x \in \mathbb{R}$. Then the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t, \quad X_0 = x$$

is said to have a *weak solution* $X = (X_t)_{t\geq 0}$, if there exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions and a $(P, \mathbb{F})$-Brownian motion $(W_t)_{t\geq 0}$ such that X is continuous, $\mathbb{F}$-adapted and satisfies

$$\int_0^t \big(|\mu(s, X_s)| + |\sigma(s, X_s)|^2\big)\,\mathrm{d}s < \infty \;\; P\text{-a.s. for each } t \geq 0,$$

as well as

$$X_t = x + \int_0^t \mu(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s \;\; P\text{-a.s., for all } t \geq 0.$$

The key difference between a weak and a strong solution is that for a *strong* solution, the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ and the $(P, \mathbb{F})$-Brownian motion $(W_t)_{t\geq 0}$ are *fixed*, whereas for a *weak* solution, the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ and the $(P, \mathbb{F})$-Brownian motion $(W_t)_{t\geq 0}$ are *part of the solution.*

Exercise 7.9. ★☆☆ Show that every strong solution to an SDE is also a weak solution to the same SDE.

As an illustration, we show that Tanaka equation (7.11) admits a weak solution.

Example 7.10. Let $X = (X_t)_{t\geq 0}$ be a Brownian motion on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. Define the process $W := (W_t)_{t\geq 0}$ by

$$W_t := \int_0^t \operatorname{sign}(X_s)\,\mathrm{d}X_s$$

Then $W \in \mathcal{M}^c_{0,\mathrm{loc}}$. Moreover,

$$\langle W \rangle_t = \int_0^t \operatorname{sign}(X_s)^2 \,\mathrm{d}s = t \ \ P\text{-a.s.}$$

It follows from Lévy's characterisation of Brownian motion (Theorem 6.16) that W is a $(P, \mathbb{F})$-Brownian motion. Moreover,

$$\mathrm{d}X_t = \operatorname{sign}(X_t)^2 \,\mathrm{d}X_t = \operatorname{sign}(X_t) \,\mathrm{d}W_t, \quad X_0 = 0,$$

whence X is a weak solution to the Tanaka equation (7.11).

Even if one can prove that an SDE has a strong solution by some theoretical result, one can sometimes construct a weak solution directly. This is due to Lévy's characterisation of Brownian motion (Theorem 6.16). This is illustrated by the following example.

Example 7.11. Let $n \in \mathbb{N}$ and $y > 0$. We proceed construct a nonnegative weak solution to the SDE

$$\mathrm{d}X_t = n \,\mathrm{d}t + 2\sqrt{X_t} \,\mathrm{d}B_t, \quad X_0 = x,$$

i.e., we construct a *squared Bessel process of dimension* n, started at x.[60]

Here we only consider the case $n \geq 2$, the case $n = 1$ is left as an exercise. Let $W = (W^1_t, \dots, W^n_t)_{t \geq 0}$ be a Brownian motion in $\mathbb{R}^n$ on some filtered probability space $(\Omega, \mathcal{F}, , \mathbb{F} = (F_t)_{t \geq 0}.P)$ satisfying the usual conditions. Fix $y = (y_1, \dots, y_n) \in \mathbb{R}^n \backslash \{0\}$ such that $\sum_{i=1}^n y_i^2 = x$ and define $\tilde{W} = (\tilde{W}^1_t, \dots, \tilde{W}^n_t)_{t \geq 0}$ by $\tilde{W}^i_t = W^i_t + y_i$.[61] Define the process $X = (X_t)_{t \geq 0}$ by

$$X_t = \sum_{i=1}^n (\tilde{W}^i_t)^2, \quad t \geq 0.$$

Applying the multidimensional Itô formula, using that $\tilde{W}^i$ (like the W^i) are independent and noting that $\langle \tilde{W}^i \rangle_t = t$ yields

$$\mathrm{d}X_t = \sum_{i=1}^n 2\tilde{W}^i_t \,\mathrm{d}\tilde{W}^i_t + \frac{1}{2} \sum_{i=1}^n 2 \,\mathrm{d}\langle \tilde{W}^i \rangle_t = 2 \sum_{i=1}^n \tilde{W}^i_t \,\mathrm{d}W^i_t + n \,\mathrm{d}t, \quad X_0 = x. \tag{7.12}$$

Since $n \geq 2$, one can show that $P[X_t > 0 \text{ for all } t \geq 0] = 1$. Hence, the continuous local martingale $B = (B_t)_{t \geq 0}$ defined by

$$B_t = \sum_{i=1}^n \int_0^t \frac{\tilde{W}^i_s}{\sqrt{X_s}} \,\mathrm{d}W^i_s, \quad t \geq 0,$$

[60] It follows from Exercise 7.7 that this SDE has a unique *strong* solution.
[61] $\tilde{W}$ is called a *Brownian motion* started at y.

is well defined. Using that the W^i are independent and recalling the definition of X, we obtain

$$\langle B\rangle_t = \sum_{i=1}^n \int_0^t \left(\frac{\tilde W^i_s}{\sqrt{X_s}}\right)^2 \mathrm{d}\langle W^i\rangle_s = \sum_{i=1}^n \int_0^t \frac{(\tilde W^i_s)^2}{X_s}\,\mathrm{d}s = \int_0^t \frac{\sum_{i=1}^n (\tilde W^i_s)^2}{X_s}\,\mathrm{d}s = t, \quad t \geq 0.$$

Lévy's characterisation of Brownian motion (Theorem 6.16), B is a Brownian motion.

Combining this with (7.12) and using linearity and associativity of the stochastic integral yields

$$\mathrm{d}X_t = 2\sqrt{X_t}\sum_{i=1}^n \frac{\tilde W^i_t}{\sqrt{X_t}}\,\mathrm{d}W^i_t + n\,\mathrm{d}t = 2\sqrt{X_t}\,\mathrm{d}B_t + n\,\mathrm{d}t, \quad X_0 = x. \tag{7.13}$$

Exercise 7.12. ★☆☆ Find a weak solution to SDE

$$\mathrm{d}X_t = \mathrm{d}t + 2\sqrt{X_t}\,\mathrm{d}B_t, \quad X_0 = x,$$

i.e., construct a squared Bessel process of dimension 1. (*Hint:* Argue as in Example 7.11 but define B slightly differently.)

Remark 7.13. The notion of strong and weak solutions to SDEs extends to multiple dimensions, where then also W is multidimensional (not necessarily of the same dimension as the number of equations). Theorem 7.4 and its proof extend naturally to multiple dimensions, the only change is that the absolute values have to be replaced by the appropriate norms. Theorem 7.6, however, is a genuine *one-dimensional* result.

7.3 Feynman–Kac formula

In this section, we establish a link between stochastic differential equations (SDEs) and (second-order linear parabolic) partial differential equations (PDEs). This link, which is known as the *Feynman-Kac formula*, is very powerful in both direction. From the perspective of Mathematical Finance, this link provides the possibility to calculate option prices by solving a PDE.

In the following result, we are rather vague on the precise conditions under which the theorem is true. In practise, the hard work is to make this precise in the given situation. To simplify the presentation, we introduce the following pieces of notation: If $U \subset \mathbb{R}^d$, we denote by $C(U)$ the set of all functions $f : U \to \mathbb{R}$ that are continuous, and if $I, J \subset \mathbb{R}$ are intervals, we denote by $C^{1,2}(I \times J)$ all functions $f : I \times J \to \mathbb{R}$ that are continuously differentiable in the first coordinate and twice continuously differentiable in the second coordinate.

Theorem 7.14 (Feynman–Kac)**.** *Let $W = (W_t)_{t\in[0,T]}$ be a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Let $\mu, \sigma \in C([0,T] \times \mathbb{R})$*

be "sufficiently nice". Suppose that for each pair $(t,x) \in [0,T) \times \mathbb{R}$, the SDE

$$\mathrm{d}X_s = \mu(s, X_s)\,\mathrm{d}s + \sigma(t, X_s)\,\mathrm{d}W_s, \quad X_t = x \tag{7.14}$$

has a unique strong solution on $[t,T]$, which we denote by $X^{t,x} = (X^{t,x}_s)_{s\in[t,T]}$. Let $h \in C(\mathbb{R})$ be "sufficiently nice" such that the second-order linear parabolic partial differential equation

$$\frac{\partial u}{\partial t}(t,x) + \mu(t,x)\frac{\partial u}{\partial x}(t,x) + \frac{1}{2}\sigma(t,x)^2\frac{\partial^2 u}{\partial x^2}(t,x) = 0, \qquad (t,x) \in [0,T) \times \mathbb{R}, \tag{7.15}$$

$$u(T,x) = h(x), \qquad x \in \mathbb{R}, \tag{7.16}$$

has a unique solution $u \in C([0,T] \times \mathbb{R}) \cap C^{1,2}([0,T) \times \mathbb{R})$ that is "sufficiently nice". Then for each $(t,x) \in [0,T) \times \mathbb{R}$, we have the Feynman-Kac formula

$$E\left[h(X^{t,x}_T)\right] = u(t,x). \tag{7.17}$$

Proof. Fix $(t,x) \in [0,T) \times \mathbb{R}$ and let $\varepsilon \in (0, T-t)$. Since $u \in C^{1,2}([0,T-\varepsilon] \times \mathbb{R})$, Itô's formula gives

$$\begin{aligned} u(T-\varepsilon, X^{t,x}_{T-\varepsilon}) &= u(t, X^{t,x}_t) + \int_t^{T-\varepsilon} \frac{\partial u}{\partial t}(s, X^{t,x}_s)\,\mathrm{d}s + \int_t^{T-\varepsilon} \frac{\partial u}{\partial x}(s, X^{t,x}_s)\mu(s, X^{t,x}_s)\,\mathrm{d}s \\ &\quad + \int_t^{T-\varepsilon} \frac{\partial u}{\partial x}(s, X^{t,x}_s)\sigma(s, X^{t,x}_t)\,\mathrm{d}W_s + \frac{1}{2}\int_t^{T-\varepsilon} \frac{\partial^2 u}{\partial x^2}(s, X^{t,x}_s)\sigma(s, X^{t,x}_s)^2\,\mathrm{d}s. \end{aligned}$$

Now using that u satisfies the PDE (7.15) and using that $X^{t,x}_t = x$, we obtain

$$u(T-\varepsilon, X^{t,x}_{T-\varepsilon}) = u(t,x) + \int_t^{T-\varepsilon} \frac{\partial u}{\partial x}(s, X^{t,x}_s)\sigma(s, X^{t,x}_t)\,\mathrm{d}W_s$$

Now taking expectations and using that the local martingale is a true martingale by the "niceness" assumption, we get

$$E\left[u(T-\varepsilon, X^{t,x}_{T-\varepsilon})\right] = u(t,x).$$

Now using that $\lim_{\varepsilon\to 0} u(T-\varepsilon, X^{t,x}_{T-\varepsilon}) = u(T, X^{t,x}_T) = h(X^{t,x}_T)$ by continuity of the paths of X and continuity of u on $[0,T] \times \mathbb{R}$ and interchanging the expectation and the limit by the "niceness" assumption, we get (7.17). □

Remark 7.15. (a) Theorem 7.14 remains true if there exists an open interval $U \subset \mathbb{R}$ such that $\mu, \sigma \in C([0,T] \times U)$, the SDE (7.14) has for each $(t,x) \in [0,T) \times U$ a unique strong solution on $[t,T]$ that is P-a.s. U-valued and $h \in C(U)$.

(b) Theorem 7.14 can be generalised to multidimensional SDEs (which are linked to PDEs with multiple space variables).

The identity (7.17) can be used in both directions: one can either use the solution to the PDE (7.15)–(7.16), e.g. using finite difference of finite element schemes, to calculate the expectation on the right hand side of (7.17), or one find the solution to the PDE (7.15)–(7.16) by calculating the expectation on the right hand side of (7.17), e.g. using a Monte-Carlo simulation of the corresponding SDE.

Exercise 7.16. ★☆☆ Let $\mu, \sigma \in \mathbb{R}$. Using the Feynman-Kac formula, find a solution to the PDE

$$\begin{aligned} \frac{\partial u}{\partial t}(t,x) + \mu \frac{\partial u}{\partial x}(t,x) + \frac{1}{2}\sigma^2 \frac{\partial^2 u}{\partial x^2}(t,x) &= 0, && (t,x) \in [0,T) \times \mathbb{R}, \\ u(T,x) &= x^2, && x \in \mathbb{R}. \end{aligned}$$

8 Pricing and hedging in continuous time

We now apply the results of the previous sections to describe *pricing* and *hedging* of *derivative contracts* like call or put options for financial markets in finite continuous time. While many of the concepts and results carry over from discrete time, the precise formulation of definitions and results is often (very) delicate and contains (many) pitfalls.

8.1 Financial market

We consider a financial market with $1+d$ assets, which are priced at times $t \in [0,T]$ for some *time horizon* $T \in \mathbb{N}$. We work on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. To simplify the analysis, we assume without further mentioning that $\mathcal{F}_0$ is P-trivial, i.e, $P[A] \in \{0,1\}$ for all $A \in \mathcal{F}_0$. We model the assets as continuous semimartingales $S^i = (S^i_t)_{t\in[0,T]}$, $i \in \{0,\dots,d\}$. We assume that S^0 is positive and of finite variation. More precisely, we assume that

$$S^0_t = \exp\left(\int_0^t r_s \,\mathrm{d}s\right), \quad t \in [0,T],$$

where the process $r = (r_t)_{t\geq 0}$ is predictable and satisfies $\int_0^T |r_s|\,\mathrm{d}s < \infty$ P-a.s. From a financial perspective, r_t denotes the *short rate* at time $t \in [0,T]$. We refer to S^0 also as *bank account.*

We set

$$S_t = (S^1_t,\dots,S^d_t) \quad \text{and} \quad \overline{S}_t = (S^0_t, S_t), \quad t \in [0,T],$$

and call the $\mathbb{R}^d$-valued continuous semimartingale $S = (S^1_t,\dots,S^d_t)_{t\in[0,T]}$ the *risky assets.*

We *discount* by S^0 (or take S^0 as *numéraire*) and define the *discounted assets* $X^0,\dots,X^d$ by

$$X^i_t := \frac{S^i_t}{S^0_t}, \quad t \in [0,T], \quad i \in \{0,\dots,d\}.$$

Then $X^0 \equiv 1$, and $X = (X^1_t,\dots,X^d_t)_{t\in\{0,\dots,T\}}$ expresses the value of the risky assets in units of the numéraire S^0. Note that $X^0,\dots,X^d$ are again continuous semimartingales.

Example 8.1 (Black–Scholes model)**.** The analogue of the Binomial model in discrete time is the Black–Scholes model (sometimes also called Black–Scholes–Samuelson model). Assume that $d=1$, i.e., there is only one risky asset. Also assume that $\mathbb{F}$ is the natural augmented filtration generated by some Brownian motion $W = (W_t)_{t\in[0,T]}$. Let $r,\mu \in \mathbb{R}$, $\sigma, s^1_0 > 0$. Assume that the bank account $S^0 = (S^0_t)_{t\in[0,T]}$ is given as the unique strong solution to the SDE

$$\mathrm{d}S^0_t = rS^0_t\,\mathrm{d}t, \quad S^0_0 = 1,$$

i.e., the short rate process r is constant, and the risky asset $S^1 = (S^1_t)_{t\in[0,T]}$ is given as the unique strong solution to the SDE

$$\mathrm{d}S_t^1 = \mu S_t^1 \,\mathrm{d}t + \sigma S_t^1 \,\mathrm{d}W_t, \quad S_0^1 = s_0^1,$$

for some $s > 0$. We have the explicit representation

$$\begin{aligned} S_t^0 &= \exp(rt), \\ S_t^1 &= s_0^1 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t - \sigma W_t\right). \end{aligned}$$

Exercise 8.2. ★☆☆ Let $\overline{S} = (S_t^0, S_t^1)_{t\in[0,T]}$ be a Black–Scholes model with parameters $r, \mu \in \mathbb{R}$, $\sigma, s_0^1 > 0$. Show that $X^1 = (X_t^1)_{t\in[0,T]}$ satisfies the SDE

$$\mathrm{d}X_t^1 = (\mu - r) X_t^1 \,\mathrm{d}t + \sigma X_t^1 \,\mathrm{d}W_t, \quad X_0^1 = s_0^1,$$

8.2 Admissible strategies

To describe trading in the market $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$, it is easiest to work with discounted quantities and to use the analogue of Proposition 3.8 as the definition for the self-financing property.

Definition 8.3. Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. A *self-financing strategy* for $\overline{S}$ is an $\mathbb{R}^{1+d}$-valued predictable process $\overline{\vartheta} = (\vartheta^0, \vartheta_t)_{t\in[0,T]}$ such that $\vartheta^i \in L(X^i)$ for $i \in \{1, \dots, d\}$ and[62]

$$\overline{\vartheta}_t \cdot \overline{X}_t = \overline{\vartheta}_0 \cdot \overline{X}_0 + (\vartheta \bullet X)_t := \overline{\vartheta}_0 \cdot \overline{X}_0 + \sum_{i=1}^d (\vartheta^i \bullet X^i)_t \;\; P\text{-a.s. for } t \in [0,T].$$

For a self-financing strategy $\overline{\vartheta} = (\vartheta^0, \vartheta_t)_{t\in[0,T]}$, we define the *discounted value process* $V(\overline{\vartheta}) = (V_t(\overline{\vartheta}))_{t\in[0,T]}$ and the *discounted gains process* $G(\vartheta) = (G_t(\vartheta))_{t\in[0,T]}$ by

$$\begin{aligned} V_t(\overline{\vartheta}) &:= \overline{\vartheta}_t \cdot \overline{X}_t, \quad t \in [0,T] \\ G_t(\vartheta) &:= (\vartheta \bullet X)_t := \sum_{i=1}^d (\vartheta^i \bullet X^i)_t, \quad t \in [0,T]. \end{aligned}$$

As in discrete time, we can identity self-financing strategies with pairs (V_0, ϑ), where $V_0 \in \mathbb{R}$ describes the initial wealth and $\vartheta = (\vartheta_t^1, \dots, \vartheta_t^d)_{t\in[0,T]}$ is a predictable process with $\vartheta^i \in L(X^i)$ for $i \in \{1, \dots, d\}$ that describes the holdings in the risky assets. Indeed, if (V_0, ϑ) is such a

[62] The condition $\vartheta^i \in L(X^i)$ for each $i \in \{1, \dots, d\}$ can (and should) be generalised to $\vartheta \in L(X)$ to obtain a complete theory. The latter means that the stochastic integral $\vartheta \bullet X$ can be defined but maybe can no longer be written as $\sum_{i=1}^d (\vartheta^i \bullet X^i)$ (because we may have that $\vartheta^i \notin L(X^i)$ for some i); for details on this very delicate and advanced point, we refer to [6, Section III.6]. For $d = 1$, this technical issue disappears.

pair, we can define the self-financing strategy $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in[0,T]}$ by defining the holding in the riskless asset $\vartheta^0 = (\vartheta_t^0)_{t\in[0,T]}$ by

$$\vartheta_t^0 := V_0 + \vartheta \bullet X_t - \vartheta_t \cdot X_t.$$

We will make use of this identification throughout the rest of the chapter and introduce as in discrete time the shorthand notation,

$$\overline{\vartheta} \mathrel{\widehat{=}} (V_0, \vartheta). \tag{8.1}$$

Remark 8.4. The identification (8.1) is tacitly done throughout most of the literature in Mathematical Finance and often – in an abuse of notation – ϑ *alone* is often called a self-financing strategy.

While the class of self-financing strategies in the largest possible class to describe trading in continuous time, it is too large from a financial perspective as it contains strategies that allow to create arbitrary positive values with zero initial wealth in finite time – even if the discounted risky assets are martingales. This is illustrated by the following example.

