MARTINGALE THEORY FOR FINANCE

SEBASTIAN ANDRES

CONTENTS

0. MOTIVATION

An investor needs a certain quantity of a share (or currency, good, ...), however not right now at time $t = 0$ but at a later time $t = 1$. The price of the share $S(\omega)$ at $t = 1$ is random and uncertain, but already now at time $t = 0$ one has to make some calcultations with it, which leads to the presence of risk. (Example: 500 USD are roughly 365 GBP today, but in one year?) A possible solution for the investor would be to purchase a financial derivative such as

- *Forward contract:* The owner of a forward contract has the right and the obligation to buy a share at time $t = 1$ for a delivery price K specified at time $t = 0$. Thus, the owner of the forward contract gains the difference between the actual market price $S(\omega)$ and the delivery price K if $S(\omega)$ is larger than K. If $S(\omega) < K$, the owner loses the amount $K - S(\omega)$ to the issuer of the forward contract. Hence, a forward contract corresponds to the random payoff
$$C(\omega) = S(\omega) - K.$$
- *Call option:* The owner of a call option has the right but not the obligation to buy a share at time $t = 1$ for a the strike price K specified at time $t = 0$. Thus, if $S(\omega) > K$ at time $t = 1$ the owner of the call gains again $S(\omega) - K$, but if $S(\omega) \leq K$ the owner buys the share from the market, and the call

Date: November 15, 2021.

becomes worthless in this case. Hence, at time $t = 1$ the random payoff of the call option is given by

$$C(\omega) = \big(S(\omega) - K\big)^+ = \begin{cases} S(\omega) - K & \text{if } S(\omega) > K, \\ 0 & \text{otherwise.} \end{cases}$$

What would be now a fair price for such a financial derivative?

A classical approach to this problem is to regard the random payoff $C(\omega)$ as a 'lottery' modelled as a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with some 'objective' probability measure $\mathbb{P}$. Then the fair price is given the expected discounted payoff $\mathbb{E}[\frac{C}{1+r}]$, where $r \geq 0$ is the interest rate for both fund and loans from $t = 0$ to $t = 1$. Here we implicitly assume that both interest rates are the same, which seems reasonable for large investors. The economic reason for working with discounted prices is that one should distinguish between payments at time $t = 0$ and ones at time $t = 1$. Usually, people tend to prefer a certain amount today over the same amount paid at a later time, and this preference is reflected in the interest rate r paid by the riskless bond (riskless asset, bank account). An investment of the amount $1/(1+r)$ at time zero in the bond results in value 1 at time $t = 1$.

Less classical approaches also take a subjective assessment of the risk by the involved agents (in this case buyer and seller of the derivative) into account, which leads to the usage of so-called utility functions.

In this lecture we will mainly focus on a more modern approach to option pricing. First let us assume for simplicity that the primary risk (the share in our example) can only be traded at $t = 0$ and $t = 1$. The idea is that the fair price of the derivative should equal the value of a so-called replicating strategy which may serve as a *hedging strategy*. Denote by

- θ^1 the number of shares held between $t = 0$ and $t = 1$,
- θ^0 the balance on a bank account with interest rate r.

Note that we allow both $\theta^i \geq 0$ and $\theta^i < 0$, where $\theta^1 < 0$ corresponds to a short sale of the share. Further, if π^1 denotes the price for one share at time $t = 0$, then the price of the strategy at $t = 0$ is

$$\theta^0 + \theta^1 \pi^1 =: V_0,$$

and the random value $V(\omega)$ of the strategy at $t = 1$ is given by

$$\theta^0(1 + r) + \theta^1 S(\omega) = V(\omega).$$

In order for a trading strategy (θ^0, θ^1) to be a replicating strategy for a derivative with random payoff function C, we require that for every possible event $\omega \in \Omega$ the value C of the derivative equals the value of the trading strategy, so

$$C(\omega) = V(\omega), \qquad \forall \omega \in \Omega.$$

In the example of a forward contract, i.e. $C = S - K$ this means

$$S(\omega) - K = V(\omega) = \theta^0(1+r) + \theta^1 S(\omega), \qquad \forall \omega \in \Omega,$$

which implies

$$\theta^1 = 1, \qquad \theta^0 = -\frac{K}{1+r}, \qquad V_0 = \pi^1 - \frac{K}{1+r}.$$

In particular, if the seller of C is using this strategy, all the risk is eliminated and the fair price $\pi(C)$ of C is given by V_0 since V_0 is the amount the seller needs for buying this strategy at $t = 0$. Moreover, $\pi(C) = V_0$ is the unique fair price for C as any other price would lead to *arbitrage*, i.e. a riskless opportunity to make profit, which should be excluded in any reasonable market model.

For example, consider a price $\tilde{\pi} > V_0$. Then, at time $t = 0$ one could sell the forward contract for $\tilde{\pi}$ and buy the above hedging strategy for V_0. At time $t = 1$ the strategy leads to a portfolio with one share and a balance of $-K$ in the bank account. Now we sell the share to the buyer of the forward for the delivery price K and repay the loan. We are left with a sure profit of $(\tilde{\pi} - V_0)(1+r) > 0$, so we have an arbitrage. These considerations lead us to the questions we will mainly address in this lecture course.

- How can arbitrage-free markets be characterised mathematically?
- How can one determine fair prices for options and derivatives?

1. Review of basic probability and measure theory

1.1. **Probability spaces and random variables.** First of all, we recall the notion of a probability space. Let Ω be a non-empty set. As an example, if we model the random price of a share in the context of a financial market, Ω can be thought of the set of all possible prices of the share, so for instance $\Omega_1 := \{0, 0.01, 0.02, \dots\}$ or $\Omega_2 := [0, \infty)$. If we consider several shares, say d, then suitable choices for Ω would be Ω_i^d, $i = 1, 2$. If we observe one share over T trading periods, meaning to look at the price at times $t = 0, 1, \dots, T$, we can take Ω as the set of mappings from $\{0, 1, \dots, T\}$ to Ω_i, $i = 1, 2$. Finally, if the price process evolves continuously in time (meaning that there is a price available at every time $t \in [0, T]$ rather than just for finitely many discrete times $t \in \{0, 1, \dots, T\}$ as in the previous case), then Ω can be chosen as set of function on $[0, T]$ with values in Ω_i, or, if we suppose the price process always to be continuous, the set of continuous functions on $[0, T]$ with values in Ω_2.

Definition 1.1. Let Ω be a non-empty set. A family $\mathcal{F}$ of subsets of Ω is called a *σ-algebra* if:

(i) $\Omega \in \mathcal{F}$.
(ii) If $A \in \mathcal{F}$ then $A^c := \Omega \setminus A \in \mathcal{F}$.
(iii) If $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ then $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

A set Ω together with a σ-algebra $\mathcal{F}$ of subsets of Ω is called a *measurable space* $(\Omega, \mathcal{F})$.

It is easy to see that the intersection of an arbitrary non-empty family of σ-algebras is again a σ-algebra. This gives rise to the following typical construction of σ-algebras. Let $\mathcal{C}$ be an arbitrary system of subsets of Ω. Then we say that $\mathcal{C}$ *generates* a σ-algebra $\sigma(\mathcal{C})$ defined as the smallest σ-algebra that contains $\mathcal{C}$,

$$\sigma(\mathcal{C}) := \bigcap_{\substack{\mathcal{A}\ \sigma\text{-algebra} \\ \mathcal{A} \supset \mathcal{C}}} \mathcal{A}.$$

Examples. (i) Let $\Omega = \mathbb{R}$. The *Borel σ-algebra $\mathcal{B}(\mathbb{R})$* is the σ-algebra generated by the open intervals,

$$\mathcal{B}(\mathbb{R}) = \sigma\big(\{(a,b),\ a, b \in \mathbb{R}, a < b\}\big).$$

Its elements are called *Borel sets*. The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is also generated by other families of subsets of $\mathbb{R}$, for instance the set of closed intervals, half-open intervals or the system of sets $\{(-\infty, x],\ x \in \mathbb{R}\}$.

(ii) For any Ω and any function $X : \Omega \to \mathbb{R}$,

$$\sigma(X) := \big\{\{X \in A\}, A \in \mathcal{B}(\mathbb{R})\big\},$$

is a σ-algebra, called the *σ-algebra generated by X*. (Here and below we write $\{X \in A\}$ shorthand for $\{\omega \in \Omega : X(\omega) \in A\}$ or $X^{-1}(A)$, respectively.)

(iii) Let $\Omega = \{\omega \,|\, \omega : \{0, \dots, T\} \to [0, \infty)\}$ be set of all non-negative functions on $\{0, \dots, T\}$. Define

$$X_t : \Omega \to [0, \infty) \quad \omega \mapsto X_t(\omega) = \omega(t), \qquad t \in \{0, \dots, T\}.$$

Then, $\sigma(X_t)$ is the σ-algebra of all events observable at time t, and

$$\mathcal{F}_t := \sigma\Big(\bigcup_{s=0}^{t} \sigma(X_s)\Big)$$

the σ-algebra of all events observable up to time t.

Definition 1.2. Let $(\Omega, \mathcal{F})$ be a measurable space. A function $X : \Omega \to \mathbb{R}$ is called *measurable* or a *random variable* if $\sigma(X) \subset \mathcal{F}$, i.e. $\{X \in A\} \in \mathcal{F}$ for all $A \in \mathcal{B}(\mathbb{R})$.

Remark 1.3. In order to verify that $X : \Omega \to \mathbb{R}$ is measurable it is sufficient to check that $\{X \in A\} \in \mathcal{F}$ for all $A \in \mathcal{E}$ for any system $\mathcal{E}$ such that $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$. For instance, X is measurable if and only if $\{X \le x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$.

Definition 1.4. Let $(\Omega, \mathcal{F})$ be a measurable space. A map $\mathbb{P} : \mathcal{F} \to [0,1]$ is called a *probability measure* if

(i) $\mathbb{P}[\Omega] = 1$.

(ii) For any countable family $(A_n)_{n \ge 0}$ of mutually disjoint elements of $\mathcal{F}$,

$$\mathbb{P}\Big(\bigcup_{n=0}^{\infty} A_n\Big) = \sum_{n=0}^{\infty} \mathbb{P}(A_n), \qquad (\sigma\text{-additivity}).$$

A triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

For a set A and sequence of sets $(A_n)_{n\geq 0}$ we write $A_n \uparrow A$ if $A_0 \subseteq A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_{n=0}^{\infty} A_n$, and we write $A_n \downarrow A$ if $A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots$ and $A = \bigcap_{n=0}^{\infty} A_n$.

Proposition 1.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Further, let $A \in \mathcal{F}$ and $(A_n)_{n\geq 0}$ be a sequence of elements in $\mathcal{F}$.*

(i) If $A_n \uparrow A$ then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.
(ii) If $A_n \downarrow A$ then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.
(iii) $\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \sum_{n=0}^{\infty} \mathbb{P}(A_n)$ (sub-additivity).

Proof. Exercise. □

Next we recall a very powerful uniqueness criterion for measures.

Theorem 1.6. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$, and let $\mathcal{C} \subset \mathcal{F}$ be closed under finite intersections (i.e. for any $A, B \in \mathcal{C}$ we also have that $A \cap B \in \mathcal{C}$). Suppose that $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \mathcal{C}$. Then $\mathbb{P}(A) = \mathbb{Q}(A)$ for all $A \in \sigma(\mathcal{C})$.*

Example 1.7. (i) Let $(\Omega, \mathcal{F})$ be a measurable space and $\omega_0 \in \Omega$ be fixed. Then, the probability measure δ_{ω_0} defined as

$$\delta_{\omega_0}(A) := \begin{cases} 1 & \text{if } \omega_0 \in A, \\ 0 & \text{if } \omega_0 \in A^c, \end{cases} \qquad A \in \mathcal{F},$$

is called *Dirac-measure* in ω_0.

(ii) Let Ω be a *countable* set and $\mathcal{F}$ be the power set of Ω (i.e. the family of all subsets of Ω). Then, every probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is uniquely determined by a weight function $p : \Omega \to [0, 1]$ satisfying $\sum_{\omega\in\Omega} p(\omega) = 1$. More precisely, for any $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega).$$

(iii) *Uniform distribution on* $(0, 1)$. Let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}((0, 1))$ be the Borel σ-algebra and $\mathbb{P}$ be the *Lebesgue measure* on $(0, 1)$. Then, for any $0 < a < b < 1$,

$$\mathbb{P}\big((a, b]\big) = b - a,$$

which uniquely determines $\mathbb{P}$ by Theorem 1.6.

Definition 1.8. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) The probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined as

$$\mu(A) := \mathbb{P}(X \in A), \qquad A \in \mathcal{B}(\mathbb{R}),$$

is called the *distribution* or the *law* of X (notation: $\mu = \mathbb{P} \circ X^{-1}$).

(ii) The function $F : \mathbb{R} \to [0, 1]$ defined by

$$F(x) := \mathbb{P}(X \leq x) = \mu((-\infty, x]), \qquad x \in \mathbb{R},$$

is called the *distribution function* of X.

Note that by Theorem 1.6 the law of a random variable is uniquely determined by its distribution function. Further, it is easy to see that the distribution function F of a random variable X is monotone increasing, right-continuous with $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

1.2. **Integration with respect to probability measures.**

1.2.1. *Definition of the integral.* We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will now review the definition of the integral of a random variable X against a probability measure $\mathbb{P}$, which will be denoted by $\int X \, d\mathbb{P}$. This will be carried out in three steps. For any $A \in \mathcal{F}$, we recall that the indicator function $\mathbb{1}_A : \Omega \to \{0, 1\}$ is defined as

$$\mathbb{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \in A^c. \end{cases}$$

Step 1: Integral of non-negative discrete random variables. Let X be a non-negative discrete random variable taking values in $\{x_1, x_2, \ldots x_n\}$ and set $A_i := \{X = x_i\}$. Then X is of the form

$$X(\omega) = \sum_{i=1}^{n} x_i \, \mathbb{1}_{A_i}(\omega).$$

Then we define the integral $\int X \, d\mathbb{P} = \int X(\omega) \, \mathbb{P}(d\omega)$ of X by

$$\int X \, d\mathbb{P} := \sum_{i=1}^{n} x_i \, \mathbb{P}(A_i) = \sum_{i=1}^{n} x_i \, \mathbb{P}(X = x_i).$$

One can check that this definition does not depend on the chosen representation of X. The integral is also monotone, i.e. $\int X \, d\mathbb{P} \leq \int Y \, d\mathbb{P}$ for any two discrete random variables X and Y with $X \leq Y$.