Example 8.5. Let $\overline{S} = (S_t^0, S_t^1)_{t\in[0,T]}$ be a Black–Scholes model with parameters $\mu = r$, $\sigma, s_0^1 > 0$. By Exercise 8.2, it follows that X^1 satisfies the SDE

$$\mathrm{d}X_t^1 = \sigma X_t^1 \,\mathrm{d}W_t, \quad X_0^1 = s_0^1.$$

Hence, it is a local martingale and even a (true) martingale by Novikov's condition (Theorem 6.28). One can show (with some effort) that the SDE

$$\mathrm{d}Z_t = -\frac{Z_t}{\sqrt{T-t}} \mathbf{1}_{\{t<T\}} \,\mathrm{d}W_t, \quad Z_0 = 1, \tag{8.2}$$

has a unique strong solution $Z = (Z_t)_{t\in[0,T]}$ that is nonnegative and satisfies $Z_T = 0$ P-a.s. Fix $c > 0$ and define the process $\vartheta^1 = (\vartheta_t^1)_{t\in[0,T]}$ by

$$\vartheta_t^1 = \frac{cZ_t}{\sigma X_t^1 \sqrt{T-t}} \mathbf{1}_{\{t<T\}}.$$

It follows from associativity and linearity of the stochastic integral and the SDE (8.2) that $\vartheta^1 \in L^2_{\mathrm{loc}}(X^1)$ and

$$\begin{aligned}\vartheta^1 \bullet X_T^1 &= \int_0^T \frac{cZ_t}{\sqrt{T-t}} \mathbf{1}_{\{t<T\}} \,\mathrm{d}W_t = -c \int_0^T -\frac{Z_t}{\sqrt{T-t}} \mathbf{1}_{\{t<T\}} \,\mathrm{d}W_t \\ &= -c \int_0^T \mathrm{d}Z_t = -c(Z_T - Z_0) = c \quad P\text{-a.s.}\end{aligned}$$

Thus, the self-financing strategy $\overline{\vartheta} \mathrel{\widehat{=}} (0, \vartheta^1)$ satisfies $V_0(\overline{\vartheta}) = 0$ and $V_T(\overline{\vartheta}) = \vartheta^1 \bullet X_T^1 = c$, i.e., it makes money out of nothing – despite the fact that X^1 is a (true) martingale.

To understand what goes wrong in Example 8.5, we need the following result on continuous local martingales that are bounded from below. Its proof is left as an exercise.

Proposition 8.6. *Let $M = (M_t)_{t\in[0,T]}$ be a continuous local martingale on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Suppose that $M_t \geq -a$ P-a.s. for all $t \in [0,T]$ for some $a \geq 0$. Then M is a $(P, \mathbb{F})$-supermartingale.*

Exercise 8.7. ★✫☆ Prove Proposition 8.6.

It follows from Proposition 8.6 that the local martingale $\vartheta^1 \bullet X^1$ in Example 8.5 is not bounded from below. This means that the sure positive value at time T is achieved by *taking excessive risk* at intermediate times. This type of strategy is reminiscent of the famous "doubling strategy" from roulette: you "set" on "red". If you win, you stop, and if you lose you double your bet and "set" on "red" again. You repeat until "red" appears. Since "red" finally appears with probability 1 (by Borel-Cantelli), you always win. It is clear that this type of gambling strategy should be and *is forbidden* in financial markets by so-called *margin requirements* which set clear limits on intermediate losses.[63]

The following definition introduces the class of self-financing strategies that should be allowed from a financial perspective.[64]

Definition 8.8. Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. A *self-financing strategy* $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in[0,T]}$ for $\overline{S}$ is called an *admissible strategy* for $\overline{S}$ if there exists $a \geq 0$ such that

$$G_t(\vartheta) = \vartheta \bullet X_t \geq -a \;\; P\text{-a.s.} \quad \text{for all } t \in [0,T].$$

The following exercise show that it does not matter whether we require the lower bound on the gains process or the value process of a self-financing strategy.

Exercise 8.9. ✫☆☆ Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Let $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in[0,T]}$ be a *self-financing strategy* for $\overline{S}$. Show that $\overline{\vartheta}$ is an admissible strategy for $\overline{S}$ if and only if there exists $b \geq 0$ such that

$$V_t(\overline{\vartheta}) = \overline{\vartheta} \cdot \overline{X}_t \geq -b \;\; P\text{-a.s.} \quad \text{for all } t \in [0,T].$$

[63]It is a sad fact that most of the spectacular losses in the financial industry can be traced back to that sort of strategy employed by "rogue traders". Also note that casinos have their own "margin requirements".

[64]One might argue that even this class is slightly too large and there should be a an upper boundary for a. Clearly taking $a = 0$ is financially prudent but maybe too restrictive.

8.3 No-arbitrage and the fundamental theorem of asset pricing

We proceed to extend the notion of arbitrage to continuous time.

Definition 8.10. Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. An admissible strategy $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in[0,T]}$ is called an *arbitrage opportunity* if

$$\overline{\vartheta}_0 \cdot \overline{S}_0 \leq 0, \quad \overline{\vartheta}_T \cdot \overline{S}_T \geq 0 \ \ P\text{-a.s.} \quad \text{and} \quad P[\overline{\vartheta}_T \cdot \overline{S}_T > 0] > 0.$$

The financial market $\overline{S}$ is called *arbitrage-free* if there are no arbitrage opportunities. In this case one also says that $\overline{S}$ satisfies *NA*.

Remark 8.11. As in discrete time, one can show that if the market $\overline{S}$ admits arbitrage, there always exists an arbitrage opportunity with $\overline{\vartheta}_0 \cdot \overline{S}_0 = 0$.

As in discrete time, we have the following equivalent characterisation of NA in terms of stochastic integrals. Its proof is left as an exercise.

Proposition 8.12. *Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. The following are equivalent:*

(a) *The market $\overline{S}$ satisfies NA.*

(b) *There does not exist a predictable process $\vartheta = (\vartheta_t^1, \ldots, \vartheta_t^d)_{t\in[0,T]}$ with $\vartheta^i \in L(X^i)$ for $i \in \{1, \ldots, d\}$ such that*

$$\vartheta \bullet X_t \geq -1 \ P\text{-a.s. for all } t \in [0,T], \quad \vartheta \bullet X_T \geq 0 \ P\text{-a.s.} \quad \text{and} \quad P[\vartheta \bullet X_T > 0] > 0.$$

Exercise 8.13. ★☆☆ Prove Proposition 8.12.

We now turn to the fundamental theorem of asset pricing. To this end, we need to slightly extend the notion of an equivalent martingale measure.

Definition 8.14. Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Denote by $X := S/S_0$ the discounted risky assets. A probability measure Q on $(\Omega, \mathcal{F})$ is called an *equivalent local martingale measure* (ELMM) for X if $Q \approx P$ and each X^i is a local Q-martingale, i.e., a local martingale under the measure Q.

We now formulate and prove the "easy" direction of the fundamental theorem of asset pricing in continuous time.[65] The "difficult" direction requires to extend the notion of no-arbitrage to the notion of so-called *No Free Lunch with Vanishing Risk*; see [2] for details.

[65]Note, however, that the "easy" direction is the relevant one in practice because it allows to check that a given model is arbitrage free.

Theorem 8.15 (Fundamental Theorem of Asset Pricing). *Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Denote by $X := S/S_0$ the discounted risky assets. If there exists an ELMM Q for X, then the market $\overline{S}$ satisfies NA.*

Proof. Let Q be an ELMM for X and $\vartheta = (\vartheta_t^1, \ldots, \vartheta_t^d)_{t\in[0,T]}$ a predictable process with $\vartheta^i \in L(X^i)$, $i \in \{1, \ldots, d\}$, $\vartheta \bullet X_t \geq -1$ P-a.s. for all $t \in [0,T]$ and $\vartheta \bullet X_T \geq 0$ P-a.s. By Proposition 8.12 it suffices to show that $\vartheta \bullet X_T = 0$ Q-a.s. Since $\vartheta^i \in L(X^i)$ for each $i \in \{1, \ldots, d\}$ and each X^i is a local Q-martingale, it follows from the definition of the stochastic integral (also using Remark 6.26(a)), that each $\vartheta^i \bullet X^i$ is a local Q-martingale. Since the finite sum of local Q-martingales is again a local Q-martingale, it follows that $\vartheta \bullet X$ is a local Q-martingale. Since it is bounded from below by -1, it follows from Proposition 8.6 that $\vartheta \bullet X$ is a Q-supermartingale. This gives

$$E^Q[\vartheta \bullet X_T] \leq E^Q[\vartheta \bullet X_0] = 0.$$

This together with the fact that $\vartheta \bullet X_T \geq 0$ Q-a.s. implies that $\vartheta \bullet X_T = 0$ Q-a.s. □

We proceed to illustrate the FTAP in the case of the Black–Scholes model.

Example 8.16. Let $\overline{S} = (S_t^0, S_t^1)_{t\in[0,T]}$ be a Black–Scholes market with parameters $\mu, r \in \mathbb{R}$, $\sigma, s_0^1 > 0$, cf. Example 8.1. It follows from Exercise 8.2 that the discounted risky asset X^1 satisfies the SDE

$$\begin{aligned} \mathrm{d}X_t^1 &= (\mu - r)X_t^1 \,\mathrm{d}t + \sigma X_t^1 \,\mathrm{d}W_t \\ &= \sigma X^1(\lambda \,\mathrm{d}t + \mathrm{d}W_t), \quad X_0^1 = s_0^1, \end{aligned}$$

where $\lambda := \frac{\mu - r}{\sigma}$ denotes the *market price of risk* or *Sharpe ratio*. Since λ is a constant, it follows from Novikov's condition (Theorem 6.28) that $\mathcal{E}(-\lambda W)$ is a true P-martingale. Define the measure $Q \approx P$ on $\mathcal{F}$ by

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \mathcal{E}(-\lambda W)_T.$$

Then it follows from the removal-of-drift result (Theorem 6.27) that the process $\tilde{W} = (\tilde{W}_t)_{t\in[0,T]}$ defined by

$$\tilde{W}_t := W_t + \lambda t, \quad t \in [0,T], \tag{8.3}$$

is a Q-Brownian motion. It follows that X^1 satisfies the SDE

$$\mathrm{d}X_t^1 = \sigma X^1 \,\mathrm{d}\tilde{W}_t, \quad X_0^1 = s_0^1,$$

and is hence a local Q-martingale. Moreover, since $X^1 = s_0^1 \mathcal{E}(\sigma \tilde{W})$ it follows from Novikov's criterion (applied under Q) that X^1 is a true Q-martingale. Hence Q is an ELMM and even an EMM for X^1. By Theorem 8.15, this implies that the Black–Scholes model satisfies NA.

8.4 Valuation of contingent claims

We now turn to the definition of contingent claims. Apart from changing the index set, the definition of a contingent claim or derivate security carries over verbatim from Definition 3.20, and we do not repeat it here.

As in discrete time, we only study the valuation of attainable contingent claim. Maybe surprisingly, the definition of an attainable contingent claim is much more delicate in continuous time. This is due to the existence of "crazy" self-financing strategies as in Example 8.5.

Definition 8.17. Let $\overline{S} = (S^0_t, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Assume that there exists an ELMM for the discounted risky assets $X = S/S^0$. A contingent claim C is called *attainable* or *replicable* if there exists an admissible strategy $\overline{\vartheta} = (\vartheta^0_t, \vartheta_t)_{t\in[0,T]}$ such that

$$\overline{\vartheta}_T \cdot \overline{S}_T = C \quad P\text{-a.s.} \tag{8.4}$$

and the corresponding discounted value process $V(\overline{\vartheta}) = V_0(\overline{\vartheta}) + \vartheta \bullet X$ is a true Q-martingale under some ELMM $Q \approx P$. In this case $\overline{\vartheta}$ is called a *replication strategy* or *(perfect) hedge* for C.

A few comments are in order.

Remark 8.18. (a) It may seem odd that we require the existence of a dual object (an ELMM) for the definition of attainability. Indeed, there is a primal definition that does not rely on ELMMs but this is more involved. (It requires that $\overline{\vartheta}$ is so-called *maximal*.)

(b) If C is a contingent claim, for any admissible trading strategy $\overline{\vartheta}$ satisfying (8.4), the corresponding discounted value process $V(\overline{\vartheta}) = V_0(\overline{\vartheta}) + \vartheta \bullet X$ is a local Q-martingale under each ELMM Q. But it might not be a true Q-martingale under any ELMM Q. Therefore, the additional condition is important. (This issue does not arise in discrete time by virtue of Proposition 3.17.)

(c) A contingent claim C is attainable if and only if the corresponding *discounted contingent claim* $H := \frac{C}{S^0_T}$ satisfies $H = \overline{\vartheta}_T \cdot \overline{X}_T = V_T(\overline{\vartheta})$ and the corresponding discounted value process $V(\overline{\vartheta})$ is a true Q-martingale under some ELMM $Q \approx P$. In this case, we say that the discounted contingent claim H is *attainable* and call $\overline{\vartheta}$ a *replication strategy* for H.

We now turn to the valuation of attainable contingent claim. This turns out to be very delicate. First of all, our argument requires the existence of an EMM (and not just an ELMM) for the discounted risky assets, and we also need to require that the discounted contingent claim is bounded. With these provisos, the following result is an analogue to Theorem 3.26 in discrete time. We leave the proof to the reader.

Theorem 8.19. *Let $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual condition and H an attainable discounted contingent claim that is* bounded. *Denote by $\mathcal{P}$ the set of all ELMMs for X and assume that $\mathcal{P}$ contains at least one EMM.*

(a) *There exists a P-a.s. unique continuous adapted process $(V_t^H)_{t\in[0,T]}$ with $V_T^H = H$ P-a.s., called the (discounted)* fair value process *of H, such that the extended $(1 + d + 1)$-dimensional market*

$$(\overline{S}, S^0 V^H) = (S_t^0, S_t^1, \dots, S_t^d, S_t^0 V_t^H)_{t\in[0,T]}$$

admits an EMM for (X, V^H). It is given by

$$V_t^H = E^Q [H \mid \mathcal{F}_t] \quad P\text{-a.s.}, \quad t \in [0, T], \quad \text{for all } Q \in \mathcal{P}$$

(b) *If $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in\{1,\dots,T\}}$ is any replication strategy for H, then*

$$V_t^H = V_t(\overline{\vartheta}) = V_0^H + \vartheta \bullet X_t \quad P\text{-a.s.}, \quad t \in [0, T], \tag{8.5}$$

and $\vartheta \bullet X$ is a $(Q, \mathbb{F})$-martingale for any $Q \in \mathcal{P}$.

Remark 8.20. If C is an undiscounted contingent claim such that the corresponding discounted contingent claim $H := C/S_T^0$ is bounded and $V^H = (V_t^H)_{t\in[0,T]}$ is defined as in Theorem 8.19, we define – in a slight abuse of notation – the process $V^C = (V_t^C)_{t\in[0,T]}$ by $V_t^C := V_t^H$, $t \in [0, T]$, and call V^C the (discounted) fair value process of C.

8.5 Complete markets and the predictable representation property

We now turn to the notion of complete markets and the predictable representation property. Again these concepts are conceptually the same but technically more involved than in discrete time.

Definition 8.21. A financial market $\overline{S} = (S_t^0, S_t)_{t\in[0,T]}$ on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions is called *complete* if each discounted *bounded* contingent claim H is attainable. Otherwise, it is called *incomplete*.

Definition 8.22. Let $M^1 = (M_t^1)_{t\in[0,T]}, \dots, M^d = (M_t^d)_{t\in[0,T]}$ be continuous local $(P, \mathbb{F})$-martingales on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Then $M = (M^1, \dots, M^d)$ is said to have the *predictable representation property* for the pair $(P, \mathbb{F})$, if every right-continuous local $(P, \mathbb{F})$-martingale $N = (N_t)_{t\in[0,T]}$ has continuous paths and there exists an $\mathbb{F}$-predictable process $H = (H_t^1, \dots, H_t^d)_{t\in[0,T]}$ with values in $\mathbb{R}^d$ such that $H^i \in L^2_{\mathrm{loc}}(M^i)$ for $i \in \{1, \dots, d\}$ and[66]

$$N_t = N_0 + H \bullet M_t \quad P\text{-a.s.}, \quad t \in [0, T].$$

We then get the following version of the second-fundamental theorem of asset pricing in continuous time.

Theorem 8.23. *Let $\overline{S} = (S^0_t, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Denote by $X = S/S^0$ the discounted risky assets and assume that there exists an ELMM for X. Then the following are equivalent:*

(a) *The market $\overline{S}$ is complete.*

(b) *X has the predictable representation property for the pair $(Q^*, \mathbb{F})$ for some ELMM Q^*.*

(c) *There exists a unique ELMM Q^* on $\mathcal{F}_T$ for X.*

We note an important corollary for the valuation of contingent claims in complete markets, provided that the unique ELMM Q^* for X is even an EMM.[67]

Corollary 8.24. *Let $\overline{S} = (S^0_t, S_t)_{t\in[0,T]}$ be a financial market on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, P)$ satisfying the usual conditions. Assume that there exists a unique ELMM Q for $X = S/S^0$ that is an EMM. Then*

(a) *A discounted contingent claim H is attainable if and only if $E^Q[H] < \infty$.*

(b) *For each attainable discounted contingent claim H, there exists a P-a.s. unique continuous adapted process $(V^H_t)_{t\in[0,T]}$ with $V^H_T = H$ P-a.s., called the (discounted)* fair value process *of H, such that the extended $(1 + d + 1)$-dimensional market*

$$(\overline{S}, S^0 V^H) = (S^0_t, S^1_t, \dots, S^d_t, S^0_t V^H_t)_{t\in[0,T]}$$

admits an EMM for (X, V^H). It satisfies

$$V^H_t = E^Q[H \,|\, \mathcal{F}_t] \quad P\text{-a.s.}, \quad t \in [0, T].$$

Moreover, if $\overline{\vartheta} = (\vartheta^0_t, \vartheta_t)_{t\in[0,T]}$ is any replication strategy for H, then

$$V^H_t = V_t(\overline{\vartheta}) = V^H_0 + \vartheta \bullet X_t \quad P\text{-a.s.}, \quad t \in [0, T]. \tag{8.6}$$

[66] Strictly speaking, this is not fully correct. The correct assertion is that there exists $H \in L^2_{\mathrm{loc}}(M)$. In this case, the stochastic integral $H \bullet M$ can be defined but may not always be written as $\sum_{i=1}^d H^i \bullet M^i$ (because we may have that $H^i \notin L^2_{\mathrm{loc}}(M^i)$ for some i); for details on this very delicate and advanced point, we refer to [6, Section III.6]. The same caveat applies to the replicating strategies in Corollary 8.24, which may only be in $L(X)$. For $d = 1$, this technical issue disappears.

[67] If the unique ELMM Q^* for X in not an EMM, the valuation of contingent claims is surprisingly difficult. Note that many practitioners (and textbooks) just ignore the issue.

Proof. We only establish part (a), part (b) is left as an exercise.

First assume that H is attainable. By Definition 8.17, there exists an admissible strategy $\overline{\vartheta} = (\vartheta_t^0, \vartheta_t)_{t\in[0,T]}$ such that $H = V_T(\overline{\vartheta})$ and such that $V(\overline{\vartheta})$ is a Q-martingale (since there exist only one ELMM on $\mathcal{F}_T$). By Q-integrability of $V(\overline{\vartheta})$, it follows that $H = V_T(\overline{\vartheta})$ is Q-integrable.

Conversely, assume that $E^Q[H] < \infty$. Define the right-continuous and nonnegative Q-martingale $V^H = (V_t^H)_{t\in[0,T]}$ by $V_t^H := E^Q[H \mid \mathcal{F}_t]$.[68] By Theorem 8.23, X has the predictable representation property for the pair $(Q, \mathbb{F})$. This implies that V^H has continuous paths and there exists a predictable process $\vartheta = (\vartheta_t^1, \dots, \vartheta_t^d)_{t\in[0,T]}$ such that $V^H = V_0^H + \vartheta \bullet X$. Let $\overline{\vartheta} \mathrel{\widehat{=}} (V^H, \vartheta)$. Since $V(\overline{\vartheta}) = V_0^H + \vartheta \bullet X = V^H$ is a nonnegative Q-martingale with $V_T(\overline{\vartheta}) = V_T^H = H$ Q-a.s., it follows that $\overline{\vartheta}$ is an admissible strategy and a replication strategy for H. $\square$

Exercise 8.25. ★☆☆ Prove Corollary 8.24(b).

Remark 8.26. In the setting of Corollary 8.24, if C is an undiscounted contingent claim such that the corresponding discounted contingent claim $H := C/S_T^0$ satisfies $E^Q[H] < \infty$ and $V^H = (V_t^H)_{t\in[0,T]}$ is defined as in Corollary 8.24, we define – in a slight abuse of notation – the process $V^C = (V_t^C)_{t\in[0,T]}$ by $V_t^C := V_t^H$, $t \in [0,T]$, and call V^C the (discounted) fair value process of C.

8.6 Pricing and hedging in the Black–Scholes model

We conclude this chapter by applying the preceding theory to the special but important case of a Black–Scholes model.