Step 2: Integral of non-negative random variables. Let X be a non-negative random variable. Then there exists a sequence $(X_n)_{n\geq 0}$ of non-negative discrete random variables such that $X_n \uparrow X$ pointwise as $n \to \infty$. Indeed, one can take, for instance,

$$X_n := \sum_{k=1}^{n2^n} (k-1) \, 2^{-n} \, \mathbb{1}_{\{(k-1)2^{-n} \leq X < k2^{-n}\}} + n\mathbb{1}_{\{X \geq n\}}.$$

Then the integral $\int X \, d\mathbb{P} = \int X(\omega) \, \mathbb{P}(d\omega)$ is defined as

$$\int X \, d\mathbb{P} := \lim_{n\to\infty} \int X_n \, d\mathbb{P} \in [0, \infty].$$

Note that the right hand side is well defined as a limit of a monotone sequence. Again, one can check that the integral does not depend on the particular choice of the sequence (X_n). Thus, the integral of a non-negative random variable is always defined, but it is possibly infinite. In particular, the integral is *monotone*, i.e. if $X \leq Y$ then $\int X \, d\mathbb{P} \leq \int Y \, d\mathbb{P}$, and *linear*, i.e. for $a, b > 0$ we have $\int (aX + bY) \, d\mathbb{P} = a \int X \, d\mathbb{P} + b \int Y \, d\mathbb{P}$. For $A \in \mathcal{F}$ we also write $\int_A X \, d\mathbb{P} := \int (\mathbb{1}_A X) \, d\mathbb{P}$.

Step 3: The general case.

Definition 1.9. A random variable is called $\mathbb{P}$-integrable if $\int |X|\, d\mathbb{P} < \infty$. In this case we write $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ (or $X \in \mathcal{L}^1(\mathbb{P})$, $X \in \mathcal{L}^1$ in short).

Since $|X| = X^+ + X^-$ (with $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$), we have $X \in \mathcal{L}^1(\mathbb{P})$ if and only if $\int X^+\, d\mathbb{P} < \infty$ and $\int X^-\, d\mathbb{P} < \infty$. For $X \in \mathcal{L}^1(\mathbb{P})$ the integral $\int X\, d\mathbb{P} = \int X(\omega)\, \mathbb{P}(d\omega)$ is defined by

$$\int X\, d\mathbb{P} := \int X^+\, d\mathbb{P} - \int X^-\, d\mathbb{P}\,.$$

For $A \in \mathcal{F}$ we also write $\int_A X\, d\mathbb{P} := \int (\mathbb{1}_A X)\, d\mathbb{P}$ for the integral of X over A. We now collect a few properties of the integral.

Proposition 1.10. *Let* $X, Y \in \mathcal{L}^1$.

(i) *Monotonicity:* $X \le Y \Rightarrow \int X\, d\mathbb{P} \le \int Y\, d\mathbb{P}$.
(ii) *Linearity:* $aX + bY \in \mathcal{L}^1$ *and* $\int (aX + bY)\, d\mathbb{P} = a \int X\, d\mathbb{P} + b \int Y\, d\mathbb{P}$ *for all* $a, b \in \mathbb{R}$.
(iii) *Additivity:* $\int_{A \cup B} X\, d\mathbb{P} = \int_A X\, d\mathbb{P} + \int_B X\, d\mathbb{P}$ *for all* $A, B \in \mathcal{F}$ *with* $A \cap B = \emptyset$.

Definition 1.11. A property, which the elements of Ω may or may not have, is said to hold $\mathbb{P}$-*almost surely* ($\mathbb{P}$-a.s. in short), if the set of all $\omega \in \Omega$, for which the property does not hold, is contained in a set $N \in \mathcal{F}$ with $\mathbb{P}(N) = 0$.

Example 1.12. Let X, Y and X_n, $n \in \mathbb{N}$, be random variables. Then

(i) $X = Y$ $\mathbb{P}$-a.s. $\iff \mathbb{P}(\{\omega : X(\omega) \ne Y(\omega)\}) = 0$.
(ii) $\lim_{n\to\infty} X_n = X$ $\mathbb{P}$-a.s. $\iff \mathbb{P}(\{\omega : \lim_{n\to\infty} X_n(\omega)$ does not exist or is not equal to $X(\omega)\}) = 0$.

Theorem 1.13. *Let* $X, Y \in \mathcal{L}^1$.

(i) $X = Y$ $\mathbb{P}$-*a.s.* $\Rightarrow \int X\, d\mathbb{P} = \int Y\, d\mathbb{P}$.
(ii) $X \ge Y$ *and* $\int X\, d\mathbb{P} = \int Y\, d\mathbb{P} \Rightarrow X = Y$ $\mathbb{P}$-*a.s.*

1.2.2. *Convergence theorems.* We will now address the following question. Suppose that $X_n \to X$ $\mathbb{P}$-a.s. Is it true that $\int X_n\, d\mathbb{P} \to \int X\, d\mathbb{P}$? In other words, are we allowed to interchange limit and integration? In general the answer is no as the following example shows.

Example 1.14. Let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}((0, 1))$ be the Borel σ-algebra and $\mathbb{P}$ be the *Lebesgue measure* on $(0, 1)$ (cf. Example 1.7-(iii) above). Consider

$$X_n(\omega) = \begin{cases} n & \text{if } \frac{1}{n} \le \omega \le \frac{2}{n}, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $X_n(\omega) \to X(\omega) := 0$ for all $\omega \in (0, 1)$, but

$$\int X_n\, d\mathbb{P} = 1 \not\to \int X\, d\mathbb{P} = 0.$$

But there are many cases where the question above has a positive answer.

Theorem 1.15 (Monotone convergence)**.** *Let $X, X_1, X_2, \dots$ be random variables such that*

$$X_1 \le X_2 \le X_3 \le \dots \ \mathbb{P}\text{-a.s.} \quad \text{and} \quad \lim_{n\to\infty} X_n = X \ \mathbb{P}\text{-a.s.}$$

Then

$$\lim_{n\to\infty} \int X_n \, d\mathbb{P} = \int X \, d\mathbb{P} \,.$$

Lemma 1.16 (Fatou's lemma)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of non-negative random variables. Then*

$$\int \liminf_{n\to\infty} X_n \, d\mathbb{P} \le \liminf_{n\to\infty} \int X_n \, d\mathbb{P} \,.$$

The proof is by applying monotone convergence to the non-decreasing sequence of random variables $(\inf_{m\ge n} X_m : n \in \mathbb{N})$.

Theorem 1.17 (Dominated convergence)**.** *Let $X, X_1, X_2, \dots$ be random variables such that $\lim_{n\to\infty} X_n = X$ $\mathbb{P}$-a.s. Suppose that there exists a random variable $Y \in \mathcal{L}^1$ such that $|X_n| \le Y$ for all n. Then*

$$\lim_{n\to\infty} \int X_n \, d\mathbb{P} = \int X \, d\mathbb{P} \,.$$

The proof is by applying Fatou's lemma to the two sequences of non-negative random variables $(Y \pm X_n)_{n\in\mathbb{N}}$, we omit the details here.

1.2.3. *Product spaces and Fubini's theorem.* Fubini's theorem is an always important tool for the computation of integrals on product spaces. Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be two probability spaces. The product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ is the σ-algebra on $\Omega_1 \times \Omega_2$ generated by subsets of the form $A_1 \times A_2$ for $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$.

Theorem 1.18 (Product measure)**.** *There exists a unique measure $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ on $\mathcal{F}_1 \otimes \mathcal{F}_2$ such that, for all $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$,*

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\, \mathbb{P}_2(A_2).$$

We only state here Fubini's theorem for non-negative random variables.

Theorem 1.19 (Fubini-Tonnelli)**.** *Let X be a non-negative random variable on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$. Then*

$$Y(\omega_1) := \int_{\Omega_2} X(\omega_1, \omega_2)\ \mathbb{P}_2(d\omega_2) \quad \text{and} \quad Z(\omega_2) := \int_{\Omega_1} X(\omega_1, \omega_2)\ \mathbb{P}_1(d\omega_1)$$

are measurable with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, and

$$\int_{\Omega_1 \times \Omega_2} X \, d(\mathbb{P}_1 \otimes \mathbb{P}_2) = \int_{\Omega_1} Y \, d\mathbb{P}_1 = \int_{\Omega_2} Z \, d\mathbb{P}_2 \,.$$

This is more usually written as

$$\int_{\Omega_1}\left(\int_{\Omega_2} X(\omega_1,\omega_2)\,\mathbb{P}_2(d\omega_2)\right)\mathbb{P}_1(d\omega_1) = \int_{\Omega_2}\left(\int_{\Omega_1} X(\omega_1,\omega_2),\mathbb{P}_1(d\omega_1)\right)\mathbb{P}_2(d\omega_2),$$

so the order of integration can be exchanged if X is non-negative. A similar statement holds if $X \in \mathcal{L}^1(\mathbb{P}_1 \otimes \mathbb{P}_2)$.

1.2.4. *Expected value and Jensen's inequality.* In the context of random variables on probability spaces the integral is usually called the expected value.

Definition 1.20. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $X \geq 0$ or $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathbb{E}[X] := \int X\,d\mathbb{P}$ is called the *expected value* or *expectation of* X.

From now on we will mostly use this terminology for the integral of a random variable. While most of the above discussion on integration directly extends to more general measure spaces, the next result is restricted to probability measures (or at least to finite measures).

Theorem 1.21 (Jensen's inequality). *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, and let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function. Then,*

$$\phi\big(\mathbb{E}[X]\big) \leq \mathbb{E}\big[\phi(X)\big].$$

1.3. **Measures with densities.** Consider a non-negative measurable function $\phi : \Omega \to [0, \infty)$ with $\mathbb{E}[\phi] = 1$. Then, the mapping $\mathbb{Q} : \mathcal{F} \to [0, \infty)$ defined by

$$\mathbb{Q}[A] := \int_A \phi\,\mathrm{d}\mathbb{P} = \mathbb{E}\big[\phi \mathbb{1}_A\big], \qquad A \in \mathcal{F}, \tag{1.1}$$

is a probability measure on $(\Omega, \mathcal{F})$. Indeed, $\mathbb{Q}(\Omega) = \mathbb{E}[\phi] = 1$, so $\mathbb{Q}$ only takes values in $[0, 1]$, and for any sequence $(A_n)_{n\geq 1}$ of mutually disjoint sets in $\mathcal{F}$ and $A = \bigcup_{n=1}^{\infty} A_n$,

$$\mathbb{Q}(A) = \int_A \phi\,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_{\bigcup_{k=1}^n A_k} \phi\,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \sum_{k=1}^{n} \int_{A_k} \phi\,\mathrm{d}\mathbb{P} = \sum_{k=1}^{\infty} \mathbb{Q}(A_k),$$

so $\mathbb{Q}$ is σ-additive. Here we used the monotone convergence theorem in the second step and the additivity of the integral (see Proposition 1.10-(iii)) in the third step. Moreover, note that for any $A \in \mathcal{F}$,

$$\mathbb{P}[A] = 0 \;\Rightarrow\; \mathbb{Q}[A] = 0,$$

since $\mathbb{P}(A) = 0$ implies that $\phi\mathbb{1}_A = 0$ $\mathbb{P}$-a.s., which in turn implies that $0 = \mathbb{E}[\phi\mathbb{1}_A] = \mathbb{Q}(A)$. Hence, the measure $\mathbb{Q}$ defined in (1.1) above is absolutely continuous with respect to $\mathbb{P}$ in the following sense.

Definition 1.22. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$. We say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$ if

$$\mathbb{P}[A] = 0 \;\Rightarrow\; \mathbb{Q}[A] = 0, \qquad \forall A \in \mathcal{F}.$$

In this case we also write $\mathbb{Q} \ll \mathbb{P}$.

It turns out that all probability measures that are absolutely continuous with respect to $\mathbb{P}$ are of the form (1.1).

Theorem 1.23 (Radon-Nikodym)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$. Then the following are equivalent.*

(i) $\mathbb{Q} \ll \mathbb{P}$.

(ii) There exists an $\mathcal{F}$-measurable function $\phi \geq 0$ with $\mathbb{E}[\phi] = 1$ such that

$$\mathbb{Q}[A] = \int_A \phi \, \mathrm{d}\mathbb{P} = \mathbb{E}\big[\phi \mathbb{1}_A\big], \qquad \forall A \in \mathcal{F}.$$

The function ϕ is called density *or Radon-Nikodym derivative and is often denoted by $\phi(\omega) = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(\omega)$.*

Proof. We have shown the implication (ii) $\Rightarrow$ (i) right before Definition 1.22. The other implication is more difficult, see, for instance, in [2, Chapter 17]. □

Remark 1.24. (i) The density $\phi = d\mathbb{Q}/d\mathbb{P}$ is uniquely determined by $\mathbb{P}$ and $\mathbb{Q}$ up to $\mathbb{P}$-null sets, i.e. for any other $\mathcal{F}$-measurable function $\tilde{\phi}$ satisfying $\mathbb{Q}(A) = \mathbb{E}[\tilde{\phi}\mathbb{1}_A]$ we have $\mathbb{P}[\tilde{\phi} = \phi] = 1$.

(ii) Theorem 1.23 also holds for a more general class of measures. For instance, if we still assume $\mathbb{P}$ to be a probability measure then the statement holds for any measure $\mathbb{Q}$. In this case the $\mathbb{P}$-integral of the density is not necessarily equal to one.

From now on we write $\mathbb{E}_{\mathbb{Q}}$ and $\mathbb{E}_{\mathbb{P}}$ for the expectation w.r.t. to the probability measures $\mathbb{Q}$ and $\mathbb{P}$, respectively. When writing $\mathbb{E}$ without subscript we mean, as before, the expectation w.r.t. $\mathbb{P}$.

Proposition 1.25. *Let $\mathbb{Q} \ll \mathbb{P}$ with density ϕ and let X be a random variable.*

(i) If $X \geq 0$ then

$$\mathbb{E}_{\mathbb{Q}}[X] = \mathbb{E}_{\mathbb{P}}[\phi X]. \tag{1.2}$$

(ii) $X \in \mathcal{L}^1(\mathbb{Q})$ if and only if $\phi X \in \mathcal{L}^1(\mathbb{P})$ and in this case (1.2) holds.

Sketch of proof. The proof follows the usual extension argument in measure theory. First, it is clear from the definition of the density that (1.2) holds if X is an indicator function. By the linearity of the expectation this immediately extends to random variables of the form $X = \sum_{i=1}^n x_i \, \mathbb{1}_{A_i}$ with $A_i \in \mathcal{F}$, $x_i \geq 0$, $n \in \mathbb{N}$. For every $X \geq 0$ there exists a sequence (X_n) of such 'simple' functions such that $X_n \uparrow X$, so we apply the monotone convergence theorem twice to see that

$$\mathbb{E}_{\mathbb{Q}}[X] = \lim_{n\to\infty} \mathbb{E}_{\mathbb{Q}}[X_n] = \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}}[\phi X_n] = \mathbb{E}_{\mathbb{P}}[\phi X],$$

which proves (i). Statement (ii) follows from (i) by decomposing X into its positive and negative part. □

Example 1.26. (i) Let Ω be a countable set, $\mathcal{F}$ be the power set of Ω and $\mathbb{P}$ be any probability measure on $(\Omega,\mathcal{F})$ with $\mathbb{P}[\{\omega\}]>0$ for all $\omega\in\Omega$. Then every probability measure $\mathbb{Q}$ on $(\Omega,\mathcal{F})$ is absolutely continuous with respect to $\mathbb{P}$ and the density is given by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(\omega)=\phi(\omega):=\frac{\mathbb{Q}[\{\omega\}]}{\mathbb{P}[\{\omega\}]},\qquad \omega\in\Omega.$$

Indeed, for any $A\in\mathcal{F}$,

$$\mathbb{Q}[A]=\sum_{\omega\in A}\mathbb{Q}\big[\{\omega\}\big]=\sum_{\omega\in A}\frac{\mathbb{Q}\big[\{\omega\}\big]}{\mathbb{P}\big[\{\omega\}\big]}\ \mathbb{P}\big[\{\omega\}\big]=\sum_{\omega\in A}\phi(\omega)\ \mathbb{P}\big[\{\omega\}\big]=\mathbb{E}_{\mathbb{P}}\big[\phi\mathbb{1}_A\big].$$

(ii) Let $\Omega=(0,1)$, $\mathcal{F}=\mathcal{B}((0,1))$ be the Borel σ-algebra and $\mathbb{P}$ be the *Lebesgue measure* on $(0,1)$. Then, $\mathbb{Q}\ll\mathbb{P}$ if and only if there exists a Borel-measurable function $\phi:(0,1)\to[0,\infty)$ satisfying $\int\phi\,d\mathbb{P}=\int_0^1\phi(x)\,dx=1$ so that $\mathbb{Q}(A)=\int_A\phi(x)\,dx$.[1]

Remark 1.27. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega,\mathcal{F})$ with $\mathbb{Q}\ll\mathbb{P}$. Then there exists an $\mathcal{F}$-measurable function $\phi\geq 0$ such that $\mathbb{E}_{\mathbb{Q}}[Y]=\mathbb{E}[\phi Y]$ for all random variables $Y\geq 0$. Consider now a σ-algebra $\mathcal{F}_0\subseteq\mathcal{F}$. Then, $\mathbb{P}$ and $\mathbb{Q}$ are also probability measures on $(\Omega,\mathcal{F}_0)$ with $\mathbb{Q}\ll\mathbb{P}$ on $\mathcal{F}_0$, i.e. $\mathbb{Q}(A)=0$ for all $A\in\mathcal{F}_0$ with $\mathbb{P}(A)=0$. By the Radon-Nikodym theorem there exists an $\mathcal{F}_0$-measurable function $\phi_0\geq 0$ such that $\mathbb{E}_{\mathbb{Q}}[Y_0]=\mathbb{E}[\phi_0 Y_0]$ for all $\mathcal{F}_0$ measurable random variables $Y_0\geq 0$. In particular, $\mathbb{E}[\phi Y_0]=\mathbb{E}[\phi_0 Y_0]$ or all $\mathcal{F}_0$ measurable random variables $Y_0\geq 0$, but $\phi\neq\phi_0$ in general.

Lemma 1.28. *Let $\mathbb{Q}\ll\mathbb{P}$. Then $\frac{d\mathbb{Q}}{d\mathbb{P}}>0$, $\mathbb{Q}$-a.s. (but not necessarily $\mathbb{P}$-a.s.)*

Proof. Set $\phi:=\frac{d\mathbb{Q}}{d\mathbb{P}}$. Then, for any random variable $X\geq 0$,

$$\mathbb{E}_{\mathbb{Q}}[X]=\mathbb{E}[\phi X]=\mathbb{E}[\phi X\mathbb{1}_{\{\phi>0\}}]=\mathbb{E}_{\mathbb{Q}}[X\mathbb{1}_{\{\phi>0\}}].$$

We get the claim by choosing $X=\mathbb{1}_{\{\phi=0\}}$. $\square$

Definition 1.29. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega,\mathcal{F})$. We say that $\mathbb{Q}$ is *equivalent* to $\mathbb{P}$ if $\mathbb{Q}\ll\mathbb{P}$ and $\mathbb{P}\ll\mathbb{Q}$, that is

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

In this case we write $\mathbb{Q}\approx\mathbb{P}$.

Proposition 1.30. *Let $\mathbb{Q}\ll\mathbb{P}$. Then $\mathbb{Q}\approx\mathbb{P}$ if and only if $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}>0$, $\mathbb{P}$-a.s. In this case*

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}=\Big(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big)^{-1}.$$

[1] In analysis, a function $f:(0,1)\to\mathbb{R}$ of the form $f(x)=\int_0^x\phi(y)\,dy$ is absolutely continuous in the following sense. For all $\varepsilon>0$ exists $\delta>0$ such that $\sum_{i=1}^n|f(b_i)-f(a_i)|<\varepsilon$ for all disjoint intervals $[a_1,b_1],\dots,[a_n,b_n]\subset(0,1)$ with $\sum_{i=1}^n|b_i-a_i|<\delta$. This is the reason for the terminology.

Proof. First note that $\phi := \frac{d\mathbb{Q}}{d\mathbb{P}} > 0$, $\mathbb{Q}$-a.s., by Lemma 1.28. Hence, for all measurable $X \geq 0$,

$$\mathbb{E}_{\mathbb{Q}}\left[\phi^{-1}X\right] = \mathbb{E}_{\mathbb{Q}}\left[\phi^{-1}X\mathbb{1}_{\{\phi>0\}}\right] = \mathbb{E}\left[\phi^{-1}X\mathbb{1}_{\{\phi>0\}}\phi\right] = \mathbb{E}\left[X\mathbb{1}_{\{\phi>0\}}\right],$$

and this is equal to $\mathbb{E}[X]$ for all $X \geq 0$ if and only if $\mathbb{P}[\phi = 0] = 0$. □

2. Conditional expectations

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, its expected value $\mathbb{E}[X]$, provided it exists, serves as a prediction for the random outcome of X.

Our goal is now to introduce an object, which allows us to improve the prediction for X if additional information is available. In the special case where this additional information can be encoded in a single event B having positive probability, this can be achieved rather easily by conditioning on B.

Definition 2.1. Let $B \in \mathcal{F}$ with $\mathbb{P}[B] > 0$. Then, for any $A \in \mathcal{F}$,

$$\mathbb{P}\left[A \,|\, B\right] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

is called *conditional probability of A given B* and for a random variable X,

$$\mathbb{E}\left[X \,|\, B\right] = \frac{\mathbb{E}[X\mathbb{1}_B]}{\mathbb{P}[B]}$$

is called the *conditional expectation of X given B.*

$\mathbb{P}\left[\,\cdot \,|\, B\right]$ is again a probability distribution and $\mathbb{E}\left[X \,|\, B\right]$ is the expected value of X under $\mathbb{P}\left[\,\cdot\,|\, B\right]$. If we regard $\mathbb{P}[A]$ as a prediction about the occurence of A and the expected value as a prediction for the value of a random variable, then the conditional probability and the conditional expectation are improved predictions under the assumption that we know that the event B occurs.

We will now generalise the notion of conditional expectations and conditional probabilities considerably, because so far it only allows us to condition on events of positive probability which is too restrictive. We will first discuss the easier discrete case before we will give the general definition.

2.1. **The elementary case.** A typical problem might be the following situation.

Example 2.2. The day after tomorrow it will be decided whether a certain event A occurs (for instance $A = \{\text{Dow Jones} \geq 10000\}$). Already today we can compute $\mathbb{P}[A]$. But what prediction would we make tomorrow night, when we have more information available (e.g. the value of the Dow Jones in the evening)? Then we would like to consider the conditional probability

$$\mathbb{P}\left[A \,|\, \text{Dow Jones tomorrow} = x\right], \qquad x = 0, 1, \ldots$$

as a function of x.

As mentioned before, our goal is to formalise predicitions under additional information. But how do we model additional information? We will use a σ-algebra $\mathcal{F}_0 \subset \mathcal{F}$. This σ-algebra contains the events, about which we will know tomorrow (in the context of Example 2.2 above) if they occur or not, so for instance

$$\mathcal{F}_0 = \sigma\big(\{\text{Dow Jones tomorrow} = x\},\ x = 0, 1, \dots\big).$$

More generally, let now $B_1, B_2, \dots$ be a decomposition of Ω into w.l.o.g. disjoint sets $B_i \in \mathcal{F}$ and set

$$\mathcal{F}_0 := \sigma(B_1, B_2, \dots) = \big\{\text{all possible unions of } B_i\text{'s}\big\} \subseteq \mathcal{F}.$$

Recall that by definition $\sigma(B_1, B_2, \dots)$ denotes the smallest σ-algebra in which all the sets $B_1, B_2, \dots$ are contained.

Definition 2.3. The random variable

$$\mathbb{E}\big[X \,|\, \mathcal{F}_0\big](\omega) := \sum_{i:\mathbb{P}[B_i]>0} \mathbb{E}\big[X \,|\, B_i\big]\, \mathbb{1}_{B_i}(\omega) \tag{2.1}$$

is called *conditional expectation of X given $\mathcal{F}_0$*.

Example 2.4. If $\mathcal{F}_0 = \{\emptyset, \Omega\}$, then $\mathbb{E}\big[X \,|\, \mathcal{F}_0\big](\omega) = \mathbb{E}[X]$.

We briefly recall what it means for a real-valued random variable to be measurable with respect to a σ-algebra.

Definition 2.5. Let $\mathcal{A} \subseteq \mathcal{F}$ be a σ-algebra over Ω. Then, a random variable $Y : \Omega \to \mathbb{R}$ is $\mathcal{A}$-measurable if $\{Y \le c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$.

Proposition 2.6. *The random variable $X_0 = \mathbb{E}\big[X \,|\, \mathcal{F}_0\big]$ has the following properties.*

(i) X_0 is $\mathcal{F}_0$-measurable.
(ii) For all $A \in \mathcal{F}_0$,

$$\mathbb{E}\big[X \mathbb{1}_A\big] = \mathbb{E}\big[X_0 \mathbb{1}_A\big].$$

Proof. (i) For every i we have that $\mathbb{1}_{B_i}$ is $\mathcal{F}_0$-measurable. Since X_0 is a linear combination of such functions, it is $\mathcal{F}_0$-measurable as well.

(ii) Let us first consider the case that $A = B_i$ for any i such that $\mathbb{P}[B_i] > 0$. Then,

$$\begin{aligned}\mathbb{E}\big[X\mathbb{1}_A\big] &= \mathbb{E}\big[X\mathbb{1}_{B_i}\big] = \mathbb{E}\big[X \,|\, B_i\big]\ \mathbb{P}[B_i] = \mathbb{E}\big[X \,|\, B_i\big]\ \mathbb{E}[\mathbb{1}_{B_i}] \\ &= \mathbb{E}\Big[\underbrace{\mathbb{E}\big[X \,|\, B_i\big]}_{=X_0 \text{ on } B_i} \mathbb{1}_{B_i}\Big] = \mathbb{E}\big[X_0\mathbb{1}_A\big].\end{aligned}$$

For general $A \in \mathcal{F}_0$, $\mathbb{1}_A$ can be written as a (possibly infinite) sum of $\mathbb{1}_{B_i}$'s (recall that the sets $B_1, B_2, \dots$ are disjoint), so (ii) follows from the linearity of the expectation and the monotone convergence theorem. $\square$

Example 2.7. (i) Consider the probability space $((0,1], \mathcal{B}((0,1]), \lambda)$, where $\mathcal{B}((0,1])$ denotes the Borel-σ-algebra and λ the Lebesgue-measure. For any $n \in \mathbb{N}$, let $\mathcal{F}_0 = \sigma((\frac{k}{n}, \frac{k+1}{n}], k = 0, \dots, n-1)$. Then, on each interval $(\frac{k}{n}, \frac{k+1}{n}]$ the random variable $\mathbb{E}\big[X \,|\, \mathcal{F}_0\big]$ is constant and coincides with the average of X over this interval.

(ii) Let $Z : \Omega \to \{z_1, z_2, \dots\} \subset \mathbb{R}$ and

$$\mathcal{F}_0 = \sigma(Z) = \sigma\big(\{Z = z_i\}, i = 1, 2, \dots\big).$$

(In general, for any real-valued random variable Z, $\sigma(Z) = \sigma(\{Z \le c\}, c \in \mathbb{R})$ denotes the smallest σ-algebra with respect to which Z is measurable.) Then,

$$\mathbb{E}\big[X \,|\, Z\big] := \mathbb{E}\big[X \,|\, \sigma(Z)\big] = \sum_{i : \mathbb{P}[Z = z_i] > 0} \mathbb{E}\big[X \,|\, Z = z_i\big] \, \mathbb{1}_{\{Z = z_i\}}.$$

In particular, $\mathbb{E}\big[X \,|\, Z\big](\omega) = \mathbb{E}\big[X \,|\, Z = Z(\omega)\big]$, so $\mathbb{E}\big[X \,|\, Z\big]$ describe the expectation of X if Z is known.

However, if Z would have a continuous distribution (e.g. $\mathcal{N}(0,1)$, then $\mathbb{P}[Z = z] = 0$ for all $z \in \mathbb{R}$ and $\mathbb{E}\big[X \,|\, Z\big]$ is not defined yet.

2.2. **The general case.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_0 \subseteq \mathcal{F}$ be a σ-algebra.

Definition 2.8. Let $X \ge 0$ be a random variable. A random variable X_0 is called (a version of) the *conditional expectation of X given $\mathcal{F}_0$* if

(i) X_0 is $\mathcal{F}_0$-measurable.
(ii) For all $A \in \mathcal{F}_0$,

$$\mathbb{E}\big[X \mathbb{1}_A\big] = \mathbb{E}\big[X_0 \mathbb{1}_A\big]. \tag{2.2}$$

In this case we write $X_0 = \mathbb{E}\big[X \,|\, \mathcal{F}_0\big]$.

If $X \in \mathcal{L}^1(\Omega, \mathbb{P})$, i.e. $\mathbb{E}[|X|] < \infty$, (but not necessarily non-negative) we decompose X into its positive and negative part $X = X^+ - X^-$ and define

$$\mathbb{E}\big[X \,|\, \mathcal{F}_0\big] := \mathbb{E}\big[X^+ \,|\, \mathcal{F}_0\big] - \mathbb{E}\big[X^- \,|\, \mathcal{F}_0\big].$$

Remark 2.9. (i) If $\mathcal{F}_0 = \sigma(\mathcal{C})$ for any $\mathcal{C} \subseteq \mathcal{F}$, then it suffices to check condition (ii) for all $A \in \mathcal{C}$.

(ii) If $\mathcal{F}_0 = \sigma(Z)$ for any random variable Z, then $\mathbb{E}\big[X \,|\, Z\big] := \mathbb{E}\big[X \,|\, \sigma(Z)\big]$ is $\sigma(Z)$-measurable by condition (i). In particular, by the so-called factorisation lemma (see e.g. [2]) it is of the form $f(Z)$ for some function f. It is then common to define

$$\mathbb{E}\big[X \,|\, Z = z\big] := f(z).$$

(iii) If $X \in \mathcal{L}^1$ then $\mathbb{E}\big[X \,|\, \mathcal{F}_0\big] \in \mathcal{L}^1$. Indeed, if $X \ge 0$, by choosing $A = \Omega$ in (2.2) we have

$$\mathbb{E}\big[X\big] = \mathbb{E}\Big[\mathbb{E}\big[X \,|\, \mathcal{F}_0\big]\Big].$$

For general $X \in \mathcal{L}^1$ we can use again the decompostion $X = X^+ - X^-$.