Theorem 8.27. *Let $\overline{S} = (S_t^0, S_t^1)_{t\in[0,T]}$ be a Black–Scholes model with parameters $r, \mu \in \mathbb{R}$, $\sigma, s_0^1 > 0$; cf. Example 8.1. Then the market $\overline{S}$ admits a unique ELMM Q on $\mathcal{F}_T$ for $X^1 = S^1/S^0$, which is even an EMM for X^1. It is given by $\frac{\mathrm{d}Q}{\mathrm{d}P} = \mathcal{E}(-\lambda W)_T$, where $\lambda := \frac{\mu - r}{\sigma}$ denotes the* market price of risk. *Consequently, the market $\overline{S}$ is arbitrage-free and complete and X^1 satisfies the predictable representation property for $(Q, \mathbb{F})$. Moreover, if H is a discounted contingent claim with $E^Q[H] < \infty$, its unique fair value process $(V_t^H)_{t\in[0,T]}$ satisfies*

$$V_t^H = E^Q[H \mid \mathcal{F}_t] \quad P\text{-a.s.} \quad t \in [0,T]. \tag{8.7}$$

Proof. Apart from uniqueness of Q on $\mathcal{F}_T$, the claims follow from Example 8.16, Theorems 8.15 and 8.23 and Corollary 8.24.

Let $\tilde{Q} \approx P$ on $\mathcal{F}$ be such that X^1 is a local $\tilde{Q}$-martingale. Denote by $\tilde{Z} = (\tilde{Z}_t)_{t\in[0,T]}$ the corresponding density process. It follows from Corollary 6.31 that $\tilde{Z} = \mathcal{E}(\nu \bullet W)$ for some $\nu \in L^2_{\text{loc}}(W)$. Since X^1 is a $\tilde{Q}$-local martingale, it follows from Bayes' theorem (Theorem

[68] The precise claim is that there exist a Q-martingale $V^H = (V_t^H)_{t\in[0,T]}$ such that $V_t^H := E^Q[H \mid \mathcal{F}_t]$ Q-a.s. for all $t \in [0,T]$.

6.23(a)) that $\tilde{Z}X^1$ is a local P-martingale. Using the dynamics of $\tilde{Z}$ and the dynamics of X^1 from Exercise 8.2, the product formula gives

$$\begin{aligned}\mathrm{d}(\tilde{Z}_t X_t^1) &= \tilde{Z}_t \,\mathrm{d}X_t^1 + X_t^1 \,\mathrm{d}\tilde{Z}_t + \mathrm{d}\langle \tilde{Z}, X^1\rangle_t \\ &= \tilde{Z}_t(\mu - r)X_t^1 \,\mathrm{d}t + \tilde{Z}_t\sigma X_t^1 \,\mathrm{d}W_t + X_t^1 \tilde{Z}_t \nu_t \,\mathrm{d}W_t + \tilde{Z}_t \nu_t \sigma X_t^1 \,\mathrm{d}t \\ &= \tilde{Z}_t X_t^1 \left((\mu - r) + \nu_t\sigma\right) \mathrm{d}t + \tilde{Z}_t X_t^1(\sigma + \nu_t)\,\mathrm{d}W_t.\end{aligned}$$

Since $\tilde{Z}X^1$ is a local P-martingale, it follows that the finite variation dt-term must disappear. This implies that the integrand of the dt-term must be 0 outside a $(P \otimes \mathrm{d}t)$-nullsets.[69] Since $\tilde{Z}$, X^1 and σ are positive, dividing by $\tilde{Z}$, X^1 and σ shows that $\nu = -\frac{\mu - r}{\sigma} = \lambda$ outside a $(P \otimes \mathrm{d}t)$-nullset. It follows that $\tilde{Z} = \mathcal{E}(-\lambda \bullet W)$. Proposition 6.22 and the definition of $\tilde{Q}$ give

$$\left.\frac{\mathrm{d}\tilde{Q}}{\mathrm{d}P}\right|_{\mathcal{F}_T} = \tilde{Z}_T = \mathcal{E}(-\lambda \bullet W)_T = \left.\frac{\mathrm{d}Q}{\mathrm{d}P}\right|_{\mathcal{F}_T} \quad P\text{-a.s.} \tag{8.8}$$

It follows that $\tilde{Q} = Q$ on $\mathcal{F}_T$. □

In the special case that $H = h(X_T^1)$, we can describe the value process by a PDE and calculate the hedging strategy by the *delta hedge*.

Corollary 8.28. *Let $\overline{S} = (S_t^0, S_t^1)_{t\in[0,T]}$ be a Black–Scholes model with parameters $r, \mu \in \mathbb{R}$, $\sigma, s_0^1 > 0$; cf. Example 8.1. Let $h \in C((0,\infty))$ be of polynomial growth, i.e., there are $K, p > 0$ such that*

$$h(x) \leq K(1 + x^p), \quad x \in (0, \infty).$$

Then the PDE

$$\frac{\partial v}{\partial t}(t,x) + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v}{\partial x^2}(t,x) = 0, \qquad (t,x) \in [0,T) \times (0,\infty), \tag{8.9}$$

$$v(T,x) = h(x), \qquad x \in (0,\infty) \tag{8.10}$$

has a unique solution (of polynomial growth) $v \in C([0,T] \times (0,\infty)) \cap C^{1,2}([0,T) \times (0,\infty))$. It can be computed explicitly by

$$v(t,x) = E\left[h\left(x \exp\left(\sigma\sqrt{T-t}Z - \frac{1}{2}\sigma^2(T-t)\right)\right)\right], \quad (t,x) \in [0,T] \times (0,\infty), \tag{8.11}$$

[69] Here $(P \otimes \mathrm{d}t)$ is a (standard) shorthand for the product measure $(P \otimes \Lambda_{[0,T]})$, where $\Lambda_{[0,T]}$ is the Lebesgue-measure on $[0,T]$. The rigorous argument goes as follows: Define the process $H = (H_t)_{t\geq 0}$ by $H_t = \tilde{Z}_t X_t^1((\mu - r) - \nu\sigma)$ and the process $A = (A_t)_{t\geq 0}$ by $A_t = \int_0^t H_s \,\mathrm{d}t$. Then by Theorem 4.33, the paths of A are P-a.s. constant 0. This implies that for all ω outside a P-null set $N \subset \Omega$, $A_t(\omega) = 0$ for all $t \in [0,T]$. So fix $\omega \in \Omega \setminus N$. Then by the fundamental theorem of calculus for the Lebesgue integral (see e.g. [7, Theorem 2.15]), it follows that $H_t(\omega) = 0$ for all t outside a $\Lambda_{[0,T]}$-nullset (depending on ω). This implies that $\int_0^T |H_t(\omega)|\,\mathrm{d}t = 0$. This in turn (together with Fubini's theorem) implies that $E\left[\int_0^T |H_t(\omega)|\,\mathrm{d}t\right] = \int_{\Omega\times[0,T]} |H_t(\omega)|(P \otimes \Lambda_{[0,T]})(\mathrm{d}(\omega,t)) = 0$, which in turn implies that $H = 0$ outside a $(P \otimes \Lambda_{[0,T]})$-nullset.

where $Z \sim \mathcal{N}(0,1)$*. If* $H = h(X^1_T)$ *is a discounted contingent claim, then its fair value process* $V^H = (V^H_t)_{t\in[0,T]}$ *satisfies*

$$V^H_t = v(t, X^1_t) \;\; P\text{-a.s.}, \quad t \in [0,T]. \tag{8.12}$$

Moreover, the corresponding hedging strategy $\overline{\vartheta}^H \mathrel{\widehat{=}} (V^H_0, \vartheta^{1,H})$ *is unique (up to* $(P \otimes \mathrm{d}t)$*-nullset). It can be calculated explicitly by the the* delta hedge

$$\vartheta^{1,H}_t := v_x(t, X^1_t), \quad t \in [0,T). \tag{8.13}$$

Proof. Let $Q \approx P$ be the unique ELMM on $\mathcal{F}_T$ for X^1. It follows from Example 8.16 that X^1 satisfies the SDE

$$\mathrm{d}X^1_t = \sigma X^1 \,\mathrm{d}\tilde{W}_t, \quad X^1_0 = s^1_0,$$

where $\tilde{W} = (\tilde{W}_t)_{t\in[0,T]}$ is a Q-Brownian motion. Moreover, for $(t,x) \in [0,T) \times (0,\infty)$, denote by $X^{1,t,x} = (X^{1,t,x}_s)_{s\in[t,T]}$ the unique strong solution to the SDE

$$\mathrm{d}X^{1,t,x}_s = \sigma X^{1,t,x}_s \,\mathrm{d}\tilde{W}_s, \quad X^{1,t,x}_t = x.$$

Proposition 6.6 (and a straightforward extension) gives

$$X^1_T = s^1_0 \exp\left(\sigma \tilde{W}_T - \frac{\sigma^2}{2} T\right), \tag{8.14}$$

$$X^{1,t,x}_T = x \exp\left(\sigma(\tilde{W}_T - \tilde{W}_t) - \frac{\sigma^2}{2}(T-t)\right). \tag{8.15}$$

It follows from the Feynman–Kac formula (7.17) and (8.15) that[70]

$$v(t,x) = E^Q\left[h(X^{1,t,x}_T)\right] = E^Q\left[h\left(x \exp\left(\sigma(\tilde{W}_T - \tilde{W}_t) - \frac{\sigma^2}{2}(T-t)\right)\right)\right]$$

Now (8.11) follows from the fact that $\tilde{W}_T - \tilde{W}_t \sim \mathcal{N}(0, T-t)$ under Q.

Next, (8.7) and Corollary 7.5 gives[71]

$$V^H_t = E^Q\left[h(X^1_T) \,\middle|\, \mathcal{F}_t\right] = v(t, X^1_t) \;\; Q\text{-a.s.}, \quad t \in [0,T],$$

which together with $Q \approx P$ on $\mathcal{F}$ gives (8.12).

Finally, let $\overline{\vartheta} = (\vartheta^0_t, \vartheta^1_t)_{t\in[0,T]}$ be any replication strategy for C (which exists by Corollary

[70] It is not difficult to check that the "niceness assumptions" are satisfied in this case.
[71] Note that Corollary 7.5 extends here to the case that h has polynomial growth.

8.24). Fix $T' \in [0, T)$. Then (8.6) gives

$$V_t^C - V_0^C = \int_0^t \vartheta_s^1 \, \mathrm{d}X_s^1, \quad P\text{-a.s.} \quad t \in [0, T']. \tag{8.16}$$

Moreover, (8.12), the two-dimensional Itô formula, the fact that $\langle X \rangle_t = \int_0^t \sigma^2 X_s^2 \, \mathrm{d}s$ and the PDE (8.9) for v give

$$\begin{aligned} V_t^H - V_0^H &= v(t, X_t^1) - v(0, X_0^1) \\ &= \int_0^t \frac{\partial v}{\partial t}(s, X_s^1) \, \mathrm{d}s + \int_0^t \frac{\partial v}{\partial x}(s, X_s^1) \, \mathrm{d}X_s^1 + \frac{1}{2} \int_0^t \frac{\partial^2 v}{\partial x^2}(s, X_s^1) \, \mathrm{d}\langle X \rangle_s \\ &= \int_0^t \frac{\partial v}{\partial t}(s, X_s^1) \, \mathrm{d}s + \int_0^t \frac{\partial v}{\partial x}(s, X_s^1) \, \mathrm{d}X_s^1 + \frac{1}{2} \int_0^t \frac{\partial^2 v}{\partial x^2}(s, X_s^1) \sigma^2 (X_s^1)^2 \, \mathrm{d}s \\ &= \int_0^t v_x(s, X_s^1) \, \mathrm{d}X_s^1, \quad P\text{-a.s.}, \quad t \in [0, T']. \end{aligned} \tag{8.17}$$

Subtracting the right-hand side of (8.17) from the right-hand side of (8.16) it follows that the continuous local Q-martingale $(\int_0^t (\vartheta_s^1 - v_x(s, X_s^1)) \, \mathrm{d}X_s)_{t \in [0,T]}$ is constant 0 Q-a.s. This implies that also its quadratic variation is constant 0 Q-a.s., and since $Q \approx P$ on $\mathcal{F}$ also P-a.s. It follows that

$$0 = \int_0^{T'} (\vartheta_s^1 - v_x(s, X_s^1))^2 \, \mathrm{d}\langle X^1 \rangle_s = \int_0^{T'} (\vartheta_s^1 - v_x(s, X_s^1))^2 \sigma^2 (X_s^1)^2 \, \mathrm{d}s \;\; P\text{-a.s.}$$

Taking the limit $T' \uparrow T$, monotone convergence gives

$$\int_0^T (\vartheta_s^1 - v_x(s, X_s^1))^2 \sigma^2 (X_s^1)^2 \, \mathrm{d}s = 0 \;\; P\text{-a.s.}$$

It follows that $(\vartheta_s^1 - v_x(s, X_s^1))^2 \sigma^2 (X_s^1)^2 = 0$ P-a.s. for Lebesgue-a.e. $s \in [0, T)$. Since σ^2 and $(X_s^1)^2$ are positive, it follows that $\vartheta_s^1 = v_x(s, X_s^1)$ P-a.s. for Lebesgue-a.e. $s \in [0, T)$. Since $\overline{\vartheta}$ was an arbitrary replication strategy for H, this gives uniqueness of the replication strategy (up to $(P \otimes \mathrm{d}t)$-nullsets). Moreover, this yields (8.13). □

Remark 8.29. One can show that the assumption that h is continuous in Corollary 8.28 can be dropped. All assertions remain true – apart from the fact that v is no longer continuous on $[0, T] \times (0, \infty)$. (It requires some more advanced concepts from PDE theory to interpret the terminal condition (8.10) in the right way in that setting.)

We illustrate the above result by calculating the value process and the hedging strategy of a European call option in the Black–Scholes model, which gives the celebrated *Black–Scholes formula*.

Example 8.30. Let $\overline{S} = (S_t^0, S_t^1)_{t \in [0,T]}$ be a Black–Scholes model with parameters $r, \mu \in \mathbb{R}$,

$\sigma, s_0^1 > 0$. Let C denote the undiscounted payoff of a call option on S^1 with strike K and maturity T, i.e., $C = (S_T^1 - K)^+$. The corresponding discounted contingent claim $H = C/S_T^0$ satisfies

$$H = (X_T^1 - K\exp(-rT)) = h(X_T^1),$$

where $h : (0, \infty) \to [0, \infty)$ is given by $h(x) = (x - K\exp(-rT))^+$. Let $Z \sim \mathcal{N}(0, 1)$ and fix $(t, x) \in [0, T] \times (0, \infty)$. By (8.11) and the change of variable $y = -z$, we obtain

$$\begin{aligned}
v(t,x) &= E\left[h\left(x\exp\left(\sigma\sqrt{T-t}Z - \frac{1}{2}\sigma^2(T-t)\right)\right)\right] \\
&= \int_{-\infty}^{\infty}\left(x\exp\left(\sigma\sqrt{T-t}z - \frac{1}{2}\sigma^2(T-t)\right) - K\exp(-rT))\right)^+ \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty}\exp(-rT)\left(x\exp\left(-\sigma\sqrt{T-t}y + rT - \frac{1}{2}\sigma^2(T-t)\right) - K\right)^+ \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)\,\mathrm{d}y.
\end{aligned}$$

Noting that $x\exp\left(-\sigma\sqrt{T-t}y + rT - \frac{1}{2}\sigma^2(T-t)\right) - K \geq 0$ if and only if

$$y \leq d_-(t,x) := \frac{1}{\sigma\sqrt{T-t}}\left(\log\frac{x}{K} + rT - \frac{1}{2}\sigma^2(T-t)\right),$$

and denoting the cdf of a standard normal distribution by Φ, we obtain

$$\begin{aligned}
v(t,x) &= \int_{-\infty}^{d_-(t,x)} x\exp\left(-\sigma\sqrt{T-t}y - \frac{1}{2}\sigma^2(T-t)\right)\exp\left(-\frac{y^2}{2}\right)\,\mathrm{d}y \\
&\quad - \int_{-\infty}^{d_-(t,x)}\exp(-rT)K\exp\left(-\frac{y^2}{2}\right)\,\mathrm{d}y \\
&= \int_{-\infty}^{d_-(t,x)} x\exp\left(-\frac{(y+\sigma\sqrt{T-t})^2}{2}\right)\,\mathrm{d}y - \exp(-rT)K\Phi(d_-(t,x)).
\end{aligned}$$

Now using the change of variable $w = y + \sigma\sqrt{T-t}$ and setting

$$d_+(t,x) := d_-(t,x) + \sigma\sqrt{T-t} = \frac{1}{\sigma\sqrt{T-t}}\left(\log\frac{x}{K} + rT + \frac{1}{2}\sigma^2(T-t)\right),$$

we obtain

$$\begin{aligned}
v(t,x) &= \int_{-\infty}^{d_+(t,x)} x\exp\left(-\frac{w^2}{2}\right)\,\mathrm{d}w - \exp(-rT)K\Phi(d_-(t,x)) \\
&= x\Phi(d_+(t,x)) - \exp(-rT)K\Phi(d_-(t,x)),
\end{aligned}$$

which is the *Black–Scholes formula.*

Next, denoting the pdf of standard normal distribution by φ, this implies that

$$v_x(t,x) = \Phi(d_+(t,x)) + \frac{1}{\sigma\sqrt{T-t}}\varphi(d_+(t,x)) - \frac{\exp(-rT)K}{\sigma\sqrt{T-t}x}\varphi(d_-(t,x)).$$

Noting that

$$\begin{aligned}\varphi(d_+(t,x)) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{d_+(t,x)^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(d_-(t,x)+\sigma\sqrt{T-t})^2}{2}\right)\\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{d_-(t,x)^2 + 2\sigma\sqrt{T-t}d_-(t,x)+\sigma^2(T-t)}{2}\right)\\ &= \frac{\exp(-rT)K}{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{d_-(t,x)^2}{2}\right) = \frac{\exp(-rT)K}{x}\varphi(d_-(t,x)),\end{aligned}$$

it follows that the *delta* of v is given by

$$v_x(t,x) = \Phi(d_+(t,x)).$$

We may conclude that the fair value process of C satisfies

$$V_t^C = v(t,X_t^1) = X_t^1\Phi(d_+(t,X_t^1)) - \exp(-rT)K\Phi(d_-(t,X_t^1)) \;\; P\text{-a.s.}, \quad t\in[0,T).$$

An explicit version of the hedging strategy is given by $\overline{\vartheta}^C \mathrel{\widehat{=}} (V_0^C, \vartheta^{1,C})$, where

$$\begin{aligned}V_0^C &= s_0^1\Phi(d_+(0,s_0^1)) - \exp(-rT)K\Phi(d_-(0,s_0^1)),\\ \vartheta_t^{1,C} &= v_x(t,X_t^1) = \Phi(d_+(t,X_t^1)), \quad t\in[0,T).\end{aligned}$$

Note that the hedging strategy is long-only (in the stock) and bounded above by 1.

Remark 8.31. Note that most textbooks look at the *undiscounted* value function $\tilde{v}(t,s)$ and consider $\tilde{d}_+(t,s)$ and $\tilde{d}_-(t,s)$, where s denotes the undiscounted stock price. For this reason, the formula looks slightly different. It is a good exercise to check that it is indeed the same. Also note that most practitioners parametrise the value function in terms of *time to maturity* $\tau := T-t$ and s.[72]

Exercise 8.32. ★☆☆ Let $\overline{S} = (S_t^0, S_t^1)_{t\in[0,T]}$ be a Black–Scholes model with parameters $r,\mu\in\mathbb{R}$, $\sigma, s_0^1 > 0$. Let P denote the undiscounted payoff of a put option on S^1 with strike K and maturity T, i.e., $P = (K - S_T^1)^+$. Calculate the discounted value process $V^P = (V_t^P)_{t\in[0,T]}$ as well as the corresponding hedging strategy $\overline{\vartheta}^P \mathrel{\widehat{=}} (V_0^P, \vartheta^{1,P})$. (*Hint:* Use put-call parity.)

The End

[72] The same remarks pertain to the calculation of delta or other *greeks*.

A Fundamental concepts of Probability Theory

In this chapter, we study some fundamental concepts from Measure Theory and Probability Theory that are foundational for various applications in Mathematical Finance.

A.1 Probability spaces

First we look at the most basic object in Probability Theory, a *probability space.* This has three components. The first component is a *sample space.*

Definition A.1. A *sample space* Ω is a (finite or infinite) set.

Each $\omega \in \Omega$ describes a possible "state of the world". Key examples are $\Omega = \{\omega_1, \ldots, \omega_N\}$ for $N \in \mathbb{N}$ (finite sample space), $\Omega = \{\omega_1, \omega_2, \ldots\}$ (countable sample space), and $\Omega = \mathbb{R}$.