(iv) The weakest possible condition on X under which a definition of conditional expectation can make sense is that $\mathbb{E}[X^+] < \infty$ or $\mathbb{E}[X^-] < \infty$. (Note that $X \in \mathcal{L}^1$ if and only if both hold.) In this case $\mathbb{E}\big[X \,|\, \mathcal{F}_0\big]$ can still be defined to be a random variable X_0 satisfying (i) and (ii) in Definition 2.8.

Theorem 2.10 (Existence and uniqueness)**.** *For any $X \geq 0$ the following hold.*

(i) The conditional expectation $\mathbb{E}\big[X \,|\, \mathcal{F}_0\big]$ exists.
(ii) Any two versions of $\mathbb{E}\big[X \,|\, \mathcal{F}_0\big]$ coincide $\mathbb{P}$-a.s.

Proof. (i) The existence follows from the Radon-Nikodym theorem. Define $\mu(A) = \int_A X \,\mathrm{d}\mathbb{P}$, $A \in \mathcal{F}$. Then μ is a measure and $\mu \ll \mathbb{P}$ on $\mathcal{F}$. In particular, $\mu \ll \mathbb{P}$ also on $\mathcal{F}_0$. We apply Theorem 1.23 (in the more general version mentioned in Remark 1.24-(ii)) on the space $(\Omega, \mathcal{F}_0)$ to obtain that there exists an $\mathcal{F}_0$-measurable function

$$X_0 = \frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_0} \qquad \text{such that} \qquad \mu(A_0) = \int_{A_0} X_0 \,\mathrm{d}\mathbb{P}, \quad \forall A_0 \in \mathcal{F}_0.$$

In other words, $\int_{A_0} X \,\mathrm{d}\mathbb{P} = \int_{A_0} X_0 \,\mathrm{d}\mathbb{P}$ or $\mathbb{E}\big[X \mathbb{1}_A\big] = \mathbb{E}\big[X_0 \mathbb{1}_A\big]$ for all $A_0 \in \mathcal{F}_0$ (cf. Remark 1.27 choosing $\mathbb{Q} = \mu$, $\phi_0 = X_0$ and $Y_0 = \mathbb{1}_{A_0}$).

(ii) follows from (2.2) and the fact that the Radon-Nikodym density is unique up to $\mathbb{P}$- null sets. If $X \in \mathcal{L}^1$ this can also be seen directly as follows.

Let X_0 and $\tilde{X}_0$ be as in Definition 2.8. By Remark 2.9 we have $X_0, \tilde{X}_0 \in \mathcal{L}^1$. Then $A_0 := \{X_0 > \tilde{X}_0\} \in \mathcal{F}_0$ and

$$\mathbb{E}\big[X_0 \mathbb{1}_{A_0}\big] = \mathbb{E}\big[X \mathbb{1}_{A_0}\big] = \mathbb{E}\big[\tilde{X}_0 \mathbb{1}_{A_0}\big].$$

Thus,

$$\mathbb{E}\big[\underbrace{(X_0 - \tilde{X}_0)}_{>0 \text{ on } A_0} \mathbb{1}_{A_0}\big] = 0,$$

which implies $\mathbb{P}[A_0] = 0$. Similarly it can be shown that $\mathbb{P}[X_0 < \tilde{X}_0] = 0$. □

2.3. Properties of conditional expectations.

Proposition 2.11. *The conditional expectation has the following properties.*

(i) If $\mathcal{F}_0$ is $\mathbb{P}$-trivial, i.e. $\mathbb{P}[A] \in \{0, 1\}$ for all $A \in \mathcal{F}_0$, then $\mathbb{E}\big[X \,|\, \mathcal{F}_0\big] = \mathbb{E}[X]$ $\mathbb{P}$-a.s.
(ii) Linearity: $\mathbb{E}\big[aX + bY \,|\, \mathcal{F}_0\big] = a\,\mathbb{E}\big[X \,|\, \mathcal{F}_0\big] + b\,\mathbb{E}\big[Y \,|\, \mathcal{F}_0\big]$ $\mathbb{P}$-a.s.
(iii) Monotonicity: $X \leq Y$ $\mathbb{P}$-a.s. $\Rightarrow \mathbb{E}\big[X \,|\, \mathcal{F}_0\big] \leq \mathbb{E}\big[Y \,|\, \mathcal{F}_0\big]$ $\mathbb{P}$-a.s.
(iv) Monotone continuity: If $0 \leq X_1 \leq X_2 \leq \ldots$ $\mathbb{P}$-a.s., then

$$\mathbb{E}\big[\lim_{n\to\infty} X_n \,|\, \mathcal{F}_0\big] = \lim_{n\to\infty} \mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big] \quad \mathbb{P}\text{-a.s.}$$

(v) Fatou: If $0 \leq X_n$ $\mathbb{P}$-a.s. for all $n \in \mathbb{N}$, then

$$\mathbb{E}\big[\liminf_{n\to\infty} X_n \,|\, \mathcal{F}_0\big] \leq \liminf_{n\to\infty} \mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big] \quad \mathbb{P}\text{-a.s.}$$

(vi) Dominated convergence: If there exists $Y \in \mathcal{L}^1$ such that $|X_n| \le Y$ $\mathbb{P}$-a.s. for all $n \in \mathbb{N}$, then

$$\lim_{n\to\infty} X_n = X \quad \mathbb{P}\text{-a.s.} \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big] = \mathbb{E}\big[X \,|\, \mathcal{F}_0\big] \quad \mathbb{P}\text{-a.s.}$$

(vii) Jensen's inequality: Let $h : \mathbb{R} \to \mathbb{R}$ be convex, then

$$h\big(\mathbb{E}\big[X \,|\, \mathcal{F}_0\big]\big) \le \mathbb{E}\big[h(X) \,|\, \mathcal{F}_0\big] \quad \mathbb{P}\text{-a.s.}$$

Proof. (i) follows directly from the definition of the conditional expectation.

(ii) The right hand side is $\mathcal{F}_0$-measurable, and for any $A \in \mathcal{F}_0$, using the linearity of $\mathbb{E}$ we have

$$\begin{aligned}\mathbb{E}\Big[\mathbb{1}_A\Big(a\,\mathbb{E}\big[X \,|\, \mathcal{F}_0\big] + b\,\mathbb{E}\Big[X \,|\, \mathcal{F}_0\Big]\Big)\Big] &= a\,\mathbb{E}\big[\mathbb{1}_A\,\mathbb{E}\big[X \,|\, \mathcal{F}_0\big]\big] + b\,\mathbb{E}\big[\mathbb{1}_A\,\mathbb{E}\big[Y \,|\, \mathcal{F}_0\big]\big] \\ &= a\,\mathbb{E}\big[\mathbb{1}_A X\big] + b\,\mathbb{E}\big[\mathbb{1}_A Y\big] = \mathbb{E}\big[\mathbb{1}_A(aX + bY)\big].\end{aligned}$$

Hence, the right hand side fulfils the conditions in Definition 2.8.

(iii) We have $Y = X + Z$ for some random variable $Z \ge 0$. By linearity $\mathbb{E}\big[Y \,|\, \mathcal{F}_0\big] = \mathbb{E}\big[X \,|\, \mathcal{F}_0\big] + \mathbb{E}\big[Z \,|\, \mathcal{F}_0\big]$, and $\mathbb{E}\big[Z \,|\, \mathcal{F}_0\big]$ is easily seen to be non-negative from its definition.

(iv) By monotonicity $\mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big] \uparrow \lim_{n\to\infty} \mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big]$ which is $\mathcal{F}_0$-measurable. For any $A \in \mathcal{F}_0$, we use the monotone convergence theorem to obtain that

$$\begin{aligned}\mathbb{E}\Big[\mathbb{1}_A \lim_{n\to\infty} \mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big]\Big] &= \lim_{n\to\infty} \mathbb{E}\Big[\mathbb{1}_A\,\mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big]\Big] = \lim_{n\to\infty} \mathbb{E}\big[\mathbb{1}_A X_n\big] \\ &= \mathbb{E}\big[\mathbb{1}_A \lim_{n\to\infty} X_n\big].\end{aligned}$$

Hence, $\lim_{n\to\infty} \mathbb{E}\big[X_n \,|\, \mathcal{F}_0\big]$ fulfils the conditions in Definition 2.8.

Statements (v) and (vi) follow now exactly as the corresponding properties of the expected value. Jensen's inequality (vii) can be shown similarly as for the usual expected value. □

Proposition 2.12. *Let $Y_0 \ge 0$ be $\mathcal{F}_0$-measurable. Then,*

$$\mathbb{E}\big[Y_0\, X \,|\, \mathcal{F}_0\big] = Y_0\, \mathbb{E}\big[X \,|\, \mathcal{F}_0\big] \quad \mathbb{P}\text{-a.s.,} \tag{2.3}$$

so $\mathcal{F}_0$-measurable random variables behave like constants. In particular,

$$\mathbb{E}\big[Y_0 \,|\, \mathcal{F}_0\big] = Y_0 \quad \mathbb{P}\text{-a.s.}$$

Proof. Clearly the right hand side of (2.3) is $\mathcal{F}_0$-measurable, so we only need to check condition (ii) in Definition 2.8. Let us first consider the case $Y_0 = \mathbb{1}_{A_0}$ for any $A_0 \in \mathcal{F}_0$. Then for any $A \in \mathcal{F}_0$,

$$\mathbb{E}\big[Y_0\, X\, \mathbb{1}_A\big] = \mathbb{E}\big[X\, \mathbb{1}_{\underbrace{A\,\cap\, A_0}_{\in\mathcal{F}_0}}\big] = \mathbb{E}\big[\,\mathbb{E}[X \,|\, \mathcal{F}_0]\, \mathbb{1}_{A\cap A_0}\big] = \mathbb{E}\,\big[\big(Y_0\, \mathbb{E}[X \,|\, \mathcal{F}_0]\big)\, \mathbb{1}_A\big]\,.$$

For general Y_0 the statement follows by linearity and approximation. □

Proposition 2.13 ('Projectivity' or 'Tower property' of conditional expectations)**.** *Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$ be σ-algebras. Then,*

$$\mathbb{E}\big[X \,|\, \mathcal{F}_0\big] = \mathbb{E}\Big[\mathbb{E}\big[X \,|\, \mathcal{F}_1\big] \,\big|\, \mathcal{F}_0\Big] \quad \mathbb{P}\text{-}a.s.$$

Proof. Again, since conditional expectations are $\mathbb{P}$-a.s. unique, it suffices to identify the right hand side as the conditional expectation of X given $\mathcal{F}_0$ by verifying the (i) and (ii) in Definition 2.8. The right hand side is clearly $\mathcal{F}_0$ measurable since it is itself a conditional expectation given $\mathcal{F}_0$. To see (ii) let $A \in \mathcal{F}_0$. Then, clearly $A \in \mathcal{F}_1$ and therefore

$$\mathbb{E}\big[X \,\mathbb{1}_A\big] = \mathbb{E}\big[\mathbb{E}\left[X \,|\, \mathcal{F}_1\right] \mathbb{1}_A\big] = \mathbb{E}\Big[\mathbb{E}\left[\mathbb{E}\big[X \,|\, \mathcal{F}_1\big] \,\big|\, \mathcal{F}_0\right] \mathbb{1}_A\Big].$$

□

Proposition 2.14. *Let X be independent of $\mathcal{F}_0$[2]. Then,*

$$\mathbb{E}\big[X \,|\, \mathcal{F}_0\big] = \mathbb{E}[X] \quad \mathbb{P}\text{-}a.s.$$

Proof. $\mathbb{E}[X]$ is constant and therefore $\mathcal{F}_0$-measurable. For $A \in \mathcal{F}_0$ we have by independence and the linearity of the expected value that

$$\mathbb{E}\big[X \,\mathbb{1}_A\big] = \mathbb{E}[\mathbb{1}_A]\; \mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X]\, \mathbb{1}_A\big].$$

□

In practice, conditional expectations are difficult to compute explicitly. However, in two situations there are explicit formulas, namely in the discrete case discussed at the beginning, see (2.1), or when the random variables involved admit densities, which we now state without proof.

Proposition 2.15. *Let X and Y be real-valued random variables with densities f_X and f_Y. Assume that (X, Y) admits a joint density f_{XY}. Then the conditional distribution of X given Y is a random distribution with density*

$$f_{X|Y}(x) := \begin{cases} \frac{f_{XY}(x, Y(\omega))}{f_Y(Y(\omega))} & \text{if } f_Y(Y(\omega)) \neq 0, \\ 0 & \text{else}, \end{cases}$$

and the conditional expectation of X given Y is

$$\mathbb{E}\big[X \,|\, Y\big] = \int_{\mathbb{R}} x\, f_{X|Y}(x)\, \mathrm{d}x.$$

For later use we end this section with another useful result on conditional expectations.

[2] i.e. $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$ for all $A \in \sigma(X)$ and all $B \in \mathcal{F}_0$. If for instance $\mathcal{F}_0 = \sigma(Y)$ this means that X and Y are independent random variables.

Proposition 2.16. *Let $F : \mathbb{R}^2 \to [0, \infty)$ be measurable, X be independent of $\mathcal{F}_0$ and Y be $\mathcal{F}_0$-measurable. Then*

$$\mathbb{E}\big[F(X,Y)\,|\,\mathcal{F}_0\big](\omega) = \mathbb{E}\big[F(X, Y(\omega))\big] \quad \mathbb{P}\text{-a.s.}$$

More precisely, if we set $\Phi(y) := \mathbb{E}\big[F(X,y)\big]$, $y \in \mathbb{R}$, then

$$\mathbb{E}\big[F(X,Y)\,|\,\mathcal{F}_0\big](\omega) = \Phi(Y(\omega)) \quad \mathbb{P}\text{-a.s.}$$

Proof. Let first F be of the form $F(x,y) = f(x)g(y)$ for any measurable $f, g : \mathbb{R} \to [0, \infty)$. Then,

$$\begin{aligned}\mathbb{E}\big[F(X,Y)\,|\,\mathcal{F}_0\big](\omega) &= g(Y(\omega))\, \mathbb{E}\big[f(X)\,|\,\mathcal{F}_0\big](\omega) = g(Y(\omega))\, \mathbb{E}\big[f(X)\big] \\ &= \mathbb{E}\big[g(Y(\omega))\, f(X)\big] = \Phi(Y(\omega)).\end{aligned}$$

For general F the statement now follows from a monotone class argument. □

3. Martingales in discrete time

3.1. **Definition and examples.** In this section we introduce the fundamental concept of martingales, which will keep playing a central role in our investigation of models for financial markets. Martingales are "truly random" stochastic processes, in the sense that their observation in the past does not allow for useful prediction of the future. By useful we mean here that no gambling strategies can be devised that would allow for systematic gains.

By a stochastic process (in discrete time) we mean a sequence of random variables $(X_n)_{n\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We now equip this probability space with a filtration.

Definition 3.1. A *filtration* $(\mathcal{F}_n)_{n\geq 0}$ is an increasing family of σ-algebras, that is $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}$ for all $n \geq 0$.