The second component of a probability space is a *σ-algebra.*

Definition A.2. Given a sample space Ω, a *σ-algebra* $\mathcal{F}$ on Ω is a collection of subsets of Ω such that

(1) $\Omega \in \mathcal{F}$;

(2) $A \in \mathcal{F} \implies A^c = \Omega \setminus A \in \mathcal{F}$;

(3) $A_1, A_2, \ldots \in \mathcal{F} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is called a *measurable space* and each $A \in \mathcal{F}$ is called *$\mathcal{F}$-measurable* or an *$\mathcal{F}$-measurable* event.

Example A.3. Let $\mathcal{F}$ be a σ-algebra on Ω. Then

(a) the empty set $\emptyset$ is in $\mathcal{F}$. Indeed, this follows from (1) and (2) via $\emptyset = \Omega \setminus \Omega$.

(b) if $A_1, A_2, \ldots \in \mathcal{F}$, then the countable intersection $\bigcap_{n \in \mathbb{N}} A_n$ is in $\mathcal{F}$. Indeed, this follows from the de Morgan laws,[73] (2), and (3) via $\bigcap_{n \in \mathbb{N}} A_n = (\bigcup_{n \in \mathbb{N}} A_n^c)^c$.

(c) if $A_1, \ldots, A_n \in \mathcal{F}$, then the finite union $A_1 \cup A_2 \cup \ldots \cup A_n$ and the finite intersection $A_1 \cap A_2 \cap \ldots \cap A_n$ are in $\mathcal{F}$. Indeed, this follows from (a) and (3) and (a) and (b), respectively, by setting $A_{n+1}, A_{n+2}, \ldots := \emptyset$.

(d) If $A, B \in \mathcal{F}$, then $A \setminus B$ and $B \setminus A$ are in $\mathcal{F}$. Indeed, this follows from (2) and (c) via $A \setminus B = A \cap B^c$ and $B \setminus A = B \cap A^c$.

(e) if $A_1, A_2, \ldots \in \mathcal{F}$, then $\{A_n \text{ infinitely often}\} := \limsup_{n \to \infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$ is in $\mathcal{F}$. Indeed this follows from (3) and (b).

[73] The de Morgan's laws say that if $(A_n)_{n \in \mathbb{N}}$ is any collection of subsets of Ω, then $(\bigcup_{n \in \mathbb{N}} A_n)^c = \bigcap_{n \in \mathbb{N}} A_n^c$ and $(\bigcap_{n \in \mathbb{N}} A_n)^c = \bigcup_{n \in \mathbb{N}} A_n^c$.

(f) if $A_1, A_2, \ldots \in \mathcal{F}$, then $\{A_n \text{ eventually}\} := \liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k$ is in $\mathcal{F}$. Indeed, this follows from (b) and (3).

If Ω is finite (or countable), i.e., $\Omega = \{\omega_1, \ldots, \omega_N\}$ (or $\Omega = \{\omega_1, \omega_2, \ldots\}$), then the usual choice for a σ-algebra on Ω is $\mathcal{F} := 2^\Omega := \{A : A \subset \Omega\}$, the *power set* of Ω.

If Ω is uncountable, e.g. $\Omega = \mathbb{R}$, it turns out that the power set 2^Ω is "too big" to be chosen as σ-algebra.[74] For this reason, one uses instead the following procedure: One chooses a *generator* $\mathcal{A}$ of "good subsets" of Ω that one wants to be measurable. One then denotes by $\sigma(\mathcal{A})$ the smallest σ-algebra on Ω that contains $\mathcal{A}$. It is called the *σ-algebra generated by* $\mathcal{A}$. It is explicitly given by

$$\sigma(\mathcal{A}) = \bigcap_{\substack{\mathcal{G}\ \sigma\text{-algebra,}\\ \mathcal{A}\subset\mathcal{G}}} \mathcal{G}.$$

Note that different generators $\mathcal{A}$ may generate the same σ-algebra; one usually chooses a "small" generator or a generator with "good properties".

Example A.4. (a) If $\Omega = \mathbb{R}$, one wants that all open sets $\mathcal{O}_\mathbb{R}$ in $\mathbb{R}$ are measurable. One then sets $\mathcal{B}_\mathbb{R} := \sigma(\mathcal{O}_\mathbb{R})$ and calls this the *Borel σ-algebra* on $\mathbb{R}$. One can show that $\mathcal{B}_\mathbb{R}$ contains all closed and open sets in $\mathbb{R}$ and that is is also generated by the generator $\mathcal{A}$, where $\mathcal{A}$ contains all (closed) sets of the form $(-\infty, a]$ for $a \in \mathbb{R}$.

(b) More generally, if Ω is a subset of $\mathbb{R}^d$ for $d \geq 1$, one denotes the open sets in Ω by $\mathcal{O}_\Omega$, and sets $\mathcal{B}_\Omega := \sigma(\mathcal{O}_\Omega)$ and calls this the *Borel σ-algebra* on Ω.[75]

The third component of a probability space is a *probability measure.*

Definition A.5. Given a measurable space $(\Omega, \mathcal{F})$, a *probability measure* P on $(\Omega, \mathcal{F})$ is a map $\mathcal{F} \to [0, 1]$ such that

(1) $P[\emptyset] = 0$ and $P[\Omega] = 1$;

(2) $A_1, A_2, \ldots \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j \implies P[\bigcup_{n\in\mathbb{N}} A_n] = \sum_{n=1}^\infty P[A_n]$.

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

Remark A.6. The properties (1) and (2) in Definition A.5 are called the *axioms of Kolmogorov* after the Russian mathematician A. Kolmogorov (1903 - 1987). Property (2) is referred to as *σ-additivity*, where the σ stands for "countable".

Example A.7. (a) Let $(\Omega, \mathcal{F})$ be a measurable space and suppose that $\mathcal{F}$ contains all *elementary events*, i.e., $\{\omega\} \in \mathcal{F}$ for all $\omega \in \Omega$. Fix $\omega^* \in \Omega$ and define the map $P : \mathcal{F} \to [0, 1]$

[74] See [4, Theorem 1.5] for a precise formulation of this statement.

[75] Even more generally, if (Ω, τ) is a topological space, one sets $B_\Omega := \sigma(\tau)$, and calls this the *Borel σ-algebra* on Ω.

by

$$P[A] = \begin{cases} 1 & \text{if } \omega^* \in A, \\ 0 & \text{if } \omega^* \notin A. \end{cases}$$

One can check that P is a probability measure on $(\Omega, \mathcal{F})$. It is called the *Dirac measure* for ω^* and often denoted by δ_{ω^*}. The Dirac measure for ω^* models the "deterministic case", where we know with probability 1 that the state of the world will be ω^*.[76]

(b) Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ for some $N \in \mathbb{N}$ and $\mathcal{F} = 2^\Omega$. Define the map $P : \mathcal{F} \to [0, 1]$ by

$$P[A] = \frac{|A|}{|\Omega|},$$

where $|A|$ denotes the number of elements in A. One can check that P is a probability measure on $(\Omega, \mathcal{F})$. It is called the *discrete uniform distribution* on Ω. For $N = 2$ (and $\omega_1 = H$ and $\omega_2 = T$), this is a good model for the flipping of a fair coin, and for $N = 6$ (and $\omega_1 = 1, \ldots, \omega_6 = 6$), this is a good model for the rolling of a fair die.

The following result collects some fundamental rules for calculating probabilities.

Proposition A.8. *Let $(\Omega, \mathcal{F}, P)$ be a probability space.*

(a) *If $A \in \mathcal{F}$, then*

$$P[A^c] = P[\Omega \setminus A] = 1 - P[A].$$

(b) *If $A \subset B \in \mathcal{F}$, then*

$$P[B] = P[A] + P[B \setminus A] \geq P[A]. \tag{A.1}$$

(c) *If $A, B \in \mathcal{F}$, then*

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] \leq P[A] + P[B].$$

(d) *If $A_1 \subset A_2 \subset \cdots \in \mathcal{F}$, then*

$$P\left[\bigcup_{n \in \mathbb{N}} A_n\right] = \lim_{n \to \infty} P[A_n].$$

(e) *If $A_1 \supset A_2 \supset \cdots \in \mathcal{F}$, then*

$$P\left[\bigcap_{n \in \mathbb{N}} A_n\right] = \lim_{n \to \infty} P[A_n].$$

[76] Note that unless $\Omega = \{\omega^*\}$, it is false to say that we are sure that the state of the world will be ω^*; we can only say that the state of the world will be ω^* P-almost surely; cf. Definition A.9 below.

(f) *If* $A_1, A_2, \ldots \in \mathcal{F}$, *then*

$$P\left[\bigcup_{n\in\mathbb{N}} A_n\right] \le \sum_{n=1}^{\infty} P[A_n].$$

Proof. We only prove parts (a), (d) and (f). The other parts are left as an exercise.

(a) Set $B_1 := A, B_2 := A^c, B_3 := \emptyset, B_4 := \emptyset, \ldots$. Then $B_1, B_2, \ldots \in \mathcal{F}$ with $B_i \cap B_j = \emptyset$ for $i \neq j$. Moreover, $\bigcup_{n\in\mathbb{N}} B_n = \Omega$. Hence, σ-additivity of P and together with $P[\emptyset] = 0$ and $P[\Omega] = 1$ give

$$1 = P[\Omega] = P\left[\bigcup_{n\in\mathbb{N}} B_n\right] = \sum_{n\in\mathbb{N}} P[B_n] = P[A] + P[A^c] + 0 = P[A] + P[A^c].$$

Rearranging yields (a).

(d) Set $B_1 := A_1, B_2 := A_2 \setminus A_1, B_3 := A_3 \setminus A_2, \ldots$. Then $B_1, B_2, \ldots \in \mathcal{F}$ with $B_i \cap B_j = \emptyset$ for $i \neq j$. Moreover, $\bigcup_{k=1}^{n} B_k = A_n$ for all $n \in \mathbb{N}$ and $\bigcup_{k\in\mathbb{N}} B_k = \bigcup_{k\in\mathbb{N}} A_k$. Hence, σ-additivity of P yields

$$P\left[\bigcup_{k\in\mathbb{N}} A_k\right] = P\left[\bigcup_{k\in\mathbb{N}} B_k\right] = \sum_{k=1}^{\infty} P[B_k] = \lim_{n\to\infty} \sum_{k=1}^{n} P[B_k] = \lim_{n\to\infty} P\left[\bigcup_{k=1}^{n} B_k\right] = \lim_{n\to\infty} P[A_n].$$

(f) Set $B_1 := A_1, B_2 := A_1 \cup A_2, B_3 := A_1 \cup A_2 \cup A_3, \ldots$ Then $B_1 \subset B_2 \subset \cdots \in \mathcal{F}$ and $\bigcup_{k\in\mathbb{N}} B_k = \bigcup_{k\in\mathbb{N}} A_k$. Moreover, $P[B_n] \le \sum_{k=1}^{n} P[A_k]$ by repeated application of part (c). Hence, by part (d),

$$P\left[\bigcup_{k\in\mathbb{N}} A_k\right] = P\left[\bigcup_{k\in\mathbb{N}} B_k\right] = \lim_{n\to\infty} P[B_n] \le \lim_{n\to\infty} \sum_{k=1}^{n} P[A_k] = \sum_{k=1}^{\infty} P[A_k]. \qquad \square$$

Events with probability zero or one are of particular importance because of the obvious interpretation of "impossible" and "sure" events. However, this is not fully correct as $P[A] = 0$ does in general not imply that $A = \emptyset$, and $P[A] = 1$ does in general not imply that $A = \Omega$. The following definition makes this precise.

Definition A.9. Let $(\Omega, \mathcal{F}, P)$ be a probability space.

(a) $N \in \mathcal{F}$ is called a P-*null set* if $P[N] = 0$.

(b) Let $\mathfrak{E}(\omega)$ be a property that a state of the world $\omega \in \Omega$ can have or not have. We say that $\mathfrak{E}$ holds P-*almost surely* if there exists a P-nullset $N \in \mathcal{F}$ such that $\mathfrak{E}$ holds for all $\omega \in \Omega \setminus N$.

A.2 Random variables

The next fundamental object we are looking at is the concept of a *random variable*, which is a map between measurable spaces with a special property, called *measurability.*

Definition A.10. Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces and $X : \Omega \to \Omega'$ a map.

(a) The *preimage* of a set $A' \in \Omega'$ under X is denoted by

$$X^{-1}(A') := \{X \in A'\} := \{\omega \in \Omega : X(\omega) \in A'\}.$$

(b) the collection of sets

$$\sigma(X) := \Big\{\{X \in A'\} : A' \in \mathcal{F}'\Big\}$$

is called the *σ-algebra generated by X.*

(c) X is called *measurable with respect to $\mathcal{F}$ and $\mathcal{F}'$* (or shorter $\mathcal{F}$-$\mathcal{F}'$-measurable) if

$$X^{-1}(A') := \{X \in A'\} := \{\omega \in \Omega : X(\omega) \in A'\} \in \mathcal{F}, \quad \text{for all } A' \in \mathcal{F}'. \tag{A.2}$$

In this case, if $\Omega' = \mathbb{R}$ and $\mathcal{F}' = \mathcal{B}_{\mathbb{R}}$, then X is called an *$\mathcal{F}$-measurable random variable.*[77] If $\Omega' = \mathbb{R}^d$ and $\mathcal{F}' = \mathcal{B}_{\Omega'}$ for $d \geq 2$, we call X an *$\mathcal{F}$-measurable random vector.*

Intuitively, $\sigma(X)$ contains all the "information" about X. Clearly, X is measurable with respect to $\mathcal{F}$ and $\mathcal{F}'$ if and only if $\sigma(X) \subset \mathcal{F}$, i.e., $\mathcal{F}$ contains all the information about X.

Example A.11. Let $(\Omega, \mathcal{F})$ be a measurable space and $A \in \mathcal{F}$ an $\mathcal{F}$-measurable event. Then the *indicator function* $\mathbf{1}_A$, defined by

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \in A^c, \end{cases}$$

is an $\mathcal{F}$-measurable random variable. Indeed, let $A' \in \mathcal{B}_{\mathbb{R}}$. Then with $X = \mathbf{1}_A$, we have

$$X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\} = \begin{cases} \emptyset & \text{if } 0, 1 \notin A', \\ A & \text{if } 1 \in A', 0 \notin A', \\ A^c & \text{if } 0 \in A', 1 \notin A', \\ \Omega & \text{if } 0, 1 \in A', \end{cases}$$

Hence, $\sigma(X) = \{\emptyset, A, A^c, \Omega\}$ and since $A \in \mathcal{F}$, it follows that $\sigma(X) \subset \mathcal{F}$ and X is $\mathcal{F}$-measurable.

[77] Also if Ω' is a Borel subset of $\mathbb{R}$ and $\mathcal{F}' = \mathcal{B}_{\Omega}$, we say that X is an $\mathcal{F}$-measurable random variable valued in Ω'. Note that each $\mathcal{F}$-measurable random variable valued in Ω' is also an $\mathcal{F}$-measurable random variable (valued in $\mathbb{R}$) so that we do not need to worry too much about the exact domain. The same holds for random vectors.

Whereas Definition A.10 is important from a theoretical perspective, it is almost useless for checking in practice that a given function $X : \Omega \to \Omega'$ is $\mathcal{F}$-$\mathcal{F}'$-measurable because we usually do not have a good description of all $A' \in \mathcal{F}'$. But fortunately, one can show that it suffices to check (A.2) for all A' in a generator $\mathcal{A}'$ of $\mathcal{F}'$, which in general is much smaller than $\mathcal{F}'$. This is the content of the following result; for a proof see [9, Theorem 1.81].

Theorem A.12. *Let* $(\Omega, \mathcal{F})$ *and* $(\Omega', \mathcal{F}')$ *be measurable spaces and* $X : \Omega \to \Omega'$ *a map. Suppose that* $\mathcal{F}' = \sigma(\mathcal{A}')$.

(a) *We have*

$$\sigma(X) = \sigma\left(\{\{X \in A'\} : A' \in \mathcal{A}'\}\right).$$

(b) X *is measurable with respect to* $\mathcal{F}$ *and* $\mathcal{F}'$ *if and only if*

$$X^{-1}(A') := \{X \in A'\} := \{\omega \in \Omega : X(\omega) \in A'\} \in \mathcal{F} \quad \text{for all } A' \in \mathcal{A}'.$$

We note two important corollaries.

Corollary A.13. *Let* $(\Omega, \mathcal{F})$ *be a measurable space. Then a function* $X : \Omega \to \mathbb{R}$ *is an* $\mathcal{F}$*-measurable random variable if and only if*

$$X^{-1}((-\infty, x]) = \{X \le x\} := \{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}, \quad \text{for all } x \in \mathbb{R}. \tag{A.3}$$

Proof. Set $\mathcal{A}' := \{(-\infty, x] : x \in \mathbb{R}\}$. Note that (A.3) states that $X^{-1}(A') \in \mathcal{F}$ for all $A' \in \mathcal{A}'$. Hence, the claim follows from (A.12) and the fact that $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A}')$ by Example A.4(a). □

Corollary A.14. *Let* Ω *be a subset of* $\mathbb{R}^d$ *and* $\mathcal{F} = \mathcal{B}_\Omega$. *If* $f : \Omega \to \mathbb{R}$ *is continuous, then* f *is* $\mathcal{F}$*-measurable.*

Proof. By Corollary A.13, it suffices to show that $f^{-1}((-\infty, x])$ is in $\mathcal{B}_\Omega$ for all $x \in \mathbb{R}$. So fix $x \in \mathbb{R}$. As f is continuous, the preimage of the closed set $(-\infty, x]$ is again closed. As $\mathcal{B}_\Omega$ contains all closed sets in Ω, it follows that $f^{-1}((-\infty, x]) \in \mathcal{B}_\Omega$. □

We proceed to show that the composition of measurable maps is again measurable.

Theorem A.15. *Let* $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$, *and* $(\Omega'', \mathcal{F}'')$ *be measurable spaces. Let* $f : \Omega \to \Omega'$ *be* $\mathcal{F}$-$\mathcal{F}'$*-measurable and* $g : \Omega' \to \Omega''$ *be* $\mathcal{F}'$-$\mathcal{F}''$*-measurable. Then* $h : \Omega \to \Omega''$, *defined by* $h = g \circ f$ *is* $\mathcal{F}$-$\mathcal{F}''$*-measurable.*[78]

[78] Recall that $g \circ f$ is defined by $(g \circ f)(\omega) = g(f(\omega))$.

Proof. Let $A'' \in \mathcal{F}''$. Set $A' := g^{-1}(A'')$. Then

$$\begin{aligned} h^{-1}(A'') &= \{\omega \in \Omega : h(\omega) \in A''\} = \{\omega \in \Omega : g(f(\omega)) \in A''\} \\ &= \{\omega \in \Omega : f(\omega) \in g^{-1}(A'')\} = \{\omega \in \Omega : f(\omega) \in A'\} \\ &= f^{-1}(A'). \end{aligned}$$

Since g is $\mathcal{F}'$-$\mathcal{F}''$-measurable, it follows that $A' = g^{-1}(A'') \in \mathcal{F}'$, and since f is $\mathcal{F}$-$\mathcal{F}'$-measurable, it follows that $f^{-1}(A') \in \mathcal{F}$. Hence h is $\mathcal{F}$-$\mathcal{F}''$-measurable. □

Our next result states that sums and products of random variables are again random variables; for a proof see [9, Theorem 1.91]

Theorem A.16. *Let $(\Omega, \mathcal{F})$ be a measurable space and X_1 and X_2 be $\mathcal{F}$-measurable random variables. Then $X_1 + X_2$, $X_1 - X_2$, X_1X_2, and X_1/X_2 are again $\mathcal{F}$-measurable random variables.*[79]

Finally, we show that *countable* suprema and infima or random variables are again measurable.[80] To this end, we need to extend the real line $\mathbb{R}$ by the points $-\infty$ and $+\infty$. Thus, we set

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} \quad \text{and} \quad \mathcal{B}_{\overline{\mathbb{R}}} := \sigma\left(\{[-\infty, x] : x \in \mathbb{R}\}\right).$$

One can check that $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\overline{\mathbb{R}}}$, so that every real-valued random variable X can be in a canonical way identified with an $\overline{\mathbb{R}}$-valued measurable map.[81] For this reason, we will not always carefully distinguish between $\mathbb{R}$-valued and $\overline{\mathbb{R}}$-valued measurable maps and call both random variables.

Theorem A.17. *Let $(\Omega, \mathcal{F})$ be a measurable space and $X_1, X_2, \dots$ be $\mathcal{F}$-measurable $\overline{\mathbb{R}}$-valued random variables. Then the following maps are also $\mathcal{F}$-measurable:*

(a) $\sup_{n\in\mathbb{N}} X_n$.

(b) $\inf_{n\in\mathbb{N}} X_n$.

(c) $\limsup_{n\to\infty} X_n$.

(d) $\liminf_{n\to\infty} X_n$.

Proof. We only prove (a) and (c); the proofs of (b) and (d) are similar.

(a) Let $x \in \mathbb{R}$. Then by the fact that each of the X_n are $\mathcal{F} - \mathcal{B}_{\overline{\mathbb{R}}}$-measurable, we have

$$\left\{\sup_{n\in\mathbb{N}} X_n \in [-\infty, x]\right\} = \bigcap_{n\in\mathbb{N}} \{X_n \in [-\infty, x]\} \in \mathcal{F}.$$

[79] Here, we agree that $x/0 := 0$ for all $x \in \mathbb{R}$, which is a standard convention in measure theory.
[80] Note that *uncountable* suprema and infima are in general not measurable.
[81] See [9, Corollary 1.87] for a precise formulation of this statement.

Now the claim follows from Theorem A.12(b).

(c) For any $n \in \mathbb{N}$ set $Y_n := \sup_{m \geq n} X_m$. Then each Y_n is $\mathcal{F}$-measurable by part (a). Hence, $\limsup_{n\to\infty} X_n = \inf_{n\in\mathbb{N}} Y_n$ is $\mathcal{F}$-measurable by part (b). $\square$

A.3 Distribution of random variables: univariate case

The definition of a random variable does not mention any probability measure P at all. If we add a probability measure P, we get some further notions.

Definition A.18. Let $(\Omega, \mathcal{F}, P)$ be a probability space and X an $\mathcal{F}$-measurable random variable.

(a) The *distribution* (or *law* or *image*) of X under P is the probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ defined by

$$P_X[B] := P[X \in B], \quad B \in \mathcal{B}_\mathbb{R}.$$

(b) The *distribution function* (or *cumulative distribution function* (cdf)) of X under P is the map $F_X : \mathbb{R} \to [0,1]$ given by

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

The following result lists the three defining properties of a distribution function; its proof is left as an exercise.

Lemma A.19. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X an $\mathcal{F}$-measurable random variable. Then the distribution function $F_X : \mathbb{R} \to [0,1]$ has the following properties:*

(a) *F_X is nondecreasing.*

(b) *F_X is right-continuous.*

(c) $\lim\limits_{x\to-\infty} F_X(x) = 0$ *and* $\lim\limits_{x\to\infty} F_X(x) = 1$.

The above lemma shows that for each random variable X, there exists a function F that is nondecreasing, right-continuous, and with limits 0 at $-\infty$ and 1 at $+\infty$, respectively. The next result show that also the converse is true. For each function F with these three properties, there exists a random variable X with distribution function F; for a proof see [9, Theorem 1.104].

Theorem A.20. *Let $F : \mathbb{R} \to [0,1]$ be a function that is nondecreasing, right-continuous and satisfies $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X such that $F_X = F$.*[82]

[82] The proof of Theorem A.20 shows that one can always take $\Omega = \mathbb{R}$, $\mathcal{F} := \mathcal{B}_\mathbb{R}$ and $X : \Omega \to \mathbb{R}$ given by $X(\omega) = \omega$, i.e., X is the identity map.

The next definition looks at the concept that two random variables X and Y have the same distribution.

Definition A.21. Let $(\Omega, \mathcal{F}, P)$ be a probability space and X and Y $\mathcal{F}$-measurable random variables.[83] Then X and Y are said to be *identically distributed* if

$$P_X = P_Y.$$

The next result shows that the distribution function of a random variable does indeed describe the whole distribution; for a proof see [5, Theorem 7.1].

Theorem A.22. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X and Y $\mathcal{F}$-measurable random variables. Then the following are equivalent:*

(1) *X and Y are identically distributed.*

(2) *$F_X = F_Y$.*

We proceed to introduce the important class of *discrete* distributions.

Definition A.23. A real-valued random variable X (or more precisely its distribution) is called *discrete* if there exists a *finite* or *countable* set $B \in \mathcal{B}_{\mathbb{R}}$ such that $P^X[B] = P[X \in B] = 1$. In this case, the *probability mass function* (pmf) $p_X : \mathbb{R} \to [0, 1]$ of X is given by

$$p_X(x) = P_X[\{x\}] = P[X = x], \quad x \in \mathbb{R}.$$

Remark A.24. (a) If X is discrete, is straightforward to check that

$$p_X(x) = 0 \text{ for } x \in B^c \quad \text{and} \quad \sum_{x \in B} p_X(x) = 1. \tag{A.4}$$

More generally, for any Borel set $A \in \mathcal{B}_{\mathbb{R}}$, we have the formula

$$P_X[A] = P[X \in A] = \sum_{x \in A} p_X(x).$$

Note that due to (A.4), we always sum over a countable set and this sum is finite and bounded above by 1.

(b) One can show that a random variable X has a discrete distribution if and only if its distribution function is piecewise constant, i.e., for each $x \in \mathbb{R}$, there is $\varepsilon > 0$ such that $F_X(y) = F_X(x)$ for all $y \in [x, x + \varepsilon]$. Moreover, $p_X(x) = F_X(x) - F_X(x-)$ for all $x \in \mathbb{R}$, where $F_X(x-) := \lim_{y \uparrow\uparrow x} F_X(y) := \lim_{y \to x, y < x} F_X(y)$ denotes the *left limit* of F_X at x.[84] This

[83] More generally, X and Y might be defined on different probability spaces.
[84] This left limit always exists because F_X is nondecreasing.

together with Theorem A.22 also shows that the distribution of a discrete random variable is uniquely described by its pmf.

We proceed to list some important examples of discrete distributions.

Example A.25. A discrete random variable X is said to have a

(a) *Bernoulli distribution* with parameter $p \in [0,1]$ if its pmf is given by

$$p_X(x) = \begin{cases} p & \text{if } x = 1, \\ 1-p & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases}$$

(b) *binomial distribution* with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ if its pmf is given by

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x \in \{0,\dots,n\}, \\ 0 & \text{otherwise.} \end{cases}$$

(c) *geometric distribution* with parameter $p \in (0,1)$ if its pmf is given by[85]

$$p_X(x) = \begin{cases} p(1-p)^x & \text{if } x \in \mathbb{N}_0, \\ 0 & \text{otherwise.} \end{cases}$$

(d) *Poisson distribution* with parameter $\lambda > 0$ if its pmf is given by

$$p_X(x) = \begin{cases} \mathrm{e}^{-\lambda} \dfrac{\lambda^x}{x!} & \text{if } x \in \mathbb{N}_0, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we introduce the equally important class of *continuous* distributions.

Definition A.26. A real-valued random variable X (or more precisely its distribution) is called (absolutely) *continuous* if there exists a nonnegative, measurable function $f_X : \mathbb{R} \to [0,\infty)$ satisfying

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,\mathrm{d}y, \quad x \in \mathbb{R}.$$

In this case, the function f_X is called the *probability density function* (pdf) of X.

[85]Warning: In parts of the literature, the geometric distribution is shifted by one to the right, i.e., it is a distribution on $\mathbb{N}$.

Remark A.27. (a) One can show that if X is continuous with pdf f_X, then for any Borel set $A \in \mathcal{B}_{\mathbb{R}}$, we have the formula

$$P_X[A] = P[X \in A] = \int_A f_X(y)\,\mathrm{d}y. \tag{A.5}$$

This together with Theorem A.22 and fundamental properties of the (Lebesgue) integral show that the distribution of a random variable is uniquely characterised by its pdf (up to Lebesgue-null sets).

(b) Note that there are many random variables which have neither a discrete nor a continuous distribution.

We proceed to list some important examples of continuous distributions.

Example A.28. A continuous random variable X is said to have a

(a) *uniform distribution* on (a, b), where $-\infty < a < b < \infty$ if its pdf is given by

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } x \in (a,b), \\ 0 & \text{otherwise.} \end{cases}$$

We then also write $X \sim \mathcal{U}(a, b)$.

(b) *exponential distribution* with rate parameter $\lambda > 0$ if its pdf is given by

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

(c) *Normal distribution* with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if its pdf is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

We then also write $X \sim \mathcal{N}(\mu, \sigma^2)$.

A.4 Distribution of random variables: multivariate case

In this section, we study the distribution of (multivariate) random vectors. This is very similar to the case of (univariate) random variables, and so we will be brief.

Definition A.29. Let $X_1, \dots, X_N$ be random variables on some probability space $(\Omega, \mathcal{F}, P)$.

(a) The *joint distribution function* of $X_1, \dots, X_N$ under P is the map $F_{(X_1,\dots,X_N)} : \mathbb{R}^N \to [0, 1]$ given by

$$F_{(X_1,\dots,X_N)}(x_1, \dots, x_N) = P[X_1 \le x_1, \dots, X_N \le x_N], \quad x_1, \dots, x_N \in \mathbb{R}.$$

(b) $X_1, \dots, X_N$ are called *jointly continuous* if there is a measurable function $f_{(X_1,\dots,X_N)} : \mathbb{R}^N \to [0, \infty)$ such that

$$\begin{aligned} &F_{(X_1,\dots,X_N)}(x_1, \dots, x_N) \\ &\quad = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} f_{(X_1,\dots,X_N)}(y_1, \dots, y_N)\, \mathrm{d}y_N \cdots \mathrm{d}y_1, \quad x_1, \dots, x_N \in \mathbb{R}. \end{aligned}$$

In this case, the function $f_{(X_1,\dots,X_N)}$ is called the *joint probability density function* (joint pdf) of $X_1, \dots, X_N$.

Remark A.30. (a) To simplify the notation one often sets $X := (X_1, \dots, X_N)$ and writes $F_X(x)$ and $f_X(x)$ for $F_{(X_1,\dots,X_N)}(x_1, \dots, x_N)$ and $f_{(X_1,\dots,X_N)}(x_1, \dots, x_N)$, respectively, where $x = (x_1, \dots, x_N)$.

(b) If $X = (X_1, \dots, X_N)$ is a random vector, then the one-dimensional distributions of $X_1, \dots, X_N$ are often called *marginal distributions*. If $X_1, \dots, X_N$ are jointly continuous, one can show that the marginal pdfs can be obtained from the joint pdf by "integrating out" the other random variables. For example if X_1, X_2 are jointly continuous with joint pdf $f_{(X_1,X_2)}$, then the marginal pdfs of X_1 and X_2 are given by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x_1, y_2)\, \mathrm{d}y_2 \quad \text{and} \quad f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(y_1, x_2)\, \mathrm{d}y_1$$

That these integrals are well defined and measurable is the content of *Fubini's theorem*, see Section A.11 below.

The key example of a multivariate distribution is the *multivariate normal distribution.*

Example A.31. A random vector $X = (X_1, \dots, X_N)$ is said to have a *multivariate normal distribution* with *mean vector* $\mu \in \mathbb{R}^N$ and *covariance matrix* $\Sigma \in \mathbb{R}^{N \times N}$, where Σ is symmetric and positive definite, if $X_1, \dots, X_N$ are jointly continuous with joint pdf

$$f_X(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1} (x - \mu) \right).$$

A.5 Conditional probabilities

In this section, we briefly look at the elementary notion of conditional probabilities.

To motivate this concept, suppose we are given a probability space $(\Omega, \mathcal{F}, P)$ and we have been informed that $A \in \mathcal{F}$ has occurred. We want to find a new probability measure $P[\cdot \,|\, A]$ on $(\Omega, \mathcal{F})$ that takes this information into account. Clearly, this new probability measure should be consistent with the old probability measure P in the sense that $P[\cdot \,|\, A]$ is proportional to P. Moreover, since we already know that A has occurred, one should require that $P[A \,|\, A] = 1$. It is then not difficult to check that these two properties uniquely determine $P[\cdot \,|\, A]$. The answer is given by the following definition.

Definition A.32. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A \in \mathcal{F}$ with $P[A] > 0$. Then for any $B \in \mathcal{F}$, the *conditional probability of B given* A is denoted by $P[B \,|\, A]$ and defined by

$$P[B \,|\, A] := \frac{P[A \cap B]}{P[A]}. \tag{A.6}$$

It is straightforward to check that $P[\cdot \,|\, A]$ is again a probability measure on $(\Omega, \mathcal{F})$.

The following result lists two elementary facts about conditional probabilities.

Theorem A.33. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{I}$ a finite or countable index set,*[86] *and $(A_i)_{i \in \mathcal{I}}$ an $\mathcal{F}$-measurable partition of Ω, i.e., each A_i is $\mathcal{F}$-measurable, $\bigcup_{i \in \mathcal{I}} A_i = \Omega$ and $A_i \cap A_j = \emptyset$ for $i \neq j$. Suppose that $P[A_i] > 0$ for each $i \in \mathcal{I}$. Then*

(a) *For every $B \in \mathcal{F}$, we have the* law of total probability

$$P[B] = \sum_{i \in \mathcal{I}} P[B \,|\, A_i] P[A_i].$$

(b) *For every $B \in \mathcal{F}$ with $P[B] > 0$ and each $k \in \mathcal{I}$, we have* Bayes' formula

$$P[A_k \,|\, B] = \frac{P[B \,|\, A_k] P[A_k]}{\sum_{i \in \mathcal{I}} P[B \,|\, A_i] P[A_i]}.$$

Proof. (a) The definition of conditional probabilities and the σ-additivity of P give

$$\sum_{i \in \mathcal{I}} P[B \,|\, A_i] P[A_i] = \sum_{i \in \mathcal{I}} P[B \cap A_i] = P[B].$$

(b) The definition of conditional probabilities and part (a) give

$$\frac{P[B \,|\, A_k] P[A_k]}{\sum_{i \in \mathcal{I}} P[B \,|\, A_i] P[A_i]} = \frac{P[A_k \cap B]}{P[B]} = P[A_k \,|\, B]. \qquad \square$$

A.6 Independence

In this section, we study the key concept of (stochastic) *independence*. We start by looking at independence of families of events and then generalise this to families of σ-algebras and families of random variables.

To motivate this concept, suppose we are given a probability space $(\Omega, \mathcal{F}, P)$ and two $\mathcal{F}$-measurable events A, B with $P[A], P[B] > 0$. Then intuitively A and B are (stochastically) *independent* if the probability assigned to A is not influenced by the information that B has occurred and vice versa. In formulas, this means that

$$P[A] = P[A \,|\, B] \quad \text{and} \quad P[B] = P[B \,|\, A]. \tag{A.7}$$

[86] We always implicitly assume that index sets are nonempty.

Using the definition of conditional probabilities in (A.6), it is straightforward to check that (A.7) is equivalent to $P[A \cap B] = P[A]P[B]$. The following definition generalises this idea of stochastic independence from two to many events.

Definition A.34. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{I}$ an arbitrary index set and $(A_i)_{i \in \mathcal{I}}$ a family of events in $\mathcal{F}$. Then the A_i are said to be (stochastically) *independent* if for any finite set of distinct indices $i_1, \ldots, i_n \in \mathcal{I}$,

$$P[A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}] = \prod_{k=1}^{n} P[A_{i_k}]. \tag{A.8}$$

Note that independence is stronger than pairwise independence, i.e., requiring (A.8) only for all disjoint i_1 and i_2.

We now "lift" the definition of independence from families of events to families of σ-algebras.

Definition A.35. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{I}$ an arbitrary index set and $(\mathcal{G}_i)_{i \in \mathcal{I}}$ a family of sub-σ-algebras of $\mathcal{F}$.[87] Then the $\mathcal{G}_i$ are said to be *independent* if for any finite set of distinct indices $i_1, \ldots, i_n \in I$ and any events $A_{i_1} \in \mathcal{G}_{i_1}, \ldots, A_{i_n} \in \mathcal{G}_{i_n}$,

$$P[A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}] = \prod_{k=1}^{n} P[A_{i_k}].$$

Next, we want to define a notion of independence for random variables. We do this by requiring that their generated σ-algebras (cf. Definition A.10) are independent.

Definition A.36. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{I}$ an arbitrary index set and $(X_i)_{i \in \mathcal{I}}$ a family of $\mathcal{F}$-measurable random variables. Then the X_i are said to be *independent* if their generated σ-algebras $\sigma(X_i)$, $i \in \mathcal{I}$, are independent.

The following result shows that we need to check independence only for certain generators of a σ-algebra; for a proof see [9, Theorem 2.13].

Theorem A.37. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{I}$ an arbitrary index set and $(\mathcal{A}_i)_{i \in \mathcal{I}}$ a family of classes of events that are independent, i.e., for any finite set of distinct indices $i_1, \ldots, i_n \in \mathcal{I}$ and any events $A_{i_1} \in \mathcal{A}_{i_1}, \ldots, A_{i_n} \in \mathcal{A}_{i_n}$,*

$$P[A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}] = \prod_{k=1}^{n} P[A_{i_k}].$$

Moreover, suppose that each each $\mathcal{A}_i$ is closed under intersections*, i.e., if $A, B \in \mathcal{A}_i$ then also $A \cap B \in \mathcal{A}_i$.*

(a) *The generated σ-algebras $\sigma(\mathcal{A}_i)$ are also independent.*

[87] This means that each $\mathcal{G}_i$ is a σ-algebra on Ω and $\mathcal{G}_i \subset \mathcal{F}$.

(b) *If $\mathcal{K}$ is an arbitrary index set and $(I_k)_{k\in\mathcal{K}}$ a partition of $\mathcal{I}$ in mutually disjoint "blocks", then the generated σ-algebras $\sigma\left(\bigcup_{j\in I_k}\mathcal{A}_j\right)$ are also independent.*

We note the following important corollary for the case of random variables. Its proof is left as an exercise.

Corollary A.38. *Let $(\Omega,\mathcal{F},P)$ be a probability space and $X_1,\dots,X_N$ $\mathcal{F}$-measurable random variables. Then $X_1,\dots,X_N$ are independent if and only if*

$$F_{(X_1,\dots,X_N)}(x_1,\dots,x_N)=\prod_{n=1}^{N}F_{X_n}(x_n),\quad \textit{for all } x_1,\dots,x_N\in\mathbb{R}.$$

Moreover, if $X_1,\dots,X_N$ are jointly continuous and have a continuous joint pdf $f_{(X_1,\dots,X_N)}$ and continuous marginal pdfs $f_{X_1},\dots,f_{X_N}$, then $X_1,\dots,X_N$ are independent if and only if

$$f_{(X_1,\dots,X_N)}(x_1,\dots,x_N)=\prod_{n=1}^{N}f_{X_n}(x_n),\quad \textit{for all } x_1,\dots,x_N\in\mathbb{R}.$$

As another application of Theorem A.37, we show that if a family $(A_i)_{i\in\mathcal{I}}$ of events is independent, then so is the family of their complements $(A_i^c)_{i\in\mathcal{I}}$.

Example A.39. Let $(\Omega,\mathcal{F},P)$ be a probability space, $\mathcal{I}$ an arbitrary index set and $(A_i)_{i\in\mathcal{I}}$ a family of independent events in $\mathcal{F}$. Then also $(A_i^c)_{i\in\mathcal{I}}$ is a family of independent events. Indeed, set

$$\mathcal{A}_i:=\{A_i\},\quad i\in\mathcal{I}.$$

Then the $\mathcal{A}_i$ are trivially closed under intersection and independent because the A_i are. Moreover, by Example A.11, it follows that $\mathcal{A}_i:=\{\emptyset,A,A^c,\Omega\}$. Hence, by Theorem A.37, the generated σ-algebras $\sigma(\mathcal{A}_i)$ are also independent. By the Definition A.35 of independence of σ-algebras, it follows that the (A_i^c) are independent.

The next result computes the probability of a sequence of events $A_1,A_2,\dots$ happening infinitely often. To this end, recall that $\{A_n \text{ i.o.}\}=\limsup_{n\to\infty}A_n=\bigcap_{n\in\mathbb{N}}\bigcup_{k\ge n}A_k$.