Set

$$\mathcal{F}_\infty := \sigma(\mathcal{F}_n : n \geq 0).$$

Then $\mathcal{F}_\infty \subseteq \mathcal{F}$. We allow the possibility that $\mathcal{F}_\infty \neq \mathcal{F}$. In almost all situations the index n represents time. Then the σ-algebra $\mathcal{F}_n$ contains all the events that are observable up to time n, so $\mathcal{F}_n$ models the information available at time n. A stochastic process $X = (X_n)_{n\geq 0}$ naturally induces a filtration $(\mathcal{F}_n^X)_{n\geq 0}$ defined via $\mathcal{F}_n^X := \sigma(X_0, \ldots X_n)$. For example, if X models the price process of a risky asset, say a share, then $\mathcal{F}_n^X$ represents the information about all prices up to time n.

Example 3.2. Consider the simple symmetric random walk X on $\mathbb{Z}$ started from $X_0 := 0$, that is $X_n = \sum_{k=1}^n Z_k$, $n \geq 1$, where $(Z_k)_{k\geq 1}$ are i.i.d. random variable with $\mathbb{P}[Z_k = 1] = \mathbb{P}[Z_k = -1] = \frac{1}{2}$. Then, $\mathcal{F}_n^X = \sigma(X_1, \ldots X_n) = \sigma(Z_1, \ldots Z_n)$, $n \geq 1$, defines a filtration and

$$\big\{X_1 \leq 0,\, X_3 \geq 2\big\} \in \mathcal{F}_3^X \quad \text{but} \quad \big\{X_4 > 0\big\} \notin \mathcal{F}_3^X.$$

Since $X_0 = 0$ is deterministic, $\mathcal{F}_n^X = \sigma(X_0, \ldots X_n)$ and we could have started the filtration with the trivial σ-algebra $\mathcal{F}_0 = \sigma(X_0) = \{\emptyset, \Omega\}$.

Definition 3.3. A stochastic process $X = (X_n)_{n\geq 0}$ is said to be *adapted* to a filtration $(\mathcal{F}_n)_{n\geq 0}$ if X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$.

Note that every stochastic process $(X_n)_{n\geq 0}$ is adapted to the filtration if $\mathcal{F}_n^X \subseteq \mathcal{F}_n$ for all n. In particular, $(X_n)_{n\geq 0}$ is always adapted to $(\mathcal{F}_n^X)_{n\geq 0}$. Now we define martingales.

Definition 3.4. Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. A stochastic process $X = (X_n)_{n\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ is a *martingale* (or $\mathbb{P}$-martingale) if and only if the following hold.

(a) X is adapted, that is X_n is $\mathcal{F}_n$-measurable for every n.
(b) X is integrable, i.e. $\mathbb{E}[|X_n|] < \infty$ for every n.
(c) The *martingale property* holds, i.e. for all $n \geq 0$,

$$\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = X_n, \qquad \mathbb{P}\text{-a.s.}$$

If (a) and (b) hold, but instead of (c) we have that $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] \geq X_n$ (resp. $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] \leq X_n$), then the process X is called a *submartingale* (resp. a *supermartingale*).

Remark 3.5. (i) If X is a martingale. then $\mathbb{E}[X_m] = \mathbb{E}[X_n]$ for all $0 \leq m \leq n$, for a submartingale we have $\mathbb{E}[X_m] \leq \mathbb{E}[X_n]$, finally, for a supermartingale $\mathbb{E}[X_m] \geq \mathbb{E}[X_n]$.

(ii) The martingale property (c) is equivalent to

$$\mathbb{E}[X_n - X_m \,|\, \mathcal{F}_m] = 0, \qquad \mathbb{P}\text{-a.s.}, \quad \forall 0 \leq m \leq n,$$

so a martingale is a mathematical model for a fair game in the sense that based on the information available at time m the expected future profit is zero.

(iii) A martingale $X = (X_n)_{n=0,\dots,N}$ with a finite time range $\{0, \dots, N\}$ is determined by X_N via $X_n = \mathbb{E}[X_N \,|\, \mathcal{F}_n]$. Conversely, every $F \in \mathcal{L}^1(\Omega, \mathcal{F}_N, \mathbb{P})$ defines a martingale via

$$X_n := \mathbb{E}[F \,|\, \mathcal{F}_n], \qquad n = 0, \dots, N.$$

Example 3.6. Let $Z_1, Z_2 \dots,$ be independent random variables with $Z_k \in \mathcal{L}^1$ and $\mathbb{E}[Z_k] = 0$ for all $k \geq 1$. Set

$$X_0 := 0, \qquad X_n := \sum_{k=1}^{n} Z_k, \quad n \geq 1$$

and $\mathcal{F}_n = \mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$, $n \geq 0$. Obviously, X is adapted to $(\mathcal{F}_n)_{n\geq 0}$ and $X_n \in \mathcal{L}^1$ for all $n \geq 0$ since $\mathbb{E}[|Z_k|] < \infty$ for all k and therefore

$$\mathbb{E}[|X_n|] \leq \sum_{k=1}^{n} \mathbb{E}[|Z_k|] < \infty.$$

Further, for $n \geq 0$,

$$\mathbb{E}\big[X_{n+1} \,|\, \mathcal{F}_n\big] = \mathbb{E}\big[X_n + Z_{n+1} \,|\, \mathcal{F}_n\big] = X_n + \mathbb{E}\big[Z_{n+1} \,|\, \mathcal{F}_n\big] = X_n + \mathbb{E}\big[Z_{n+1}\big] = X_n,$$

where we used in the third step that Z_{n+1} is independent of $\mathcal{F}_n$. So X is a martingale. Note that in the special case $Z_k \in \{-1, 1\}$ with $\mathbb{P}[Z_k = 1] = \mathbb{P}[Z_k = -1] = \frac{1}{2}$ the process X becomes the simple random walk on $\mathbb{Z}$.

Definition 3.7. A stochastic process $(C_n)_{n\geq 1}$ is called *previsible*[3] (or *predictable*) with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$, if C_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$.

Proposition 3.8. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ be a filtered probability space.*

(i) Let $X = (X_n)_{n\geq 0}$ be a martingale and let $(C_n)_{n\geq 1}$ be a bounded previsible process. Then, the process $Y = (Y_n)_{n\geq 0}$ defined by

$$Y_n := \sum_{k=1}^{n} C_k \left(X_k - X_{k-1}\right), \qquad Y_0 := 0,$$

is a martingale.

(ii) If X is a submartingale (resp. supermartingale) and $(C_n)_{n\geq 1}$ is a bounded previsible and non-negative, then Y is a submartingale (resp. supermartingale).

Proof. (i) Since $(C_n)_{n\geq 1}$ is bounded, i.e. there exists $c > 0$ such that $|C_n| \leq c$ $\mathbb{P}$-a.s. for all n, we have by triangle inequality and integrability of the martingale X that

$$\mathbb{E}\left[|Y_n|\right] \;\leq\; \sum_{k=1}^{n} \mathbb{E}\left[|C_k|\,|X_k - X_{k-1}|\right] \;\leq\; c\sum_{k=1}^{n} \left(\mathbb{E}[|X_k|] + \mathbb{E}[|X_{k-1}|]\right) \;<\; \infty,$$

so $Y_n \in \mathcal{L}^1$ for all n. For all $k \leq n$ the random variables C_k, X_{k-1} and X_k are all $\mathcal{F}_n$-measurable, so Y_n is $\mathcal{F}_n$-measurable, which means that Y is adapted. Finally, for $n \geq 1$,

$$\begin{aligned} \mathbb{E}\big[Y_n - Y_{n-1} \,|\, \mathcal{F}_{n-1}\big] &= \mathbb{E}\big[C_n(X_n - X_{n-1}) \,|\, \mathcal{F}_{n-1}\big] \\ &= C_n\, \mathbb{E}\big[X_n - X_{n-1} \,|\, \mathcal{F}_{n-1}\big] = 0. \end{aligned} \tag{3.1}$$

Here we used that C_n is $\mathcal{F}_{n-1}$-measurable in the second step and the martingale property in the last step.

(ii) Since C_n is now assumed to be non-negative, we have in the last step of (3.1) that $C_n\, \mathbb{E}\big[X_n - X_{n-1} \,|\, \mathcal{F}_{n-1}\big]$ is non-negative if X is a submartingale and non-positive if M is a supermartingale. □

Remark 3.9. (i) Sometimes the process $(C_n)_{n\geq 1}$ represents a gambling strategy. If X models the price process of a share, then Y_n represents the wealth at time n.

(ii) Y is a discrete time version of the stochastic integral '$\int C \,\mathrm{d}M$'.

3.2. **Martingale convergence.** Let $X = (X_n)_{n\geq 0}$ be a real-valued stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ adapted to a filtration $(\mathcal{F}_n)_{n\geq 0}$. Consider an interval $[a, b]$. We want to count the number of times a process crosses this interval from below.

[3] The teminology previsible refers to the fact that C_n can be foreseen from the information available at time $n - 1$

Definition 3.10. Let $a < b \in \mathbb{R}$. We say that an *upcrossing* of $[a, b]$ occurs between times m and n, if

(i) $X_m < a$, $X_n > b$,
(ii) for all k such that $m < k < n$, $X_k \in [a, b]$.

We denote by $U_N(X, [a, b])$ the number of uprossings in the time interval $[0, N]$. Now we consider the previsible process $(C_n)_{n\geq 1}$ defined by

$$C_1 := \mathbb{1}_{\{X_0<a\}}, \qquad C_n := \mathbb{1}_{\{C_{n-1}=1\}}\mathbb{1}_{\{X_{n-1}\leq b\}} + \mathbb{1}_{\{C_{n-1}=0\}}\mathbb{1}_{\{X_{n-1}<a\}}, \quad n \geq 2. \tag{3.2}$$

This process represents a winning strategy: wait until the process (say, price of a share) drops below a. Buy the stock, and hold it until its price exceeds b; sell, wait until the price drops below a, and so on. The associated wealth process is given by

$$W_n = \sum_{k=1}^{n} C_k \left(X_k - X_{k-1}\right), \qquad W_0 := 0.$$

Now each time there is an upcrossing of $[a, b]$ we win at least $(b - a)$. Thus, at time N, we have

$$W_N \geq (b - a)\, U_N(X, [a, b]) - |a - X_N|\, \mathbb{1}_{\{X_N<a\}}, \tag{3.3}$$

where the last term count is the maximum loss that we could have incurred if we are invested at time N and the price is below a.

Naive intuition would suggest that in the long run, the first term must win. The next theorem says that this is false, if we are in a fair or disadvantageous game.

Theorem 3.11 (Doob's upcrossing lemma)**.** *Let X be a supermartingale. Then for any $a < b \in \mathbb{R}$,*

$$\mathbb{E}\left[U_N(X, [a, b])\right] \leq \frac{\mathbb{E}\left[\left(X_N - a\right)^-\right]}{b - a}. \tag{3.4}$$

Proof. The process $(C_n)_{n\geq 1}$ defined in (3.2) is obviously bounded, non-negative and previsible, so by Proposition 3.8 (ii) the wealth process $(W_n)_{n\geq 0}$ is a supermartingale with $W_0 = 0$. Therefore $\mathbb{E}[W_N] \leq 0$ and taking expectation in (3.3) gives (3.4). □

For any interval $[a, b]$, we define the monotone limit

$$U_\infty(X, [a, b]) := \lim_{N\to\infty} U_N(X, [a, b]).$$

Corollary 3.12. *Let $(X_n)_{n\geq 0}$ be an $\mathcal{L}^1$-bounded supermartingale, i.e.* $\sup_n \mathbb{E}[|X_n|] < \infty$*. Then*

$$\mathbb{E}\left[U_\infty(X, [a, b])\right] \leq \frac{|a| + \sup_n \mathbb{E}\left[|X_n|\right]}{b - a} < \infty. \tag{3.5}$$

In particular, $\mathbb{P}\left[U_\infty(X, [a, b]) = \infty\right] = 0.$

Note that the requirement $\sup_n \mathbb{E}[|X_n|] < \infty$ is strictly stronger than just asking that for all n, $\mathbb{E}[|X_n|] < \infty$.

Proof. This follows directly from Theorem 3.11 and the monotone convergence theorem since $\sup_n \mathbb{E}\left[(X_n - a)^-\right] \leq |a| + \sup_n \mathbb{E}\big[|X_n|\big]$. ☐

This is quite impressive: a (super-)martingale that is $\mathcal{L}^1$-bounded cannot cross any interval infinitely often. The next result is even more striking, and in fact one of the most important results about martingales.

Theorem 3.13 (Doob's supermartingale convergence theorem). *Let $(X_n)_{n\geq 0}$ be an $\mathcal{L}^1$-bounded supermartingale. Then there exists an integrable $\mathcal{F}_\infty$-measurable random variable X_∞ such that, $\mathbb{P}$-a.s., $X_n \to X_\infty$ as $n \to \infty$.*

Proof. Define

$$\begin{aligned}
\Lambda &:= \big\{\omega : X_n(\omega) \text{ does not converge to a limit in } [-\infty,\infty]\big\} \\
&= \big\{\omega : \limsup_n X_n(\omega) > \liminf_n X_n(\omega)\big\} \\
&= \bigcup_{a,b\in\mathbb{Q}:a<b} \{\omega : \limsup_n X_n(\omega) > b > a > \liminf_n X_n(\omega)\big\} =: \bigcup_{a,b\in\mathbb{Q}:a<b} \Lambda_{a,b}.
\end{aligned}$$

But

$$\Lambda_{a,b} \subset \big\{\omega : U_\infty(X,[a,b])(\omega) = \infty\big\}.$$

Therefore, by Corollary 3.12, $\mathbb{P}[\Lambda_{a,b}] = 0$, and thus also

$$\mathbb{P}\Big[\bigcup_{a,b\in\mathbb{Q}:a<b} \Lambda_{a,b}\Big] = 0,$$

since countable unions of null-sets are null-sets. Thus $\mathbb{P}[\Lambda] = 0$ and the limit $X_\infty := \lim_n X_n$ exists in $[-\infty,\infty]$ with probability one and is $\mathcal{F}_\infty$-measurable. It remains to show that it is integrable. To do this, we use Fatous lemma:

$$\mathbb{E}\big[|X_\infty|\big] = \mathbb{E}\big[\liminf_n |X_n|\big] \leq \liminf_n \mathbb{E}\big[|X_n|\big] \leq \sup_n \mathbb{E}\big[|X_n|\big] < \infty.$$

So X_∞ is integrable. ☐

Remark 3.14. Doobs convergence theorem implies that positive supermartingale always converge a.s. This is because the supermartingale property ensures in this case that $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \leq \mathbb{E}[X_0]$, so the uniform boundedness in $\mathcal{L}^1$ is always guaranteed.

3.3. **Stopping times and optional stopping.** In a stochastic process we often want to consider random times that are determined by the occurrence of a particular event. If this event depends only on what happens 'in the past', we call it a *stopping time*. Stopping times are nice, since we can determine their occurrence as we observe the process; so if we are only interested in them, we can stop the process at this moment, hence the name.

Definition 3.15. A map $\tau : \Omega \to \{0, 1, \ldots\} \cup \{\infty\}$ is called a *stopping time* (with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$) if $\{\tau \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$ or, equivalently, $\{\tau = n\} \in \mathcal{F}_n$ for all $n \geq 0$.