Lemma A.40 (Borel-Cantelli)**.** *Let $(\Omega,\mathcal{F},P)$ be a probability space and $(A_n)_{n\in\mathbb{N}}$ a sequence of $\mathcal{F}$-measurable events.*

(a) *If $\sum_{n=1}^{\infty}P[A_n]<\infty$, then $P[\{A_n \text{ i.o.}\}]=0$.*

(b) *If the A_n are independent and $\sum_{n=1}^{\infty}P[A_n]=\infty$, then $P[\{A_n \text{ i.o.}\}]=1$.*

Proof. (a) Proposition A.8(e) and (f) together with $\sum_{n=1}^{\infty}P[A_n]<\infty$ yields

$$P[\{A_n \text{ i.o.}\}]=P\left[\bigcap_{n\in\mathbb{N}}\bigcup_{k\ge n}A_k\right]=\lim_{n\to\infty}P\left[\bigcup_{k\ge n}A_k\right]\le\lim_{n\to\infty}\sum_{k=n}^{\infty}P[A_k]=0.$$

(b) It suffices to show that $P[\{A_n \text{ i.o.}\}^c] = 0$. The de Morgan laws and Proposition A.8(f) yield

$$P[\{A_n \text{ i.o.}\}^c] = P\left[\left(\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} A_k\right)^c\right] = P\left[\bigcup_{n\in\mathbb{N}}\bigcap_{k\geq n} A_k^c\right] = \lim_{n\to\infty} P\left[\bigcap_{k\geq n} A_k^c\right]$$

It suffices to show that $P[\bigcap_{k\geq n} A_k^c] = 0$ for each $n \in \mathbb{N}$. So fix $n \in \mathbb{N}$. Set $B_1 := A_n^c$, $B_2 := A_n^c \cap A_{n+1}^c, \ldots$. Then $B_1 \supset B_2 \supset \cdots \in \mathcal{F}$ and $\bigcap_{\ell\in\mathbb{N}} B_\ell = \bigcap_{k\geq n} A_k^c$. Proposition A.8(e), the fact that $(A_n^c)_{n\in\mathbb{N}}$ is independent by Example A.39, and the elementary inequality $1 - x \leq \exp(-x)$ for $x \in \mathbb{R}$ give

$$\begin{aligned} P\left[\bigcap_{k\geq n} A_k^c\right] &= P\left[\bigcap_{\ell\in\mathbb{N}} B_\ell\right] = \lim_{\ell\to\infty} P[B_\ell] = \lim_{\ell\to\infty} P\left[A_n^c \cap \cdots A_{n+\ell-1}^c\right] \\ &= \lim_{\ell\to\infty} \prod_{k=n}^{n+\ell-1} (1 - P[A_k]) \leq \liminf_{\ell\to\infty} \exp\left(-\sum_{k=n}^{n+\ell-1} P[A_k]\right) = 0. \end{aligned}$$ □

A.7 Expectation

In this section, we aim to define for a random variable X on a probability space $(\Omega, \mathcal{F}, P)$, the *expectation* of X under P (also called the *integral* of X with respect to P). We proceed in three steps.

First, we define the expectation for *simple* random variable.

Definition A.41. A random variable X on a probability space $(\Omega, \mathcal{F}, P)$ is called *simple* if there is $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R} \setminus \{0\}$ and $A_1, \ldots, A_n \in \mathcal{F}$ such that

$$X = c_1 \mathbf{1}_{A_1} + \cdots + c_n \mathbf{1}_{A_n}.$$

In this case, the *expectation of X with respect to P* is given by

$$E^P[X] := c_1 P[A_1] + \cdots + c_n P[A_n]. \tag{A.9}$$

If there is no danger of confusion, we often drop the qualifier P in E^P.

Remark A.42. (a) It is not difficult to check that each simple random variable has a representation such that the A_i are pairwise disjoint and the c_i are distinct.

(b) It is not difficult to check that (A.9) is independent of the choice of the representation of X. More precisely, if X can also be written as $X = d_1 \mathbf{1}_{B_1} + \cdots + d_m \mathbf{1}_{B_m}$ then

$$c_1 P[A_1] + \cdots + c_n P[A_n] = d_1 P[B_1] + \cdots + d_m P[B_m].$$

The following result shows that the expectation is linear on simple random variables. Its proof is left as an exercise.

Lemma A.43. *Let X and Y be simple random variables on some probability space $(\Omega, \mathcal{F}, P)$. For $a, b \in \mathbb{R}$, $aX + bY$ is again a simple random variable and*

$$E[aX + bY] = aE[X] + bE[Y].$$

We next aim to define the expectation for arbitrary *nonnegative* random variables.

Definition A.44. Let X be a $[0, \infty]$-valued random variable on some probability space $(\Omega, \mathcal{F}, P)$. Then the *expectation of X with respect to P* is given by

$$E^P[X] := \sup\left\{E^P[Y] : Y \text{ is nonnegative, simple, and satisfies } Y \leq X\right\}.$$

If there is no danger of confusion, we often drop the qualifier P in E^P.

Remark A.45. (a) The expectation of a nonnegative random variable can be ∞ even if X itself never takes the value ∞.

(b) It follows immediately from Definition (A.44) that the expectation is *monotone* in the sense that if X and Y are $[0, \infty]$-valued random variables with $X \leq Y$ then $E[X] \leq E[Y]$.

Definition A.44 suggests that in order to calculate the expectation of a nonnegative random variable X, we approximate X from below by a nondecreasing sequence of nonnegative simple random variables $(X_n)_{n\in\mathbb{N}}$ and calculate $\lim_{n\to\infty} E[X_n]$. The following two results show that this idea really works.

Lemma A.46. *Let X be a $[0, \infty]$-valued random variable on some probability space $(\Omega, \mathcal{F}, P)$. Then there exists a nondecreasing sequence of nonnegative simple random variables $(X_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} X_n = X$.*

Proof. For $n \in \mathbb{N}$, set $X_n(\omega) := \min(2^{-n}\lfloor 2^n X(\omega)\rfloor, n)$, where $\lfloor\cdot\rfloor$ denotes the floor function.[88] Then for each $\omega \in \Omega$, it is easy to check that the sequence $(X_n(\omega))_{n\in\mathbb{N}}$ is nondecreasing and satisfies $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. □

The next theorem shows that expectation and monotone limits can be interchanged. It is one of the cornerstones of modern integration theory; for a proof see [9, Lemma 4.2].

Theorem A.47 (Monotone convergence theorem)**.** *Let $X_1, X_2, \ldots$ and X be $[0, \infty]$-valued random variables on some probability space $(\Omega, \mathcal{F}, P)$. Suppose that $X_1 \leq X_2 \leq \cdots$ P-a.s. and $\lim_{n\to\infty} X_n = X$ P-a.s. Then*

$$\lim_{n\to\infty} E[X_n] = E[X]. \tag{A.10}$$

Remark A.48. Note that in (A.10) both sides may be ∞.

[88] If $X(\omega) = +\infty$, we use the conventions that $c \times \infty = \infty$ for $c > 0$ and $\lfloor\infty\rfloor = \infty$.

Theorem A.47 serves as a crucial ingredient to many proofs, where one first establishes the result for simple random variables and then passes to a monotone limit. To illustrate this approach, we prove the following generalisation of Lemma A.43.

Lemma A.49. *Let X and Y be $[0,\infty]$-valued random variables on some probability space $(\Omega, \mathcal{F}, P)$. For $a, b \geq 0$, $aX + bY$ is again a $[0,\infty]$-valued random variable and*[89]

$$E\left[aX + bY\right] = aE\left[X\right] + bE\left[Y\right].$$

Proof. First $aX + bY$ is a $[0,\infty]$-valued random variable by (a straightforward extension of) Theorem A.16. Next, let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be nondecreasing sequences of nonnegative simple random variables satisfying $\lim_{n\to\infty} X_n = X$ and $\lim_{n\to\infty} Y_n = Y$. (They exist by Lemma A.46.) Then $(aX_n + bY_n)_{n\in\mathbb{N}}$ is a nondecreasing sequences of nonnegative simple random variables with $\lim_{n\to\infty}(aX_n + bY_n) = aX + bY$. Hence, the monotone convergence theorem (Theorem A.47) and Lemma A.43(a) give

$$\begin{aligned} E\left[aX + bY\right] &= \lim_{n\to\infty} E\left[aX_n + bY_n\right] = \lim_{n\to\infty}\left(aE\left[X_n\right] + bE\left[Y_n\right]\right) \\ &= a\lim_{n\to\infty} E\left[X_n\right] + b\lim_{n\to\infty} E\left[Y_n\right] = aE\left[X\right] + bE\left[Y\right]. \end{aligned}$$ □

The following result can also be proved using the monotone convergence theorem. Its proof is left as an exercise.

Lemma A.50. *Let X be a $[0,\infty]$-valued random variable on some probability space $(\Omega, \mathcal{F}, P)$.*

(a) *We have $X = 0$ P-a.s. if and only if $E\left[X\right] = 0$.*

(b) *If $E\left[X\right] < \infty$, then $X < \infty$ P-a.s.*

As another application of the monotone convergence theorem, we prove the important *Lemma of Fatou.*

Lemma A.51 (Fatou)**.** *Let $X_1, X_2, \dots$ be $[0,\infty]$-valued random variables on some probability space $(\Omega, \mathcal{F}, P)$. Then*

$$E\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} E\left[X_n\right].$$

Proof. For $n \in \mathbb{N}$, set $Y_n := \inf_{m\geq n} X_m$. Then $Y_n \leq X_n$ for each $n \in \mathbb{N}$, and so

$$E\left[Y_n\right] \leq E\left[X_n\right], \quad n \in \mathbb{N}. \tag{A.11}$$

by Remark A.45(b). Moreover, $(Y_n)_{n\in\mathbb{N}}$ is a nondecreasing sequence of nonnegative random variables with $\lim_{n\to\infty} Y_n = \liminf_{n\to\infty} X_n$. Hence, by the monotone convergence theorem

[89] Here, we use the standard convention in measure theory that $c \times \infty =: \infty$ for $c > 0$ and $0 \times \infty =: 0$.

(Theorem A.47) and (A.11), we obtain

$$E\left[\liminf_{n\to\infty} X_n\right] = E\left[\lim_{n\to\infty} Y_n\right] = \lim_{n\to\infty} E\left[Y_n\right] = \liminf_{n\to\infty} E\left[Y_n\right] \leq \liminf_{n\to\infty} E\left[X_n\right]. \qquad \square$$

Finally, we define the expectation for a general random variable. To this end, recall that $\bar{\mathbb{R}} = [-\infty, +\infty]$.

Definition A.52. Let X be an $\bar{\mathbb{R}}$-valued random variable on some probability space $(\Omega, \mathcal{F}, P)$. X is called *integrable* or said to have *finite expectation* (with respect to P) if $E^P\left[|X|\right] < \infty$. In this case one sets

$$E^P\left[X\right] := E^P\left[X^+\right] - E^P\left[X^-\right],$$

where $X^+ = \max\{0, X\}$ denotes the *positive part* of X and $X^- = \max\{0, -X\}$ denotes the *negative part* of X.[90] If there is no danger of confusion, we often drop the qualifier P in E^P.

Remark A.53. (a) In situations where one wants to highlight the underlying sample space Ω, it is more handy to write the expectation in integral notation. For an integrable random variable X on a probability space $(\Omega, \mathcal{F}, P)$, we set

$$\int_\Omega X(\omega) P(\mathrm{d}\omega) := E^P\left[X\right],$$

and call this the *integral of* X with respect to P.

(b) The construction of the integral can be easily extended to *general measures* μ on $(\Omega, \mathcal{F})$, which are still σ-additive but do not longer satisfy $\mu(\Omega) = 1$; see [9, Chapter 4] for details.

- If $X = c_1 \mathbf{1}_{A_1} + \cdots c_n \mathbf{1}_{A_n}$ simple, we set[91]

$$\int_\Omega X(\omega)\mu(\mathrm{d}\omega) := c_1\mu(A_1) + \cdots + c_n\mu(A_n).$$

- If X is $[0, \infty]$-valued, we set

$$\int_\Omega X(\omega)\mu(\mathrm{d}\omega) := \sup\left\{\int_\Omega Y(\omega)\mu(\mathrm{d}\omega) : Y \geq 0 \text{ simple and satisfies } Y \leq X\right\}.$$

- If X is general, we say that X is *μ-integrable* if

$$\int_\Omega |X(\omega)|\mu(\mathrm{d}\omega) < \infty.$$

[90]Note that both X^+ and X^- are nonnegative random variables by Theorem A.15, Corollary A.14 and the fact that the functions $x \mapsto x^+$ and $x \mapsto x^-$ are continuous. Moreover, $X = X^+ - X^-$, and $|X| = X^+ + X^-$.

[91]Here, we use the standard convention in measure theory that $\infty + \infty =: \infty$.

In this case, we set

$$\int_\Omega X(\omega)\mu(\mathrm{d}\omega) := \int_\Omega X^+(\omega)\mu(\mathrm{d}\omega) - \int_\Omega X^-(\omega)\mu(\mathrm{d}\omega)$$

and call this the *integral* of X with respect to μ.

(c) The most important example of a general measure is the *Lebesgue-measure* λ on $\mathbb{R}$, which is the unique measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\lambda((a,b]) = b - a$ for all $-\infty < a < b < \infty$. If $f : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable function, we say that f is *Lebesgue-integrable* if

$$\int_{-\infty}^{\infty} |f(x)| \, \mathrm{d}x := \int_\mathbb{R} |f(x)| \lambda(\mathrm{d}x) < \infty$$

and write

$$\int_{-\infty}^{\infty} f(x) \, \mathrm{d}x := \int_\mathbb{R} f(x) \lambda(\mathrm{d}x)$$

for the *Lebesgue integral* of f.[92]

The following result lists important properties of the expectation operator.

Lemma A.54. *Let X and Y be integrable random variables on a probability space $(\Omega, \mathcal{F}, P)$.*

(a) *For $a, b \in \mathbb{R}$, $aX + bY$ is again an integrable random variable and*

$$E[aX + bY] = aE[X] + bE[Y].$$

(b) *If $X \geq Y$ P-a.s., then*

$$E[X] \geq E[Y]. \tag{A.12}$$

Moreover, the inequality in (A.12) *is an equality if and only if $X = Y$ P-a.s.*

Property (a) is referred to as *linearity* of the expectation and property (b) as *monotonicity* of the expectation.

Proof. (a) By the triangle inequality, Lemma A.49, and the fact that X and Y are integrable it follows that

$$E[|aX + bY|] \leq E[|a||X| + |b||Y|] = |a|E[|X|] + |b|E[|Y|] < \infty.$$

Thus, $aX + bY$ is integrable. The rest of the claim follows from splitting X, Y, and $X + Y$ into their positive and negative parts and applying Lemma A.49; for details see [9, Theorem 4.9(c)].

[92] Note that for Riemann-integrable functions, the Lebesgue integral and the Riemann integral coincide; see [9, Chapter 4.3].

(b) Set $Z := X - Y$. Then $Z \geq 0$ P-a.s. By part (a), it suffices to show that $E[Z] \geq 0$, where the inequality is an equality if and only if $Z = 0$ P-a.s. Since $Z \geq 0$ P-a.s., it follows that $Z^- = 0$ P-a.s., and so

$$E[Z] = E[Z^+] - E[Z^-] = E[Z^+].$$

by Lemma A.50(a). Since Z^+ is nonnegative, $E[Z^+] \geq 0$ by Definition A.44. This gives $E[Z] \geq 0$. Moreover, $E[Z] = E[Z^+] = 0$ if and only if $Z^+ = 0$ P-a.s. by Lemma A.50(a). Finally $Z^+ = 0$ P-a.s. if and only if $Z = 0$ P-a.s. since $Z^- = 0$ P-a.s. □

The next result is a measure theoretic change of variable formula. Its proof (which uses again the monotone convergence theorem) is left as an exercise.

Proposition A.55. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(\Omega', \mathcal{F}')$ a measurable space, and $X : \Omega \to \Omega'$ an $\mathcal{F} - \mathcal{F}'$ measurable map. Moreover, let $g : \Omega' \to \bar{\mathbb{R}}$ be an $\mathcal{F}' - \mathcal{B}_{\bar{\mathbb{R}}}$-measurable map. Let P_X be image measure of X under P.*[93] *Then $g(X)$ is P integrable if and only if g is P_X integrable. Moreover, in this case (or if $g \geq 0$) we have the* transformation formula*:*

$$\int_\Omega g(X(\omega)) P(\mathrm{d}\omega) = \int_{\Omega'} g(\omega') P_X(\mathrm{d}\omega').$$

The following lemma considers the special case that X is a random variable with discrete or continuous distribution. Its proof (which uses Remarks A.24 and A.27 together with the monotone convergence theorem) is left as an exercise.

Lemma A.56. *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $g : \mathbb{R} \to \mathbb{R}$ a measurable function.*

(a) *If X is discrete with pmf p_X, then*

$$E[|g(X)|] = \sum_x |g(x)| p_X(x). \tag{A.13}$$

If (A.13) *is finite, then*

$$E[g(X)] = \sum_x g(x) p_X(x).$$

(b) *If X is continuous with pdf f_X, then*

$$E[|g(X)|] = \int_{-\infty}^{\infty} |g(x)| f_X(x)\,\mathrm{d}x. \tag{A.14}$$

[93] This is the probability measure on $(\Omega', \mathcal{F}')$ defined by

$$P_X[A'] := P[X \in A'].$$

If (A.14) *is finite, then*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, \mathrm{d}x.$$

Finally, we link the notion of independence to the concept of expectation; see [9, Theorem 5.4] for a proof (which uses once again the monotone convergence theorem).

Theorem A.57. *Let X and Y be independent integrable random variables on some probability space $(\Omega, \mathcal{F}, P)$. Then XY is also integrable and*

$$E[XY] = E[X]E[Y].$$

A.8 L^p-spaces

In this section, we introduce the key notion of L^p-spaces.

First, we consider the case $p \in [1, \infty)$.

Definition A.58. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $p \in [1, \infty)$. For an $\mathcal{F}$-measurable random variable X, set[94]

$$\|X\|_p := E\left[|X|^p\right]^{1/p}. \tag{A.15}$$

If $\|X\|_p < \infty$, we say that X has *finite p-th moment* (with respect to P) and call $E\left[X^p\right] < \infty$ the *p-th moment* of X. We denote the collection of all random variables on $(\Omega, \mathcal{F})$ with finite p-th moment with respect to P by $\mathcal{L}^p(\Omega, \mathcal{F}, P)$. If there is no danger of confusion, we often write $\mathcal{L}^p(P)$ or just $\mathcal{L}^p$ for $\mathcal{L}^p(\Omega, \mathcal{F}, P)$.

Remark A.59. The reason for the "outside power" $\frac{1}{p}$ in (A.15) is to ensure that the map $\|\cdot\|_p : \mathcal{L}^p \to \mathbb{R}$ is *positively homogenous*. Indeed, let $\lambda \geq 0$ and $X \in \mathcal{L}^p$. Then linearity of the expectation gives

$$\|\lambda X\|_p = E\left[|\lambda X|^p\right]^{1/p} = \left(\lambda^p E\left[|X|^p\right]\right)^{1/p} = \lambda E\left[|X|^p\right]^{1/p} = \lambda \|X\|_p.$$

Next, we turn to the case $p = \infty$.

Definition A.60. Let $(\Omega, \mathcal{F}, P)$ be a probability space. For an $\mathcal{F}$-measurable random variable X, set

$$\|X\|_\infty := \inf\left\{K \geq 0 : |X| \leq K \;\; P\text{-a.s.}\right\}.$$

If $\|X\|_\infty < \infty$, we say that X is P-a.s.-bounded. We denote the collection of all real-valued random variables on $(\Omega, \mathcal{F})$ that are P-a.s.-bounded by $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$. If there is no danger of confusion, we often write $\mathcal{L}^\infty(P)$ or just $\mathcal{L}^\infty$ for $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$.

[94]Here, we use the natural convention that $\infty^{\frac{1}{p}} = \infty$.

The following result shows that L^p-spaces are naturally ordered.

Proposition A.61. *Let $1 \leq p_1 < p_2 \leq \infty$. Then $\mathcal{L}^{p_2}(\Omega, \mathcal{F}, P) \subset \mathcal{L}^{p_1}(\Omega, \mathcal{F}, P)$.*

Proof. Let $X \in \mathcal{L}^{p_2}$.

First assume that $p_2 = \infty$. Then there is a constant $K > 0$ such that $X \leq K$ P-a.s. Monotonicity of the expectation gives $E[|X|^{p_1}] \leq K^{p_1} < \infty$ and so $X \in \mathcal{L}^{p_1}$.

Next assume that $p_2 < \infty$. The elementary inequality $x^{p_1} \leq 1 + x^{p_2}$ for $x \geq 0$ together with monotonicity and linearity of the expectation give

$$E[|X|^{p_1}] \leq E[1 + |X|^{p_2}] \leq 1 + E[|X|^{p_2}] < \infty.$$

Thus, $X \in \mathcal{L}^{p_1}$. □

We proceed to state the important inequalities of *Hölder* and *Minkowski*; for a proof see [9, Theorems 7.16 and 7.17]. Hölder's inequality shows that the product of random variables that lie in L^p-spaces with *conjugate exponents* is integrable. Here, $p, q \in [1, \infty]$ are called *conjugate* if $\frac{1}{p} + \frac{1}{q} = 1$, with the convention that $1/\infty := 0$.