Example 3.16. The most important examples of stopping times are hitting times. Let $(X_n)_{n\geq 0}$ be an adapted process, and let $B \in \mathcal{B}(\mathbb{R})$. Define

$$\tau_B(\omega) := \inf\big\{n > 0 : X_n(\omega) \in B\big\}$$

with $\inf \emptyset := +\infty$. Then τ_B is a stopping time.

Definition 3.17. Let $(X_n)_{n\geq 0}$ be a stochastic process and τ be a stopping time. We define the *stopped process* X^τ via

$$X_n^\tau(\omega) := X_{n\wedge\tau(\omega)}(\omega).$$

Proposition 3.18. *Let $(X_n)_{n\geq 0}$ be a (sub-)martingale and τ be a stopping time. Then the stopped process X^τ is a (sub-)martingale.*

Proof. Exercise! □

Theorem 3.19 (Doob's Optional stopping theorem)**.** *Let $(X_n)_{n\geq 0}$ be a martingale and τ be a stopping time. Then, $X_\tau \in \mathcal{L}^1$ and*

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0],$$

if one of the following conditions holds.

(a) τ is a.s. bounded (i.e. there exists $N \in \mathbb{N}$ such that $\tau(\omega) \leq N$ for $\mathbb{P}$-a.e. $\omega \in \Omega$).
(b) X^τ is bounded and τ is a.s. finite.
(c) $\mathbb{E}[\tau] < \infty$ and for some $K < \infty$,

$$\big|X_n(\omega) - X_{n-1}(\omega)\big| \leq K, \qquad \forall n \in \mathbb{N},\ \omega \in \Omega.$$

Proof. By Proposition 3.18 the stopped process $X_n^\tau = X_{\tau\wedge n}$ is a martingale. In particular, its expected value is constant in n, so that

$$\mathbb{E}\big[X_{\tau\wedge n}\big] = \mathbb{E}\big[X_n^\tau\big] = \mathbb{E}\big[X_0^\tau\big] = \mathbb{E}\big[X_0\big]. \tag{3.6}$$

Consider the case (a). By assumption τ is a.s. bounded, so there exists $N \in \mathbb{N}$ such that $\tau(\omega) \leq N$ for $\mathbb{P}$-a.e. $\omega \in \Omega$. Then, choosing $n = N$ in (3.6) on the event $\{\tau \leq N\}$ gives the claim.

In case (b) we have $\lim_n X_{\tau\wedge n} = X_\tau$ on the event $\{\tau < \infty\}$ and therefore $\mathbb{P}$-a.s. Since X^τ is bounded, we get by the dominated convergence theorem

$$\lim_{n\to\infty} \mathbb{E}\big[X_{\tau\wedge n}\big] = \mathbb{E}\big[\lim_{n\to\infty} X_{\tau\wedge n}\big] = \mathbb{E}\big[X_\tau\big],$$

which together with (3.6) implies the result.

In the last case, (c), we observe that

$$\big|X_{\tau\wedge n} - X_0\big| = \Big|\sum_{k=1}^{\tau\wedge n}(X_k - X_{k-1})\Big| \leq K\tau,$$

and by assumption $\mathbb{E}[K\tau] < \infty$. So again by the dominated convergence theorem we can pass to the limit in (3.6). □

Remark 3.20. (i) By similar arguments Theorem 3.19 extends immediately to super- resp. submartingales in which case the conclusion reads

$$\mathbb{E}[X_\tau] \leq \mathbb{E}[X_0] \quad \text{resp.} \quad \mathbb{E}[X_\tau] \geq \mathbb{E}[X_0].$$

(ii) Theorem 3.19 may look strange and contradict the 'no strategy' idea. Take a simple random walk $(S_n)_{n\geq 0}$ on $\mathbb{Z}$ (i.e. a series of fair games), and define a stopping time $\tau = \inf\{n : S_n = 10\}$. Then clearly $\mathbb{E}[S_\tau] = 10 \neq \mathbb{E}[S_0] = 0$! So we conclude, using (c), that $\mathbb{E}[\tau] = +\infty$. In fact, the 'sure' gain if we achieve our goal is offset by the fact that on average, it takes infinitely long to reach it (of course, most games will end quickly, but chances are that some may take very very long!).

(iii) Condition (b) in Theorem 3.19 can be relaxed. In fact, it suffices that X^τ is bounded by an integrable random variable, i.e. $|X_{\tau\wedge n}| \leq Y$ a.s. for some $Y \in \mathcal{L}^1$. The proof remains unchanged as the dominated convergence theorem still applies in this situation.

The following variant of the optional stopping theorem is also often used.

Theorem 3.21 (Hunt's optional stopping theorem)**.** *Let X be a supermartingale and let σ and τ be bounded stopping times with $\sigma \leq \tau$. Then $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_\sigma]$.*

Proof. Fix $n \geq 0$ such that $\tau \leq n$. Then

$$X_\tau = X_\sigma + \sum_{\sigma \leq k < \tau} (X_{k+1} - X_k) = X_\sigma + \sum_{k=0}^{n} (X_{k+1} - X_k) \mathbb{1}_{\{\sigma \leq k < \tau\}}. \tag{3.7}$$

Now $\{\sigma \leq k\} \in \mathcal{F}_k$ and $\{\tau > k\} = \{\tau \leq k\}^c \in \mathcal{F}_k$, so that $\{\sigma \leq k < \tau\} \in \mathcal{F}_k$, and by the supermartingale property of X,

$$\begin{aligned}\mathbb{E}\big[(X_{k+1} - X_k)\mathbb{1}_{\{\sigma\leq k<\tau\}}\big] &= \mathbb{E}\Big[\mathbb{E}\big[(X_{k+1} - X_k)\mathbb{1}_{\{\sigma\leq k<\tau\}} \,\big|\, \mathcal{F}_k\big]\Big] \\ &\mathbb{E}\Big[\mathbb{E}\big[X_{k+1} - X_k \,\big|\, \mathcal{F}_k\big]\mathbb{1}_{\{\sigma\leq k<\tau\}}\Big] \leq 0.\end{aligned}$$

Hence, on taking expectations in (3.7), we obtain $\mathbb{E}[X_\tau] \leq \mathbb{E}[X_\sigma]$. □

Again, note that X is a submartingale if and only if $-X$ is a supermartingale, and X is a martingale if and only both X and $-X$ are supermartingales. So the optional stopping theorem immediately implies a submartingale version with $\mathbb{E}[X_\tau] \geq \mathbb{E}[X_\sigma]$ and a martingale version with $\mathbb{E}[X_\tau] = \mathbb{E}[X_0] = \mathbb{E}[X_\sigma]$.

3.4. **Doob's maximal inequalities.** Define, for a stochastic process $X = (X_n)_{n\geq 0}$,

$$X_n^* := \sup_{k\leq n} |X_k|.$$

In the next two theorems, we see that the martingale (or submartingale) property allows us to obtain estimates on this supremum in terms of expectations for X_n itself.

Theorem 3.22 (Doob's maximal inequality)**.** *Let X be a martingale or non-negative submartingale. Then, for all $\lambda \geq 0$,*

$$\lambda\,\mathbb{P}[X_n^* \geq \lambda] \leq \mathbb{E}\big[|X_n|\mathbb{1}_{\{X_n^* \geq \lambda\}}\big] \leq \mathbb{E}[|X_n|].$$

Proof. If X is a martingale, then $|X|$ is a non-negative submartingale. It therefore suffices to consider the case where X is non-negative. Set

$$\tau = \inf\{k \geq 0 : X_k \geq \lambda\} \wedge n.$$

Then τ is a bounded stopping time as $\tau \leq n$. So, by optional stopping,

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_\tau] = \mathbb{E}\left[X_\tau \mathbb{1}_{\{X_n^* \geq \lambda\}}\right] + \mathbb{E}\big[X_\tau \mathbb{1}_{\{X_n^* < \lambda\}}\big] \geq \lambda\,\mathbb{P}[X_n^* \geq \lambda] + \mathbb{E}\big[X_n \mathbb{1}_{\{X_n^* < \lambda\}}\big].$$

Hence

$$\lambda\,\mathbb{P}[X_n^* \geq \lambda] \leq \mathbb{E}[X_n \mathbb{1}_{\{X_n^* \geq \lambda\}}] \leq \mathbb{E}[X_n].$$

□

Theorem 3.23 (Doob's L^p-inequality)**.** *Let X be a martingale or non-negative submartingale. Then, for all $p > 1$,*

$$\mathbb{E}\big[(X_n^*)^p\big] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_n|^p\big].$$

Proof. If X is a martingale, then $|X|$ is a non-negative submartingale. So it suffices to consider the case where X is non-negative. Fix $K \in (0,\infty)$. By Fubini's theorem, Doob's maximal inequality, and Hölder's inequality,

$$\begin{aligned}
\mathbb{E}\big[(X_n^* \wedge K)^p\big] &= \mathbb{E}\int_0^K p\lambda^{p-1}\mathbb{1}_{\{X_n^* \geq \lambda\}}\,d\lambda = \int_0^K p\lambda^{p-1}\,\mathbb{P}[X_n^* \geq \lambda]\,d\lambda \\
&\leq \int_0^K p\lambda^{p-2}\,\mathbb{E}\big[X_n \mathbb{1}_{\{X_n^* \geq \lambda\}}\big]\,d\lambda = \frac{p}{p-1}\,\mathbb{E}\big[X_n (X_n^* \wedge K)^{p-1}\big] \\
&\leq \frac{p}{p-1}\,\mathbb{E}\big[X_n^p\big]^{1/p}\,\mathbb{E}\big[(X_n^* \wedge K)^p\big]^{(p-1)/p}.
\end{aligned}$$

We divide both sides by $\mathbb{E}\big[(X_n^* \wedge K)^p\big]^{(p-1)/p}$ and obtain that $\mathbb{E}\big[\big(X_n^* \wedge K\big)^p\big] \leq \big(\frac{p}{p-1}\big)^p\,\mathbb{E}\big[X_n^p\big]$. The result follows by monotone convergence on letting $K \to \infty$. □

Doob's maximal and L^p inequalities have versions which apply, under the same hypotheses, to the full supremum

$$X^* = \sup_{n \geq 0}|X_n|.$$

Since $X_n^* \uparrow X^*$, on letting $n \to \infty$, we obtain, for all $\lambda \geq 0$,

$$\lambda\,\mathbb{P}[X^* > \lambda] = \lim_{n\to\infty}\lambda\,\mathbb{P}[X_n^* > \lambda] \leq \sup_{n\geq 0}\mathbb{E}[|X_n|].$$

We can then replace $\lambda\,\mathbb{P}[X^* > \lambda]$ by $\lambda\,\mathbb{P}[X^* \geq \lambda]$ by taking limits from the right in λ. Similarly, for $p \in (1,\infty)$, by monotone convergence,

$$\mathbb{E}\big[(X^*)^p\big] \leq \Big(\frac{p}{p-1}\Big)^p \sup_{n\geq 0}\mathbb{E}\big[|X_n|^p\big].$$

3.5. **Doob decomposition.** One of the games when dealing with stochastic processes is to "extract the martingale part". There are several such decompositions, but the following Doob decomposition is very important and its continuous time analogue will be fundamental for the theory of stochastic integration.

Theorem 3.24 (Doob decomposition). *Let $X = (X_n)_{n\geq 0}$ be an adapted process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ such that $\mathbb{E}[|X_n|] < \infty$ for every n.*

(i) The process X admits a $\mathbb{P}$-a.s. unique decomposition

$$X_n = X_0 + M_n + A_n, \qquad n \geq 0,$$

where $M = (M_n)_{n\geq 0}$ is a martingale with $M_0 = 0$ and $A = (A_n)_{n\geq 0}$ is a previsible process with $A_0 = 0$.

(ii) The process X is a submartingale, if and only if A is an increasing process in the sense that, $\mathbb{P}$-a.s., $A_n \leq A_{n+1}$ for all $n \geq 0$.

Proof. (i) We first show existence. Define

$$A_0 := 0, \qquad A_{n+1} - A_n = \mathbb{E}\big[X_{n+1} - X_n \,|\, \mathcal{F}_n\big], \qquad n \geq 0.$$

Then, A is previsible and $M_n := X_n - X_0 - A_n$ defines a martingale. Indeed, M is clearly adapted and integrable, and the martingale property holds since

$$\begin{aligned} \mathbb{E}\big[M_{n+1} - M_n \,|\, \mathcal{F}_n\big] &= \mathbb{E}\big[(X_{n+1} - X_n) - (A_{n+1} - A_n) \,|\, \mathcal{F}_n\big] \\ &= \mathbb{E}\big[X_{n+1} - X_n \,|\, \mathcal{F}_n\big] - (A_{n+1} - A_n) = 0, \end{aligned} \tag{3.8}$$

by the definition of A.

To see uniqueness, suppose we have another such decomposition $X_n = X_0 + M'_n + A'_n$. Then, the difference $A'_n - A_n = M_n - M'_n$ is a martingale and $A'_n - A_n$ is $\mathcal{F}_{n-1}$-measurable for $n \geq 1$. In particular,

$$A'_n - A_n = \mathbb{E}\big[A'_n - A_n \,|\, \mathcal{F}_{n-1}\big] = A'_{n-1} - A_{n-1}.$$

By induction it follows that $A'_n - A_n = A_0 - A'_0 = 0$. This ends the proof of (i).

The assertion of (ii) is obvious from (3.8). □

The Doob decomposition gives rise to an important derived process associated with a martingale X, namely the bracket $\langle X\rangle$. More precisely, let X be a martingale in $\mathcal{L}^2$ with $X_0 = 0$. Then X^2 is a submartingale with Doob decomposition $X^2 = M + \langle X\rangle$, where M is a martingale that vanishes at zero and $\langle X\rangle$ is a previsible process that vanishes at zero. The process $\langle X\rangle$ called the *bracket* of X. We will come back to this point later when we discuss martingales in continuous time.

4. Arbitrage theory in discrete time

In this section we will give some answers to our two main questions in the context of a multiperiod model in discrete time, that is we will develop a formula for prices of financial derivative and give a characterisation of arbitrage-free market models. For the latter we will discuss the so-called 'Fundamental Theorem of Asset Prices', which states that a model is arbitrage-free if and only if the process discounted asset prices is a martingale under some measure admitting the same null sets as the original measure. As a warm-up we will discuss these results for a one-period model first.

In the following, $x \cdot y = x^T y = \sum_{i=1}^d x_i y_i$ with $x, y \in \mathbb{R}^d$ denotes the canonical scalar product in $\mathbb{R}^d$.

4.1. **Single period model.** Consider a single period market model with d risky assets and one riskless bond. Their values at time t are denoted by $\bar{S}_t = (S_t^0, S_t) = (S_t^0, S_t^1, \ldots, S_t^d)$, $t \in \{0, 1\}$. The prices $\bar{S}_0$ at time $t = 0$ are deterministic. The bond S^0 has the value $S_0^0 = 1$ at time $t = 0$ and the value $S_1^0 = 1 + r$, $r \geq 0$, at time $t = 1$. The values of the risky assets at time $t = 1$ are represented by a vector of random variables $S_1 = (S_1^1, \ldots, S_1^d)$ on $(\Omega, \mathcal{F}, \mathbb{P})$.

At $t = 0$, an investor chooses a portfolio

$$\bar{\theta} = (\theta^0, \theta) = (\theta^0, \ldots, \theta^d) \in \mathbb{R}^{d+1},$$

where θ^i represents the number of units of asset i. We allow the components θ^i to be negative. If $\theta^0 < 0$, this corresponds to a loan, and if $\theta^i < 0$ for $i \geq 1$, a quantity of $|\theta^i|$ units of asset i is sold without owning them (short sale). At time $t = 0$ the price to buy the portfolio equals $\bar{\theta} \cdot \bar{S}_0 = \sum_{i=0}^d \theta^i S_0^i$, and at time $t = 1$ the portfolio will have the value $\bar{\theta} \cdot \bar{S}_1(\omega) = \sum_{i=0}^d \theta^i S_1^i(\omega)$.

Definition 4.1. We say that a portfolio $\bar{\theta} = (\theta^0, \theta) \in \mathbb{R}^{d+1}$ is an *arbitrage opportunity* if $\bar{\theta} \cdot \bar{S}_0 \leq 0$ but $\bar{\theta} \cdot \bar{S}_1 \geq 0$ $\mathbb{P}$-a.s. and $\mathbb{P}[\bar{\theta} \cdot \bar{S}_1 > 0] > 0$.

Intuitively, an arbitrage opportunity is an investment strategy that yields with positive probability a positive profit and is not exposed to any downside risk. The existence of such an arbitrage opportunity may be regarded as a market inefficiency in the sense that certain assets are not priced in a reasonable way. In real-world markets, arbitrage opportunities are rather hard to find. If such an opportunity would show up, it would generate a large demand, prices would adjust, and the opportunity would disappear.

Remark 4.2. The probability measure $\mathbb{P}$ enters the definition of an arbitrage only through the null sets of $\mathbb{P}$. Thus, if $\bar{\theta}$ is an arbitrage under $\mathbb{P}$ then it is also an arbitrage under any probability measure $\mathbb{Q} \approx \mathbb{P}$.

The following lemma shows that in an arbitrage-free market any investment in risky assets must be open to some downside risk if it yields with positive probability a better result than investing the same amount in the risk-free asset.

Lemma 4.3. *The following are equivalent.*

(i) The market model admits an arbitrage opportunity.

(ii) There exists $\theta \in \mathbb{R}^d$ such that

$$\theta \cdot S_1 \geq (1+r)\, \theta \cdot S_0 \ \mathbb{P}\text{-a.s.} \quad \text{and} \quad \mathbb{P}\left[\theta \cdot S_1 > (1+r)\, \theta \cdot S_0\right] > 0.$$

Proof. (i) $\Rightarrow$ (ii): Let $\bar{\theta} \in \mathbb{R}^{d+1}$ be an arbitrage. Then, $\theta^0 + \theta \cdot S_0 = \bar{\theta} \cdot \bar{S}_0 \leq 0$, i.e. $\theta^0 \leq -\theta \cdot S_0$. Therefore,

$$\theta \cdot S_1 - (1+r)\theta \cdot S_0 \geq \theta \cdot S_1 + (1+r)\theta^0 = \bar{\theta} \cdot \bar{S}_1.$$

Since $\bar{\theta} \cdot \bar{S}_1$ is $\mathbb{P}$-a.s. non-negative and strictly positive with positive probability, the same is true for $\theta \cdot S_1 - (1+r)\theta \cdot S_0$.

(ii) $\Rightarrow$ (i): Let θ be as in (ii) and set $\theta^0 := -\theta \cdot S_0$ so that $\bar{\theta} \cdot \bar{S}_0 = 0$. Then,

$$\bar{\theta} \cdot \bar{S}_1 = -(1+r)\theta \cdot S_0 + \theta \cdot S_1,$$

which is $\mathbb{P}$-a.s. non-negative and strictly positive with positive probability. Hence, $\bar{\theta}$ is an arbitrage. □

Theorem 4.4 (Fundamental Theorem of Asset Pricing (FTAP))**.** *The following are equivalent.*

(i) There is no arbitrage.
(ii) There exists a probability measure $\mathbb{Q} \approx \mathbb{P}$ such that $\mathbb{E}_{\mathbb{Q}}[|S_1|] < \infty$,

$$\mathbb{E}_{\mathbb{Q}}\left[\frac{\bar{S}_1}{1+r}\right] = \bar{S}_0, \tag{4.1}$$

and the density $d\mathbb{Q}/d\mathbb{P}$ is bounded.

The probability measure $\mathbb{Q}$ is referred to as a risk-neutral measure *or an* equivalent martingale measure*.*

In (4.1) and below, for a random vector $X = (X^1, \dots, X^d)$ we use the shorthand notation $\mathbb{E}_{\mathbb{Q}}[X]$ for the d-dimensional vector with components $\mathbb{E}_{\mathbb{Q}}[X^i]$, $i = 1, \dots, d$. We first prove the simpler implication in Theorem 4.4.

Proof of Theorem 4.4. (ii) $\Rightarrow$ (i): Let $\bar{\theta} \in \mathbb{R}^{d+1}$ be such that $\bar{\theta} \cdot \bar{S}_1 \geq 0$ $\mathbb{P}$-a.s. and $\mathbb{P}[\bar{\theta} \cdot \bar{S}_1 > 0] > 0$. Then, since $\mathbb{Q} \approx \mathbb{P}$, we also have $\bar{\theta} \cdot \bar{S}_1 \geq 0$ $\mathbb{Q}$-a.s. and $\mathbb{Q}[\bar{\theta} \cdot \bar{S}_1 > 0] > 0$. In particular, $\mathbb{E}_{\mathbb{Q}}[\bar{\theta} \cdot \bar{S}_1] > 0$, and by condition (ii),

$$\bar{\theta} \cdot \bar{S}_0 = \frac{\mathbb{E}_{\mathbb{Q}}[\bar{\theta} \cdot \bar{S}_1]}{1+r} > 0,$$

so $\bar{\theta} \cdot \bar{S}_0$ is positive and thus $\bar{\theta}$ is not an arbitrage. □

Our proof of the other direction will require the following version of the separating hyperplane theorem.

Theorem 4.5. *Suppose that $\mathcal{C} \subset \mathbb{R}^d$ is a non-empty convex set with $0 \notin \mathcal{C}$. Then there exists $\theta \in \mathbb{R}^d$ with $\theta \cdot x \geq 0$ for all $x \in \mathcal{C}$, and with $\theta \cdot x_0 > 0$ for at least one $x_0 \in \mathcal{C}$. Moreover, if $\inf_{x \in \mathcal{C}} |x| > 0$, then one can find $\theta \in \mathbb{R}^d$ with $\inf_{x \in \mathcal{C}} \theta \cdot x > 0$.*

Proof. See, for instance, [6, Proposition A.1]. □

For abbreviation we introduce the vector $Y = (Y^1, \dots, Y^d)$ of discounted net gains

$$Y^i := \frac{S_1^i}{1+r} - S_0^i, \qquad i = 1, \dots, d.$$

Then, by Lemma 4.3 the market is arbitrage-free if and only if, for any $\theta \in \mathbb{R}^d$,

$$\theta \cdot Y \geq 0,\ \mathbb{P}\text{-a.s.} \ \Rightarrow \theta \cdot Y = 0,\ \mathbb{P}\text{-a.s.} \tag{4.2}$$

Moreover, note that for every risk-neutral measure $\mathbb{Q}$ we have $\mathbb{E}_{\mathbb{Q}}[|Y|] < \infty$ and $\mathbb{E}_{\mathbb{Q}}[Y] = 0$.

Proof of Theorem 4.4. (i) $\Rightarrow$ (ii): We need to show that (4.2) implies the existence of some $\mathbb{Q} \approx \mathbb{P}$ such that $\mathbb{E}_{\mathbb{Q}}[|Y|] < \infty$ and $\mathbb{E}_{\mathbb{Q}}[Y] = 0$.

Step 1. We first assume that Y is bounded. Let $\mathcal{Q}$ denote the convex set of all probability measures $\mathbb{Q} \approx \mathbb{P}$ with bounded densities $d\mathbb{Q}/d\mathbb{P}$ and

$$\mathcal{C} := \big\{ \mathbb{E}_{\mathbb{Q}}[Y],\ \mathbb{Q} \in \mathcal{Q} \big\}.$$

Since Y is bounded, all these expectations are trivially finite. Note that $\mathcal{C}$ is a convex set in $\mathbb{R}^d$. Indeed, for any $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{Q}$ and $\alpha \in [0,1]$, we have that $\mathbb{Q}_\alpha := \alpha\mathbb{Q}_1 + (1-\alpha)\mathbb{Q}_2 \in \mathcal{Q}$, and

$$\alpha\, \mathbb{E}_{\mathbb{Q}_1}[Y] + (1-\alpha)\, \mathbb{E}_{\mathbb{Q}_2}[Y] = \mathbb{E}\Big[Y\Big(\alpha \frac{d\mathbb{Q}_1}{d\mathbb{P}} + (1-\alpha)\frac{d\mathbb{Q}_2}{d\mathbb{P}}\Big)\Big] = \mathbb{E}_{\mathbb{Q}_\alpha}[Y] \in \mathcal{C},$$

where we used that

$$\frac{d\mathbb{Q}_\alpha}{d\mathbb{P}} = \alpha \frac{d\mathbb{Q}_1}{d\mathbb{P}} + (1-\alpha)\frac{d\mathbb{Q}_2}{d\mathbb{P}}.$$

Our aim is to show that $0 \in \mathcal{C}$. Let us suppose that $0 \notin \mathcal{C}$. Then, by Theorem 4.5 there exists $\theta \in \mathbb{R}^d$ such that $\theta \cdot x \geq 0$ for all $x \in \mathcal{C}$, and with $\theta \cdot x_0 > 0$ for some $x_0 \in \mathcal{C}$. In other words, $\mathbb{E}_{\mathbb{Q}}[\theta \cdot Y] \geq 0$ for all $\mathbb{Q} \in \mathcal{Q}$ and $\mathbb{E}_{\mathbb{Q}_0}[\theta \cdot Y] > 0$ for some $\mathbb{Q}_0 \in \mathcal{Q}$. Clearly, since $\mathbb{Q}_0 \approx \mathbb{P}$, the latter condition yields $\mathbb{P}[\theta \cdot Y > 0] > 0$. We claim that the first condition implies that

$$\theta \cdot Y \geq 0, \qquad \mathbb{P}\text{-a.s.}, \tag{4.3}$$

which contradicts our assumption (4.2) and will therefore prove that $0 \in \mathcal{C}$.

We now prove (4.3). Let $A := \{\theta \cdot Y < 0\}$ and

$$\phi_n := \Big(1 - \frac{1}{n}\Big)\mathbb{1}_A + \frac{1}{n}\mathbb{1}_{A^c}.$$

We define new probability measures $\mathbb{Q}_n$ via

$$\frac{d\mathbb{Q}_n}{d\mathbb{P}} := \frac{1}{\mathbb{E}[\phi_n]}\,\phi_n, \qquad n \geq 2.$$

Since $0 < \phi_n \leq 1$, we have $\mathbb{Q}_n \in \mathcal{Q}$ and therefore

$$0 \leq \mathbb{E}_{\mathbb{Q}_n}[\theta \cdot Y] = \frac{1}{\mathbb{E}[\phi_n]}\,\mathbb{E}[\theta \cdot Y \phi_n].$$

In particular, $\mathbb{E}[\theta \cdot Y\phi_n] \geq 0$ for all n. Note that $\phi_n \to \mathbb{1}_A$, $\mathbb{P}$-a.s., with $A := \{\theta \cdot Y < 0\}$. Since Y is bounded, we may apply the dominated convergence theorem to obtain that

$$\mathbb{E}\big[\theta \cdot Y \mathbb{1}_{\{\theta \cdot Y < 0\}}\big] = \lim_{n\to\infty} \mathbb{E}[\theta \cdot Y\phi_n] \geq 0.$$

This proves the claim (4.3) and completes the proof in the case where Y is bounded.

Step 2. If Y is not bounded, consider instead

$$\tilde{Y} := \frac{Y}{1+|Y|},$$

which is clearly bounded. From our assumption (4.2) it follows that for any $\theta \in \mathbb{R}^d$,

$$\theta \cdot \tilde{Y} \geq 0,\ \mathbb{P}\text{-a.s.} \ \Rightarrow \theta \cdot \tilde{Y} = 0,\ \mathbb{P}\text{-a.s.}$$

Hence, we may apply Step 1 on $\tilde{Y}$, which implies the existence of a probability measure $\tilde{\mathbb{Q}} \approx \mathbb{P}$ such that $\mathbb{E}_{\tilde{\mathbb{Q}}}[\tilde{Y}] = 0$ and $d\tilde{\mathbb{Q}}/d\mathbb{P}$ is bounded. Set

$$\phi := \frac{c}{1+|Y|}\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}}$$

with $c := 1/\mathbb{E}_{\tilde{\mathbb{Q}}}[(1+|Y|)^{-1}]$ so that $\mathbb{E}[\phi] = 1$. Then $d\mathbb{Q} := \phi\, d\mathbb{P}$ (i.e. $\mathbb{Q}(A) = \mathbb{E}[\phi \mathbb{1}_A]$ for $A \in \mathcal{F}$), defines a probability measure $\mathbb{Q} \ll \mathbb{P}$ with bounded density ϕ, that is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \phi = \frac{c}{1+|Y|}\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}} > 0, \qquad \mathbb{P}\text{-a.s.}$$

Thus, $\mathbb{Q} \approx \mathbb{P}$. Moreover, since $|Y|/(1+|Y|) \leq 1$,

$$\mathbb{E}_{\mathbb{Q}}\big[|Y|\big] = \mathbb{E}\big[|Y|\phi\big] = c\,\mathbb{E}\Big[\frac{|Y|}{1+|Y|}\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}}\Big] \leq c\,\mathbb{E}\Big[\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}}\Big] = c < \infty$$

and since $\tilde{Y} = Y/(1+|Y|)$ by its definition,

$$\mathbb{E}_{\mathbb{Q}}\big[Y\big] = \mathbb{E}\big[Y\phi\big] = c\,\mathbb{E}\Big[\frac{Y}{1+|Y|}\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}}\Big] = c\,\mathbb{E}\Big[\tilde{Y}\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}}\Big] = c\,\mathbb{E}_{\tilde{\mathbb{Q}}}[\tilde{Y}] = 0.$$

Hence, $\mathbb{Q}$ is as desired, and the theorem is proved. □

Remark 4.6. (i) The strictly positive asset S^0 above is referred to as a *numéraire*. Here we considered a situation where there is a single riskless asset (referred to variously as the money-market account, the bond, the bank account, ...) in the market, and it is very common to use this asset as numéraire. It turns out that this will serve for our present applications, but there are occasions when it is advantageous to use other numéraires. The Fundamental Theorem of Asset Pricing holds as long as S_1^0 is strictly positive. Indeed, note that $\bar{\theta}$ is an arbitrage for $\bar{S}$ if and only if it is an arbitrage for $\tilde{S}$ defined by $\tilde{S}_t^i := S_t^i/S_t^0$ for $i \in \{0,1,\dots,d\}$, $t \in \{0,1\}$. The FTAP also does not require the existence of a riskless asset, that is S_1^0 could also be random as long as it is strictly positive.