Theorem A.62 (Hölder's inequality)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X, Y be random variables with $X \in \mathcal{L}^p(P)$ and $Y \in \mathcal{L}^q(P)$, where $p, q \in [1, \infty]$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then $XY \in L^1(P)$ and*

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q$$

Minkowski's inequality shows that the map $\|\cdot\|_p : \mathcal{L}^p \to \mathbb{R}_+$ satisfies the triangle inequality.

Theorem A.63 (Minkowski's inequality)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $p \in [1, \infty]$ and $X, Y \in \mathcal{L}^p(P)$. Then*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

One important consequence of Minkowki's inequality is that the map $\|\cdot\|_p$ is a *norm* and each $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ is a normed vectors space.[95] One can even show that the metric/topology induced by $\|\cdot\|_p$ is complete and hence a *Banach space.* This is the content of the following result; for a proof see [9, Theorem 7.18]

Theorem A.64 (Fischer-Riesz)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $p \in [1, \infty]$, and $(X_n)_{n \in \mathbb{N}}$ a Cauchy sequence in $\mathcal{L}^p$, i.e., for each $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that for all $m, n \geq N$,*

$$\|X_n - X_m\|_p < \varepsilon.$$

There there is $X \in \mathcal{L}^p$ with

$$\lim_{n \to \infty} \|X_n - X\|_p = 0.$$

[95] To be precise this also requires to identify random variables that coincide P-a.s, i.e., one has to pass from random variables to equivalence classes of random variables that coincide P-a.s.. Some authors indicate this by writing $L^p(\Omega, \mathcal{F}, P)$ instead of $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ when doing this. However, we will always work with random variables.

A.9 Variance and covariance

In this section, we study the *variance* and *covariance* of random variables.

Definition A.65. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X, Y \in \mathcal{L}^2(P)$.

(a) The *variance of* X is defined by

$$\operatorname{Var}[X] := E\left[(X - E[X])^2\right].$$

(b) The *covariance of* X *and* Y is defined by

$$\operatorname{Cov}[X, Y] := E\left[(X - E[X])(Y - E[Y])\right].$$

(c) X and Y are called *uncorrelated* if $\operatorname{Cov}[X, Y] = 0$ and *correlated* if $\operatorname{Cov}[X, Y] \neq 0$.

The following result lists some elementary properties of the variance/covariance; its proof is left as an exercise.

Proposition A.66. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X, Y \in \mathcal{L}^2$ and $a, b, c, d \in \mathbb{R}$.*

(a) $\operatorname{Cov}[X, Y] = E[XY] - E[X]E[Y]$.

(b) $\operatorname{Cov}[aX + b, cX + d] = ac \operatorname{Cov}[X, Y]$.

(c) $\operatorname{Var}[X] = 0$ *if and only if X is P-a.s.-constant.*

We proceed to show that independent random variables are uncorrelated.

Proposition A.67. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X, Y \in \mathcal{L}^2$. If X and Y are independent, then they are uncorrelated.*

Proof. This follows immediately from Proposition A.66(a) and Theorem A.57. □

Remark A.68. The converse of Proposition A.67 is false.

We proceed to calculate the variance of the sum of random variables.

Proposition A.69. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X_1, \ldots, X_n \in \mathcal{L}^2$. Then*

$$\operatorname{Var}\left[\sum_{k=1}^{n} X_n\right] = \sum_{k=1}^{n} \operatorname{Var}[X_k] + \sum_{1 \leq i < j \leq n} 2 \operatorname{Cov}[X_i, X_j]. \tag{A.16}$$

In particular, if $X_1, \ldots, X_n$ are uncorrelated, the Bienaymé formula *holds:*

$$\operatorname{Var}\left[\sum_{k=1}^{n} X_k\right] = \sum_{k=1}^{n} \operatorname{Var}[X_k].$$

Proof. Linearity of the expectation and the fact that $\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X]$ yield

$$\begin{aligned}
\mathrm{Var}\left[\sum_{k=1}^{n} X_n\right] &= E\left[\left(\sum_{k=1}^{n} X_k - E[X_k]\right)\left(\sum_{k=1}^{n} X_k - E[X_k]\right)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} E\left[(X_i - E[X_i])(X_j - E[X_j])\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}[X_i, X_j] = \sum_{k=1}^{n} \mathrm{Var}[X_k] + \sum_{i,j=1, i\neq j}^{n} \mathrm{Cov}[X_i, X_j] \\
&= \sum_{k=1}^{n} \mathrm{Var}[X_k] + \sum_{1\leq i<j\leq n} 2\,\mathrm{Cov}[X_i, X_j].
\end{aligned}$$

□

A.10 The inequalities of Markov and Jensen

In this section, we study some key inequalities of probability theory.

First, we establish the elementary but important *Markov inequality.*

Theorem A.70 (Markov's inequality)**.** *Let X be a random variable on some probability space $(\Omega, \mathcal{F}, P)$ and $f : [0,\infty) \to [0,\infty)$ a nondecreasing function.*[96] *Then for any $\varepsilon > 0$ with $f(\varepsilon) > 0$, we have the* Markov inequality*:*

$$P[|X| \geq \varepsilon] \leq \frac{E\left[f(|X|)\right]}{f(\varepsilon)}.$$

Proof. Monotonicity and linearity of the expectation together with the fact that f is nondecreasing give

$$\begin{aligned}
E\left[f(|X|)\right] &\geq E\left[f(|X|)\mathbf{1}_{\{f(|X|)\geq f(\varepsilon))\}}\right] \geq E\left[f(\varepsilon)\mathbf{1}_{\{f(|X|)\geq f(\varepsilon))\}}\right] \\
&\geq E\left[f(\varepsilon)\mathbf{1}_{\{|X|\geq\varepsilon\}}\right] = f(\varepsilon)P[|X| \geq \varepsilon].
\end{aligned}$$

Now the claim follows by rearrangement. □

The case $f(x) = x^2$ is of special importance.

Corollary A.71 (Chebyshev's inequality)**.** *Let X be a real-valued random variable on some probability space $(\Omega, \mathcal{F}, P)$ and assume that $X \in \mathcal{L}^2(P)$. Then for $\varepsilon > 0$, we have the* Chebyshev inequality*:*

$$P[|X - E\left[X\right]| \geq \varepsilon] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2}.$$

For the next inequality, we need to recall the notion of a *convex function.*

[96] Note that f is then automatically measurable. Indeed, let $x \in \mathbb{R}$, then $\{f^{-1}([0,x])\} \in \mathcal{B}_{[0,\infty]}$ because it is of the form $[0,a)$ or $[0,a]$ for some $a \in [0,\infty]$.

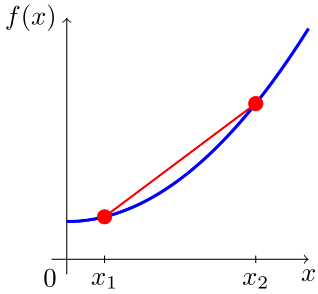

Figure 1: Example of a convex function

Definition A.72. Let $D \subset \mathbb{R}$ be a non-empty interval. A function $f : D \to \mathbb{R}$ is called *convex* if

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2), \quad x_1, x_2 \in D, \lambda \in [0,1]. \tag{A.17}$$

It is called *strictly convex* if the inequality in (A.17) is strict for $x_1 \neq x_2$ and $\lambda \in (0,1)$.

From a geometric perspective, (strict) convexity means that straight line segments joining $(x_1, f(x_1))$ to $(x_2, f(x_2))$ always lie (strictly) above the graph of f.

Remark A.73. (a) If f is convex, then f is automatically continuous in the interior of D; see [9, Theorem 7.7(i)].

(b) If f is (strictly) convex, then $-f$ is called (strictly) *concave*.

(c) If $f : D \to \mathbb{R}$ is twice continuously differentiable then f is convex if and only if $f'' \geq 0$ in the interior of D. Moreover, it is strictly convex if $f'' > 0$ in the interior of D.[97]

We proceed to state and prove the fundamental inequality for convex functions.

Theorem A.74 (Jensen's inequality)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and X an integrable random variable with values in a non-empty interval $D \subset \mathbb{R}$. Let $f : D \to \mathbb{R}$ be convex and suppose that f is nonnegative or $E[|f(X)|] < \infty$. Then*

$$E[f(X)] \geq f(E[X]).$$

Moreover, the inequality is strict when f is strictly convex and X is not P-a.s. constant.

Proof. The claim is trivial if f is nonnegative and $E[f(X)] = \infty$. So it suffices to consider the case $E[|f(X)|] < \infty$.

First, using the definition of convexity, one can show that for each $a \in D$, there is $b \in \mathbb{R}$ such that

$$f(x) \geq f(a) + b(x-a), \tag{A.18}$$

[97]The converse is not true: For example, the function $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^4$ is strictly convex, but $f''(0) = 0$.

where the inequality in (A.18) is strict for $x \neq a$ if f is strictly convex.[98]

Next, choose $a := E[X]$. One can show that $a \in D$ because D is an interval. Let $b \in \mathbb{R}$ be such that (A.18) is satisfied. Then

$$f(X) \geq f(a) + b(X - a)$$

and

$$P[f(X) > f(a) + b(X - a)] = P[X \neq a] > 0,$$

if f is strictly convex and X is not P-a.s. constant. Thus, by monotonicity and linearity of the integral and the fact that $a = E[X]$,

$$\begin{aligned} E[f(X)] &\geq E[f(a) + b(X - a)] = f(a) + b(E[X] - a) = f(a) + b(a - a) = f(a) \\ &= f(E[X]), \end{aligned}$$

where the inequality is strict if f is strictly convex and X is not P-a.s. constant. □

A.11 Product spaces and Fubini's theorem

In this section, we study products of probability spaces and formulate the important Theorem of *Fubini*.

First, we consider the product of sample spaces.

Definition A.75. Let $\Omega_1, \dots, \Omega_N$ be sample spaces. Set

$$\Omega := \{\omega = (\omega_1, \dots, \omega_N) : \omega_1 \in \Omega_1, \dots, \omega_N \in \Omega_N\}.$$

Then Ω is called the *product sample space* of $\Omega_1, \dots, \Omega_N$ and denoted by

$$\Omega := \bigtimes_{n=1}^{N} \Omega_n := \Omega_1 \times \cdots \times \Omega_N.$$

If $\Omega_i = \Omega_0$ for all $i \in \{1, \dots, N\}$, we also write $\Omega := \Omega_0^N$.

Example A.76. Consider rolling a die 3 times. Then this can be modelled by the sample space $\Omega := \{1, \dots, 6\}^3 := \{(\omega_1, \omega_2, \omega_3) : \omega_1, \omega_2, \omega_3 \in \{1, \dots, 6\}\}$.

Next, we consider the product of σ-algebras.

[98] If f is twice continuously differentiable, the (weak) inequality (A.18) can be easily derived as follows: Fix $a \in D$ and set $b := f'(a)$. By a Taylor expansion of f in a of order 1 with Lagrange remainder term, we obtain for fixed $x \in D$

$$f(x) = f(a) + b(x - a) + \frac{1}{2} f''(\xi)(x - a)^2,$$

where ξ lies in the interval with the endpoints x and a. Since $f'' \geq 0$ by convexity of f, (A.18) follows.

Definition A.77. Let $(\Omega_1, \mathcal{F}_1), \dots, (\Omega_N, \mathcal{F}_N)$ be measurable spaces. Set

$$\mathcal{F} := \sigma\left(\{A_1 \times \cdots \times A_N : A_1 \in \mathcal{F}_1, \dots, A_N \in \mathcal{F}_N\ \}\right).$$

Then $\mathcal{F}$ is called the *product σ-algebra* of $\mathcal{F}_1, \dots, \mathcal{F}_N$ and denoted by

$$\mathcal{F} := \bigotimes_{n=1}^{N} \mathcal{F}_n := \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_N.$$

If $(\Omega_n, \mathcal{F}_n) = (\Omega_0, \mathcal{F}_0)$ for all $i \in \{1, \dots, N\}$, we also write $\mathcal{F} := \mathcal{F}_0^{\otimes N}$.

Intuitively, the product σ-algebra is the smallest σ-algebra such that all *rectangular sets* $A_1 \times \cdots \times A_N$ of Ω with $A_1 \in \mathcal{F}_1, \dots, A_N \in \mathcal{F}_N$ are measurable.

Finally, we consider the product of probability measures. The following result establishes existence and uniqueness of the *product measure*; for a proof see [9, Theorem 14.14].

Theorem A.78. *Let $(\Omega_1, \mathcal{F}_1, P_1), \dots, (\Omega_N, \mathcal{F}_N, P_N)$ be probability spaces. Then there exists a unique probability measure P on $(\times_{n=1}^{N} \Omega_n, \bigotimes_{n=1}^{N} \mathcal{F}_n)$ such that*

$$P[A_1 \times \cdots \times A_N] = \prod_{n=1}^{N} P[A_n] \tag{A.19}$$

It is called the product measure *of $P_1, \dots, P_N$ and denoted by*

$$P := \bigotimes_{n=1}^{N} P_n := P_1 \otimes \cdots \otimes P_N$$

If $(\Omega_i, \mathcal{A}_i, P_i) = (\Omega_0, \mathcal{F}_0, P_0)$ for all $i \in \{1, \dots, N\}$, we also write $P := P_0^{\otimes N}$.

We proceed to study the expectation of random variables with respect to the product measure of two probability measures. In this case the integral notation is more handy; cf. Remark A.53. The following result shows that instead of integrating over the product measure (which we do not know explicitly), we can also integrate first over one measure and then over the other. Moreover, the order of integration does not matter; for a proof see [9, Theorem 14.16].

Theorem A.79 (Fubini)**.** *Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be probability spaces. Let $X : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ be $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable. Assume that either $X \geq 0$ $P_1 \otimes P_2$-almost surely or $X \in L^1(P_1 \otimes P_2)$. Then*

- *The map $\omega_1 \mapsto \int_{\Omega_2} X(\omega_1, \omega_2) P_2(\mathrm{d}\omega_2)$ is $\mathcal{F}_1$-measurable and P_1-integrable in case that $X \in L^1(P_1 \otimes P_2)$.*[99]

[99] Here, we agree that $\int_{\Omega_2} X(\omega_1, \omega_2) P_2(\mathrm{d}\omega_2) := -\infty$ if $\int_{\Omega_2} |X(\omega_1, \omega_2)| P_2(\mathrm{d}\omega_2) = \infty$.

- *The map* $\omega_2 \mapsto \int_{\Omega_1} X(\omega_1, \omega_2) P_1(\mathrm{d}\omega_1)$ *is* $\mathcal{F}_2$*-measurable and* P_2*-integrable in case that* $X \in L^1(P_1 \otimes P_2)$.[100]

- *We have the identity*

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2)(P_1 \otimes P_2)(d(\omega_1, \omega_2)) &= \int_{\Omega_1} \left(\int_{\Omega_2} X(\omega_1, \omega_2) P_2(\mathrm{d}\omega_2) \right) P_1(\mathrm{d}\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} X(\omega_1, \omega_2) P_1(\mathrm{d}\omega_1) \right) P_2(\mathrm{d}\omega_2).\end{aligned}$$

Remark A.80. (a) Fubini's theorem also holds more generally, if we replace P_1 and P_2 by *σ-finite* measures μ_1 and μ_2. Of special importance is the case if μ_1 and μ_2 are the Lebesgue measure; see [9, Section 14.2] for details.

(b) The notion of product sample spaces, product σ-algebras, and product measures can be extended to countable (and even uncountable) families of probability spaces; see [9, Chapter 14] for details.

A.12 Convergence of random variables

In this section, we look at different types of convergence for random variables.

Definition A.81. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. Then $(X_n)_{n \in \mathbb{N}}$ is said to converge to a random variable X

- *in* L^p, where $p \in [1, \infty]$, if each $X_n \in \mathcal{L}^p(P)$ and

$$\lim_{n \to \infty} \|X_n - X\|_p = 0.$$

  In this case, we write $X_n \overset{L^p}{\to} X$.

- *almost surely*, if there is a P-nullset $N \in \mathcal{F}$ such that

$$\lim_{n \to \infty} X_n(\omega) = X(\omega), \quad \text{for all } \omega \in \Omega \setminus N.$$

  In this case, we write $X_n \overset{\text{a.s.}}{\to} X$.

- *in probability*, if for each $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left[|X_n - X| > \varepsilon\right] = 0.$$

  In this case, we write $X_n \overset{P}{\to} X$.

[100] Here, we agree that $\int_{\Omega_1} X(\omega_1, \omega_2) P_1(\mathrm{d}\omega_1) := -\infty$ if $\int_{\Omega_1} |X(\omega_1, \omega_2)| P_1(\mathrm{d}\omega_1) = \infty$.

- *in distribution*, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \text{for all } x \in \mathbb{R} \text{ such that } F_X \text{ is continuous at x.}$$

In this case, we write $X_n \Rightarrow X$.

Remark A.82. (a) If $(X_n)_{n\in\mathbb{N}}$ converges in L^p, almost surely, or in probability, the limiting random variable X is P-almost surely unique. In light of Proposition A.83 below, it suffices to show this for the case of convergence in probability. So suppose that $(X_n)_{n\in\mathbb{N}}$ converges in probability to X and X'. Let $\varepsilon > 0$ be given. Using that for each $x \in \mathbb{R}$

$$\{|X - X'| > \varepsilon\} \subset \left\{|X - x| > \frac{\varepsilon}{2}\right\} \cup \left\{|X' - x| > \frac{\varepsilon}{2}\right\}.$$

we obtain

$$P[|X - X'| > \varepsilon] \leq \limsup_{n\to\infty} \left(P\left[|X - X_n| > \frac{\varepsilon}{2}\right] + P\left[|X' - X_n| > \frac{\varepsilon}{2}\right]\right) = 0.$$

We may conclude that $X = X'$ P-a.s.

(b) Since $|X| \leq |X - X_n| + |X_n|$ for each n, $X_n \in L^p$ together with $X_n \xrightarrow{L^p} X$ and Minkowski's inequality give $X \in \mathcal{L}^p$. Moreover, using also that $|X_n| \leq |X - X_n| + |X|$ for each $n \in \mathbb{N}$, we get by Minkowski's inequality and properties of the limit inferior and the limit superior[101]

$$\begin{aligned}
\|X\|_p &\leq \liminf_{n\to\infty} \||X - X_n| + |X_n|\|_p \leq \liminf_{n\to\infty} (\|X - X_n\|_p + \|X_n\|_p) \\
&= \liminf_{n\to\infty} \|X_n\|_p \leq \limsup_{n\to\infty} \|X_n\|_p \leq \limsup_{n\to\infty} \||X - X_n| + |X|\|_p \\
&\leq \limsup_{n\to\infty} (\|X - X_n\|_p + \|X\|_p) = \|X\|_p.
\end{aligned}$$

This implies that

$$\lim_{n\to\infty} \|X_n\|_p = \|X\|_p < \infty.$$

(c) Using Markov's inequality (Theorem A.70), it is not difficult to check that $X_n \xrightarrow{P} X$ if

[101] More precisely, we use that if $(a_n)_{n\in\mathbb{N}}$ and (b_n) are sequence of real numbers, where $(b_n)_{n\in\mathbb{N}}$ is convergent, then $\liminf_{n\to\infty}(a_n + b_n) = \liminf_{n\to\infty} a_n + \lim_{n\to\infty} b_n$ and $\limsup_{n\to\infty}(a_n + b_n) = \limsup_{n\to\infty} a_n + \lim_{n\to\infty} b_n$. Here, we only show the claim for the limit inferior. Since $\liminf_{n\to\infty}(a_n + b_n) \geq \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$ without any assumptions on $(b_n)_{n\in\mathbb{N}}$, in suffices to show that $\liminf_{n\to\infty}(a_n + b_n) \leq \liminf_{n\to\infty} a_n + \lim_{n\to\infty} b_n$. Let $\varepsilon > 0$ be given and set $b := \lim_{n\to\infty} b_n$. Then there is $N \in \mathbb{N}$ such that $b_n \leq b + \varepsilon$ for all $n \geq N$. Hence, by the definition of the limit inferior,

$$\liminf_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} \inf_{k\geq n} (a_k + b_k) \leq \lim_{n\to\infty} \inf_{k\geq n} (a_k + b + \varepsilon) = \lim_{n\to\infty} \left(\inf_{k\geq n} a_k + (b + \varepsilon)\right) = \liminf_{n\to\infty} a_n + b + \varepsilon$$

Now the claim follows from letting $\varepsilon \to 0$.

and only if[102]

$$\lim_{n\to\infty} E\left[|X_n - X| \wedge 1)\right] = 0.$$

This alternative characterisation shows that the topology induced by convergence in probability is *metrisable* with metric $d(X,Y) = E\left[|X - Y| \wedge 1\right]$.

(d) One can show with some effort (see [9, Theorem 13.23]), that $X_n \Rightarrow X$ if and only if for all bounded continuous functions $f : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n\to\infty} E\left[f(X_n)\right] = E\left[f(X)\right].$$

This alternative characterisation explains why convergence in distribution is also called *weak convergence.*[103]

We proceed to study the relationship between the different types of convergence. First, it is not difficult to check that L^p convergence does not imply almost sure convergence and vice versa.

Next, we show that convergence in L^p and almost sure convergence both imply convergence in probability. By contrast, it is not difficult to check that convergence in probability does neither imply L^p-convergence nor almost sure convergence.

Proposition A.83. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(X_n)_{n\in\mathbb{N}}$ a sequence of random variables and X a random variable.*

(a) *If $(X_n)_{n\in\mathbb{N}}$ converges to X in L^p for $p \in [1,\infty]$, then it converges to X in probability.*

(b) *If $(X_n)_{n\in\mathbb{N}}$ converges to X almost surely, then it converges to X in probability.*

Proof. (a) The case $p = \infty$ is easy and left as an exercise. Assume that $p < \infty$. Let $\varepsilon > 0$. Markov's inequality (Theorem A.70) with $f(x) = x^p$ and the fact that $X_n \overset{L^p}{\to} X$ give

$$\limsup_{n\to\infty} P[|X_n - X| \geq \varepsilon] \leq \limsup_{n\to\infty} \frac{E[|X - X_n|^p]}{\varepsilon^p} = \limsup_{n\to\infty} \frac{1}{\varepsilon^p}\|X_n - X\|_p^p = 0.$$

(b) Let $\varepsilon > 0$. For $n \in \mathbb{N}$, set $A_n := \{|X_n - X| > \varepsilon\}$. Since $X_n \overset{\text{a.s.}}{\to} X$, it follows that $P[\{A_n \text{ i.o.}\}] = 0$. Hence, by Exercise 1.2(c), we have

$$\limsup_{n\to\infty} P[|X_n - X| > \varepsilon] = \limsup_{n\to\infty} P[A_n] \leq P[\{A_n \text{ i.o.}\}] = 0. \qquad \square$$

Finally, we show that convergence in probability implies convergence in distribution. By contrast, it is not difficult to check that convergence in distribution does not imply convergence in probability.

[102] Recall that $x \wedge y := \min(x,y)$ and $x \vee y := \max(x,y)$ for $x, y \in \mathbb{R}$.

[103] Note, however, that from a functional analysis perspective, this "weak convergence" is in fact weak*-convergence.

Proposition A.84. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(X_n)_{n\in\mathbb{N}}$ a sequence of random variables and X a random variable. If $(X_n)_{n\in\mathbb{N}}$ converges to X in probability, then it converges to X in distribution.*

Proof. Let $x \in \mathbb{R}$ be a continuity point of F_X. We have to show that

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Using that $A = (A\cap B)\cup(A\cap B^c) \subset B\cup(A\cap B^c)$ for all $A, B \in \mathcal{F}$, we obtain for fixed $\varepsilon > 0$ and $n \in \mathbb{N}$,

$$\begin{aligned}\{X \le x-\varepsilon\} &\subset \{X_n \le x\} \cup \{X \le x-\varepsilon, X_n > x\} \subset \{X_n \le x\} \cup \{|X - X_n| \ge \varepsilon\},\\ \{X_n \le x\} &\subset \{X \le x+\varepsilon\} \cup \{X_n \le x, X > x+\varepsilon\} \subset \{X \le x+\varepsilon\} \cup \{|X - X_n| \ge \varepsilon\}.\end{aligned}$$

Taking probabilities, we get

$$\begin{aligned}F_X(x-\varepsilon) &\le F_{X_n}(x) + P[|X_n - X| \ge \varepsilon],\\ F_{X_n}(x) &\le F_X(x+\varepsilon) + P[|X_n - X| \ge \varepsilon].\end{aligned}$$

Letting $n \to \infty$ and using that $X_n \xrightarrow{P} X$, we obtain

$$F_X(x-\varepsilon) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le F_X(x+\varepsilon).$$

Now (A.12) follows by letting $\varepsilon \to 0$ and using that F_X is continuous at x. □

A.13 Uniform integrability and the dominated convergence theorem

In this section, we try to understand under which conditions almost sure convergence implies convergence in L^1, i.e., under which conditions limits and expectations can be interchanged.

The key ingredient is the notion of *uniform integrability* of a family of random variables.

Definition A.85. A family $(X_i)_{i\in\mathcal{I}}$ of random variables on some probability space $(\Omega, \mathcal{F}, P)$ is said to be *uniformly integrable* (UI) if

$$\lim_{K\to\infty} \sup_{i\in\mathcal{I}} E\left[|X_i| \mathbf{1}_{\{|X_i|\ge K\}}\right] = 0.$$

It is not difficult to check that a single random variable is uniformly integrable if and only if it is integrable. The following result lists some further simple criteria to check for uniform integrability. Its proof is left as an exercise.

Lemma A.86. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(X_i)_{i\in\mathcal{I}}$ a family of integrable random variables.*

(a) *If there is* $X \in \mathcal{L}^1$ *with* $|X_i| \leq X$ *P-a.s. for all* $i \in \mathcal{I}$*, then* $(X_i)_{i\in\mathcal{I}}$ *is uniformly integrable.*

(b) *If* $\mathcal{I}$ *is a finite index set, then* $(X_i)_{i\in\mathcal{I}}$ *is uniformly integrable.*

(c) *If* $(X_i)_{i\in\mathcal{I}}$ *is uniformly integrable and* $X \in \mathcal{L}^1$*, then* $(X_i + X)_{i\in\mathcal{I}}$ *is again uniformly integrable.*

The following result gives two equivalent useful characterisations of uniform integrability; for a proof see [9, Theorems 6.19 and 6.24]

Theorem A.87 (de la Vallée-Poussin)**.** *Let* $(X_i)_{i\in\mathcal{I}}$ *be a family of variables on some probability space* $(\Omega, \mathcal{F}, P)$*. Then the following are equivalent.*

(1) $(X_i)_{i\in\mathcal{I}}$ *is uniformly integrable.*

(2) $(X_i)_{i\in\mathcal{I}}$ *is* bounded in L^1, *i.e.,* $\sup_{i\in\mathcal{I}} E[|X_i|] < \infty$*, and for each* $\varepsilon > 0$*, there exists* $\delta > 0$ *such that*

$$P[A] \leq \delta \implies E[|X_i|\mathbf{1}_A] \leq \varepsilon \text{ for all } i \in \mathcal{I}.$$

(3) *There exists a nondecreasing convex function* $H : [0,\infty) \to [0,\infty)$ *with* $\lim\limits_{x\to\infty} \frac{H(x)}{x} = \infty$ *such that* $\sup_{i\in\mathcal{I}} E[H(|X_i|)] < \infty$.

We note an important corollary, which gives one of the most useful criterion in practice to check that a family of random variables is UI.

Corollary A.88. *Let* $(X_i)_{i\in\mathcal{I}}$ *be a family of random variables on some probability space* $(\Omega, \mathcal{F}, P)$ *and* $p \in (1, \infty]$*. Suppose that* $(X_i)_{i\in\mathcal{I}}$ *is* bounded in L^p, *i.e.,* $\sup_{i\in\mathcal{I}} \|X_i\|_p < \infty$*. Then* $(X_i)_{i\in\mathcal{I}}$ *is uniformly integrable.*

With the help of Theorem A.87, we can now show that a sequence of integrable random variable that converges almost surely converges in L^1 if and only if it is uniformly integrable

Theorem A.89. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $(X_n)_{n\in\mathbb{N}}$ *a sequence of random variables that converges almost surely to random variable* X*.*[104] *Then the following are equivalent.*

(1) $(X_n)_{n\in\mathbb{N}}$ *is uniformly integrable.*

(2) $(X_n)_{n\in\mathbb{N}}$ *converges to* X *in* L^1*.*

Proof. We shall only proof the more important direction "(1) ⇒ (2)"; the other direction is left as an exercise.

[104] With slightly more work, one can show that the result still holds if we only assume that $(X_n)_{n\in\mathbb{N}}$ converges to X in probability; see [9, Theorem 6.25] for details.

"(1) ⇒ (2)". First we show that X is integrable. Using that the X_n are bounded in L^1 by Theorem A.87(b), Fatou's lemma gives

$$E[|X|] = E\left[\liminf_{n\to\infty} |X_n|\right] \leq \liminf_{n\to\infty} E[|X_n|] \leq \sup_{n\in\mathbb{N}} E[|X_n|] < \infty.$$

Next, set $Y_n := X_n - X$ for $n \in \mathbb{N}$. Then Y_n converges to 0 almost surely, and $(Y_n)_{n\in\mathbb{N}}$ is UI by Lemma A.86. Let $\varepsilon > 0$ be given. Then for each $n \in \mathbb{N}$,

$$E[|Y_n|] = E\left[|Y_n|\mathbf{1}_{\{|Y_n|\leq\varepsilon\}}\right] + E\left[|Y_n|\mathbf{1}_{\{|Y_n|>\varepsilon\}}\right] \leq \varepsilon + E\left[|Y_n|\mathbf{1}_{\{|Y_n|>\varepsilon\}}\right]. \tag{A.20}$$

Moreover, by Theorem A.87(b), there is $\delta > 0$ such that

$$P[A] \leq \delta \implies E[|Y_n|\mathbf{1}_A] \leq \varepsilon \text{ for all } n \in \mathbb{N}. \tag{A.21}$$

Now using that Y_n converges to 0 almost surely, and hence in probability by Proposition A.83, there is $N \in \mathbb{N}$ such that $P[|Y_n| > \varepsilon] \leq \delta$ for all $n \geq N$. Combining this with (A.20) and (A.21), we obtain

$$E[|Y_n|] \leq 2\varepsilon \quad \text{for all } n \geq N.$$

Since $\varepsilon > 0$ was arbitrary, we may conclude that Y_n converges to 0 in L^1 and hence X_n converges to X in L^1. □

The following result follows immediately from Theorem A.89 and Lemma A.86(a). It is known as the *dominated convergence theorem.*

Theorem A.90 (Dominated convergence theorem)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(X_n)_{n\in\mathbb{N}}$ a sequence of integrable random variables that converges almost surely to random variable X. Suppose that there is an integrable random variable Y such that $|X_n| \leq Y$ P-a.s. for all $n \in \mathbb{N}$. Then X_n converges to X in L^1.*

Remark A.91. The direction "(1) ⇒ (2)" in Theorem A.89 is often referred to as *generalised dominated convergence theorem.*

A.14 The laws of large numbers

In this section, we study averages of random variables that are *independent and identically distributed* (i.i.d). If the random variables are integrable, these averages converge in probability and almost surely to their mean.

First, we study the convergence in probability, which is usually referred to as the *weak law of large numbers.*

Theorem A.92 (Weak law of large numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables in $\mathcal{L}^1$ with mean μ on some probability space $(\Omega, \mathcal{F}, P)$. Set $S_n := \sum_{i=1}^n X_i$ for*

$n \in \mathbb{N}$. *Then*

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

Proof. We show the result under the additional assumption that $E\left[(X_1)^2\right] < \infty$; the general case follows from Theorem A.93 below and the fact that almost sure convergence implies convergence in probability. Set $\sigma^2 := \mathrm{Var}[X^1]$. Then by the fact the X^i are i.i.d., we obtain by linearity of the expectation and the Bienaymé formula (A.16),

$$E\left[\frac{S_n}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X^i\right] = \frac{1}{n} n\mu = \mu, \tag{A.22}$$

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}, \tag{A.23}$$

Let $\varepsilon > 0$. By the Chebyshev's inequality (Corollary A.71), (A.22) and (A.23), we obtain for $n \in \mathbb{N}$,

$$P\left[\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right] = P\left[\left|\frac{S_n}{n} - E\left[\frac{S_n}{n}\right]\right| > \varepsilon\right] \leq \frac{1}{\varepsilon^2} \mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}.$$

Letting $n \to \infty$ establishes the claim. □

Next, we study the almost sure version of the law of large numbers. This is usually referred to as the *strong law of large numbers.*

Theorem A.93 (Strong law of large numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables in $\mathcal{L}^1$ with mean μ on some probability space $(\Omega, \mathcal{F}, P)$. Set $S_n := \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$. Then*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu.$$

Proof. We show the result under the additional assumption that $K := E\left[X_1^4\right] < \infty$; for the general case see [9, Theorem 5.16]. We may assume without loss of generality that $E\left[X_1\right] = 0$; otherwise consider $\tilde{X}_i := X_i - E\left[X_i\right]$ and use that $\frac{S_n}{n} = \frac{\tilde{S}_n}{n} + \mu$, where $\tilde{S}_n := \sum_{i=1}^n \tilde{X}_i$. Then independence of the X_i gives

$$\begin{aligned} E\left[X_i X_j X_k X_l\right] &= 0 \quad \text{for } i,j,k,l \in \mathbb{N} \text{ distinct},\\ E\left[X_i X_j X_k^2\right] &= 0 \quad \text{for } i,j,k \in \mathbb{N} \text{ distinct},\\ E\left[X_i X_j^3\right] &= 0 \quad \text{for } i,j \in \mathbb{N} \text{ distinct}, \end{aligned}$$

Moreover, Jensen's inequality and the fact that the X_i are i.i.d. gives

$$E\left[X_i^2 X_j^2\right] = E\left[X_i^2\right] E\left[X_j^2\right] = E\left[X_1^2\right]^2 \leq E\left[X_1^4\right] = K, \quad \text{for } i,j \in \mathbb{N} \text{ distinct}.$$

Thus, by some elementary combinatorics, we obtain

$$
\begin{aligned}
E\left[S_n^4\right] &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} E\left[X_iX_jX_kX_l\right] = \sum_{i=1}^{n} E\left[X_i^4\right] + 6\sum_{i=1}^{n}\sum_{j=1}^{i-1} E\left[X_i^2X_j^2\right] \\
&\leq nK + 6\frac{n(n-1)}{2}K \leq 3n^2K.
\end{aligned}
$$

Now using the fact that $\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty$, we obtain by monotone convergence,

$$
E\left[\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4\right] = \sum_{n=1}^{\infty}\frac{1}{n^4}E\left[S_n^4\right] \leq 3K\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty.
$$

By Lemma A.50, this implies that $\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4 < \infty$ P-a.s. It follows that $\lim_{n\to\infty}\left(\frac{S_n}{n}\right)^4 = 0$ P-a.s.[105] Hence, we have $\lim_{n\to\infty}\frac{S_n}{n} = 0$ P-a.s. □

A.15 The law of iterated logarithm and the central limit theorem

If $X_1, X_2, \ldots$ are i.i.d. random variables in $\mathcal{L}^1$ with mean μ then the strong law of large numbers implies that $\frac{\tilde{S}_n}{n}$ converges to 0 almost surely, where $\tilde{S}_n = \sum_{k=1}^{n}(X_k - \mu)$. Two natural follow-up questions are to understand the precise size of $\tilde{S}_n$ in terms of n and to find a nondegenerate limit in distribution under a different scaling in n.

The famous *law of iterated logarithm* by Hartman and Wintner answers the question on the precise size of $\tilde{S}_n$; for a proof see [9, Theorem 22.11].

Theorem A.94 (Law of iterated logarithm)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables in $\mathcal{L}^2$ with mean μ and variance $\sigma^2 > 0$ on some probability space $(\Omega, \mathcal{F}, P)$. Set $\tilde{S}_n = \sum_{k=1}^{n}(X_k - \mu)$ for $n \in \mathbb{N}$. Then*

$$
\limsup_{n\to\infty}\frac{\tilde{S}_n}{\sigma\sqrt{2n\log(\log n)}} = 1 \;\; P\text{-a.s.}
$$

The *central limit theorem* answers the second question. The correct scaling in n to get a nondegenerate weak limit is $\sqrt{n}$ and the corresponding distribution is the normal distribution. For a proof, we refer to [9, Theorem 15.37]

Theorem A.95 (Central limit theorem)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables in $\mathcal{L}^2$ with mean μ and variance $\sigma^2 > 0$ on some probability space $(\Omega, \mathcal{F}, P)$. Set $\tilde{S}_n := \sum_{k=1}^{n}(X_k - \mu)$ for $n \in \mathbb{N}$. Then*

$$
\frac{\tilde{S}_n}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}(0,1).
$$

[105] Recall that if $(a_n)_{n\in\mathbb{N}}$ is a sequence of nonnegative numbers with $\sum_{n=1}^{\infty} a_n < \infty$, then $\lim_{n\to\infty} a_n = 0$.

B The usual conditions

In this appendix, we briefly discuss the so-called usual conditions of right-continuity and completeness for a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$.

Definition B.1. Let $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space.

- The filtration $\mathbb{F}$ is said to be *right-continuous* if $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$ for all $t \geq 0$.
- $\mathcal{F}$ is said to be P-complete if every subset $\tilde{N}$ of a P-null set N in $\mathcal{F}$ is contained in $\mathcal{F}$.[106] In this case the the probability space $(\Omega, \mathcal{F}, P)$ is called *complete*.
- $\mathbb{F}$ is said to be P-complete if $\mathcal{F}$ is P-complete and $\mathcal{F}_0$ (and hence each $\mathcal{F}_t$) contains all P-null sets in $\mathcal{F}$.
- $\mathbb{F}$ is said to to satisfy the *usual conditions* if $\mathbb{F}$ is right-continuous and P-complete.

Remark B.2. (a) Working under the usual conditions has two main advantages: First, by modifying a process on a P-null set (which is always $\mathcal{F}_0$-measurable), we can assume *without loss of generality* that any property of the processes (like the continuity of its paths) that holds for P-almost all ω holds in fact for *all* ω. Moreover, by the *debut theorem* (Theorem 4.28), the first hitting time of *any* Borel set is a stopping time for continuous (or right-continuous) processes.

(b) If $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ is any filtration, the filtration $\mathbb{F}^+ := (\mathcal{F}_{t+})_{t\geq 0}$ is always right-continuous.

(c) If $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ is a filtered probability space such that $\mathcal{F}$ is P-compete, we can define the *completed filtration* $\mathbb{F}^P = (\mathcal{F}_t^P)_{t\geq 0}$ by

$$\mathcal{F}_t^P = \sigma\big(\mathcal{F}_t \cup \{N : N \text{ is } P\text{-null set in } \mathcal{F}\}\big).$$

Moreover, one can show that $(\mathbb{F}^+)^P = (\mathbb{F}^P)^+$, see [7, Lemma 7.8].

(d) If $W = (W_t)_{t\geq 0}$ is a Brownian motion on a complete probability space $(\Omega, \mathcal{F}, P)$, i.e., $\mathcal{F}$ is P-complete, then the completed natural filtration $(\mathbb{F}^W)^P$ is automatically right-continuous, i.e., it satisfies the usual conditions; see [7, Corollary 7.25]. Moreover, one can check that W is also a Brownian motion with respect to $(\mathbb{F}^W)^P$.

[106] $\tilde{N}$ is sometimes called a *P-negligible event*.

References

[1] C. Aliprantis and K. Border, *Infinite-dimensional analysis*, third ed., Springer-Verlag, Berlin, 2006.

[2] F. Delbaen and W. Schachermayer, *The mathematics of arbitrage*, Springer Finance, Springer-Verlag, Berlin, 2006.

[3] H. Föllmer and A. Schied, *Stochastic Finance*, 4th ext. ed., de Gruyter Studies in Mathematics, vol. 27, Walter de Gruyter & Co., Berlin, 2016.

[4] H.-O. Georgii, *Stochastics*, De Gruyter Textbook, Walter de Gruyter & Co., Berlin, 2008.

[5] J. Jacod and P. Protter, *Probability essentials*, second ed., Universitext, Springer-Verlag, Berlin, 2003.

[6] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003.

[7] O. Kallenberg, *Foundations of modern probability*, second ed., Springer-Verlag, New York, 2002.

[8] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.

[9] A. Klenke, *Probability theory*, Universitext, Springer-Verlag London, Ltd., London, 2008.

[10] P. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., Stochastic Modelling and Applied Probability, vol. 21, Springer, Berlin, 2005.

[11] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999.