(ii) Note that the Fundamental Theorem of Asset Pricing does not make any claim about uniqueness of $\mathbb{Q}$ when there is no arbitrage. This is because situations where there is a unique $\mathbb{Q}$ are rare and special; when $\mathbb{Q}$ is unique, the market is called *complete*.

4.2. **Multi-period model.** Consider a multi-period model in which $d+1$ assets are priced at times $t = 0, 1, \dots, T$. The price of asset i at time t is modelled by a non-negative random variable S_t^i on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will write $\bar{S}_t = (S_t^0, S_t) = (S_t^0, \dots, S_t^d)$, $t \in \{0, \dots, T\}$. The stochastic process $(\bar{S}_t)_{t\in\{0,\dots,T\}}$ is assumed to be adapted to a filtration $(\mathcal{F}_t)_{t\in\{0,\dots,T\}}$. Further, we assume that $\mathcal{F}_0$ is $\mathbb{P}$-trivial, i.e. $\mathbb{P}[A] \in \{0, 1\}$ for all $A \in \mathcal{F}_0$. This condition holds if and only if all $\mathcal{F}_0$-measurable random variable are $\mathbb{P}$-a.s. constant.

Definition 4.7. A *trading strategy* is an $\mathbb{R}^{d+1}$-valued, previsible process $\bar{\theta} = (\theta^0, \theta) = (\theta_t^0, \dots, \theta_t^d)_{t=1,\dots,T}$, i.e. $\bar{\theta}_t$ is $\mathcal{F}_{t-1}$-measurable for all $t = 1, \dots, T$.

The value θ_t^i of a trading strategy $\bar{\theta}$ corresponds to the quantity of shares of asset i held between time $t-1$ and time t. Thus, $\theta_t^i S_{t-1}^i$ is the amount invested into asset i at time $t-1$, while $\theta_t^i S_t^i$ is the resulting value at time t. The total value of the portfolio $\bar{\theta}_t$ at time $t-1$ is

$$\bar{\theta}_t \cdot \bar{S}_{t-1} = \sum_{i=0}^{d} \theta_t^i \, S_{t-1}^i.$$

By time t, the value of the portfolio $\bar{\theta}_t$ has changed to

$$\bar{\theta}_t \cdot \bar{S}_t = \sum_{i=0}^{d} \theta_t^i \, S_t^i.$$

The previsibility of $\bar{\theta}$ expresses the fact that investments must be allocated at the beginning of each trading period, without anticipating future price increments.

Definition 4.8. A trading strategy $\bar{\theta}$ is called *self-financing* if

$$\bar{\theta}_t \cdot \bar{S}_t = \bar{\theta}_{t+1} \cdot \bar{S}_t, \qquad \forall t = 1, \dots, T-1. \tag{4.4}$$

Intuitively, (4.4) means that the value of the portfolio at any time t equals the amount invested at time t. It follows that the accumulated gains and losses resulting from the price fluctuations are the only source of variations of the portfolio:

$$\bar{\theta}_{t+1} \cdot \bar{S}_{t+1} - \bar{\theta}_t \cdot \bar{S}_t = \bar{\theta}_{t+1} \cdot \big(\bar{S}_{t+1} - \bar{S}_t\big),$$

and summing up yields

$$\bar{\theta}_t \cdot \bar{S}_t = \bar{\theta}_1 \cdot \bar{S}_0 + \sum_{s=1}^{t} \bar{\theta}_s \cdot \big(\bar{S}_s - \bar{S}_{s-1}\big).$$

Here, the constant $\bar{\theta}_1 \cdot \bar{S}_0$ can be interpreted as the initial investment for the purchase of the portfolio $\bar{\theta}_1$, while the second term may be regarded as a discrete stochastic integral (cf. Proposition 3.8).

We assume from now on that

$$S_t^0 > 0 \qquad \mathbb{P}\text{-a.s. for all } t \in \{0, \dots T\}.$$

This assumption allows us to use asset 0 as *numéraire*. Now we define the *discounted price process*

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

Then $X_t^0 \equiv 1$, and $X_t = (X_t^1, \dots, X_t^d)$ expresses the value of the remaining assets in units of the numéraire.

Definition 4.9. The (discounted) *value process* $V = (V_t)_{t=0,\dots,T}$ of a trading strategy $\bar{\theta}$ is given by

$$V_0 := \bar{\theta}_1 \cdot \bar{X}_0, \qquad V_t := \bar{\theta}_t \cdot \bar{X}_t, \quad t = 1, \dots, T.$$

Proposition 4.10. *For a trading strategy $\bar{\theta}$ the following are equivalent.*

(i) *$\bar{\theta}$ is self-financing.*
(ii) *$\bar{\theta}_t \cdot \bar{X}_t = \bar{\theta}_{t+1} \cdot \bar{X}_t$ for $t = 1, \dots, T-1$.*
(iii) *$V_t = V_0 + \sum_{s=1}^t \theta_s \cdot (X_s - X_{s-1})$ for all t.*

Proof. By dividing both sides of (4.4) by S_t^0 we easily see that (i) and (ii) are equivalent. Moreover, (ii) holds if and only if

$$\bar{\theta}_{t+1} \cdot \bar{X}_{t+1} - \bar{\theta}_t \cdot \bar{X}_t = \bar{\theta}_{t+1} \cdot \big(\bar{X}_{t+1} - \bar{X}_t\big) = \theta_{t+1} \cdot \big(X_{t+1} - X_t\big)$$

for $t = 1, \dots, T-1$, and this identity is equivalent to (iii). □

We now define the notion of an arbitrage in the context of a multi-period model.

Definition 4.11. A self-financing strategy $\bar{\theta}$ is called an *arbitrage opportunity* if its value process V satisfies

$$V_0 \le 0, \qquad V_T \ge 0 \quad \mathbb{P}\text{-a.s.}, \qquad \text{and} \qquad \mathbb{P}\big[V_T > 0\big] > 0.$$

Again we are aiming to characterise those market models that do not allow arbitrage opportunities.

Definition 4.12. A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ is called an *equivalent martingale measure* if $\mathbb{Q} \approx \mathbb{P}$ and the discounted price process X is a d-dimensional martingale under $\mathbb{Q}$. The set of all equivalent martingale measures is denoted by $\mathcal{P}$.

Proposition 4.13. *Let $\mathbb{Q} \in \mathcal{P}$ and $\bar{\theta}$ be a self-financing strategy with value process V satisfying $V_T \ge 0$ $\mathbb{P}$-a.s. Then V is a $\mathbb{Q}$-martingale and $\mathbb{E}_{\mathbb{Q}}[V_T] = V_0$.*

Proof. Step 1. As a warm-up, we first suppose that $\bar{\theta} = (\theta^0, \theta)$ with θ bounded, i.e. $\max_t |\theta_t| \le c < \infty$ for some $c > 0$. Then

$$V_t = V_0 + \sum_{s=1}^t \theta_s \cdot (X_s - X_{s-1}),$$

so that

$$|V_t| \le |V_0| + c \sum_{s=1}^{t} \big(|X_s| + |X_{s-1}|\big).$$

Since X is a $\mathbb{Q}$-martingale and $\mathbb{E}_{\mathbb{Q}}[|X_k|] < \infty$ for each k, we have $\mathbb{E}_{\mathbb{Q}}[|V_t|] < \infty$ for every t. Moreover, for $0 \le t \le T-1$,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}\big[V_{t+1} \,|\, \mathcal{F}_t\big] &= \mathbb{E}_{\mathbb{Q}}\big[V_t + \theta_{t+1} \cdot (X_{t+1} - X_t) \,|\, \mathcal{F}_t\big] \\ &= V_t + \theta_{t+1} \cdot \mathbb{E}_{\mathbb{Q}}\big[(X_{t+1} - X_t) \,|\, \mathcal{F}_t\big] \\ &= V_t,\end{aligned}$$

where we used that V_t and θ_{t+1} are $\mathcal{F}_t$-measurable and X is a $\mathbb{Q}$-martingale. Thus, V is a $\mathbb{Q}$-martingale.

Step 2. Now let $\bar{\theta}$ be as in the statement. In this step we will show that $V_t \ge 0$ $\mathbb{P}$-a.s. for all $t \in \{0, \dots T\}$ by backward induction. For $t = T$ this holds by assumption. Further, note that for any t we have by induction assumption

$$V_{t-1} = V_t - \theta_t \cdot (X_t - X_{t-1}) \ge -\theta_t \cdot (X_t - X_{t-1}).$$

For any $c > 0$ let θ^c be defined via $\theta^c_t := \mathbb{1}_{\{|\theta_t| \le c\}} \theta_t$. Then $\mathbb{E}_{\mathbb{Q}}\left[V_{t-1} \mathbb{1}_{\{|\theta_t| \le c\}} \,|\, \mathcal{F}_{t-1}\right]$ is well defined since

$$V_{t-1} \mathbb{1}_{\{|\theta_t| \le c\}} = V_t \mathbb{1}_{\{|\theta_t| \le c\}} - \theta^c_t \cdot (X_t - X_{t-1}),$$

and the first term is non-negative by the induction assumption and the second term is integrable. Thus,

$$V_{t-1} \mathbb{1}_{\{|\theta_t| \le c\}} = \mathbb{E}_{\mathbb{Q}}\left[V_{t-1} \mathbb{1}_{\{|\theta_t| \le c\}} \,|\, \mathcal{F}_{t-1}\right] \ge -\,\mathbb{E}_{\mathbb{Q}}\left[\theta^c_t \cdot (X_t - X_{t-1}) \,|\, \mathcal{F}_{t-1}\right] = 0.$$

Taking $c \uparrow \infty$ yields $V_{t-1} \ge 0$ $\mathbb{P}$-a.s.

Notice that Step 2 ensures that $\mathbb{E}_{\mathbb{Q}}[V_t \,|\, \mathcal{F}_{t-1}]$ is well-defined for all t.

Step 3. We show the martingale property for V. Indeed, since θ^c_t is $\mathcal{F}_{t-1}$-measurable and X is a $\mathbb{Q}$-martingale,

$$\mathbb{E}_{\mathbb{Q}}\left[V_t \mathbb{1}_{\{|\theta_t| \le c\}} \,|\, \mathcal{F}_{t-1}\right] = \mathbb{E}_{\mathbb{Q}}\left[V_{t-1} \mathbb{1}_{\{|\theta_t| \le c\}} + \theta^c_t \cdot (X_t - X_{t-1}) \,|\, \mathcal{F}_{t-1}\right] = V_{t-1} \mathbb{1}_{\{|\theta_t| \le c\}}.$$

Letting again $c \uparrow \infty$ the monotone convergence theorem gives $\mathbb{E}_{\mathbb{Q}}[V_t \,|\, \mathcal{F}_{t-1}] = V_{t-1}$ $\mathbb{P}$-a.s.

Step 4. Since we have assumed that $\mathcal{F}_0$ is $\mathbb{P}$-trivial, i.e. $\mathbb{P}[A] \in \{0, 1\}$ for all $A \in \mathcal{F}_0$, and $\mathbb{Q} \approx \mathbb{P}$, clearly $\mathcal{F}_0$ is also $\mathbb{Q}$-trivial. Hence, by Proposition 2.11 and Step 3,

$$\mathbb{E}_{\mathbb{Q}}[V_1] = \mathbb{E}_{\mathbb{Q}}\big[V_1 \,|\, \mathcal{F}_0\big] = V_0 < \infty.$$

Moreover, we use repeatedly Step 3 to obtain

$$\mathbb{E}_{\mathbb{Q}}[V_T] = \mathbb{E}_{\mathbb{Q}}\Big[\mathbb{E}_{\mathbb{Q}}\big[V_T \,|\, \mathcal{F}_{T-1}\big]\Big] = \mathbb{E}_{\mathbb{Q}}[V_{T-1}] = \cdots = \mathbb{E}_{\mathbb{Q}}[V_1] = V_0 < \infty.$$

Thus, $\mathbb{E}_{\mathbb{Q}}[V_t] < \infty$ for all t and we have shown that V is a $\mathbb{Q}$-martingale with $\mathbb{E}_{\mathbb{Q}}[V_T] = V_0$. $\square$

Theorem 4.14 (Fundamental Theorem of Asset Pricing, FTAP). *Assume that S^0 is an a.s. strictly positive numéraire, i.e. $S^0_t > 0$ $\mathbb{P}$-a.s. for all $t = 0, \dots T$. Then, the following are equivalent.*

(i) There is no arbitrage.

(ii) $\mathcal{P} \neq \emptyset$, that is there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the discounted price process X defined by

$$X_t := \frac{S_t}{S^0_t}, \qquad t = 0, \dots T,$$

is a $\mathbb{Q}$-martingale.

Proof. (ii) $\Rightarrow$ (i): Let $\mathbb{Q} \in \mathcal{P}$ and $\bar{\theta}$ be a self-financing strategy with a value process V satisfying $V_0 \leq 0$ and $V_T \geq 0$ $\mathbb{P}$-a.s. Then, by Proposition 4.13,

$$\mathbb{E}_{\mathbb{Q}}[V_T] = V_0 \leq 0,$$

which implies $V_T = 0$ $\mathbb{P}$-a.s., so there is no arbitrage.

(i) $\Rightarrow$ (ii): A nice proof in our current discrete time setting, which is based on an application of the Hahn-Banach separation theorem, can be found in [6, Theorem 5.17, Section 1.6]. In continuous time the proof is even much more complicated, see [5]. □

4.3. **European contingent claims.**

Definition 4.15. A non-negative random variable C on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *European contingent claim* or *European option*. A European contingent claim C is called a *derivative* of the underlying assets $S^0, S^1, \dots, S^d$ if C is measurable with respect to the σ-algebra generated by the price process $(\bar{S}_t)_{t=0,\dots,T}$.

A European contingent claim has the interpretation of an asset which yields at time T the amount $C(\omega)$, depending on the scenario ω of the market evolution. T is called the *expiration date* or the *maturity* of C.

Example 4.16. (i) The owner of a *European call option* has the right, but not the obligation, to buy a unit of an asset, say asset i, at time T for a strike price K. The corresponding contingent claim is given by

$$C^{\text{call}} = \left(S^i_T - K\right)^+.$$

Conversely, a *European put option* gives the right, but not the obligation, to sell a unit of an asset at time T for a fixed price K, called *strike price*. This corresponds to a contingent claim of the form

$$C^{\text{put}} = \left(K - S^i_T\right)^+.$$

(ii) The payoff of an *Asian option* depends on the average price

$$S^i_{\text{av}} := \frac{1}{|\mathbb{T}|} \sum_{i \in \mathbb{T}} S^i_t$$

of the underlying asset during a predetermined averaging period $\mathbb{T} \subseteq \{0, \dots, T\}$. Examples are

- Average price call: $\left(S^i_{\mathrm{av}} - K\right)^+$,
- Average price put: $\left(K - S^i_{\mathrm{av}}\right)^+$,
- Average strike call: $\left(S^i_T - S^i_{\mathrm{av}}\right)^+$,
- Average strike put: $\left(S^i_{\mathrm{av}} - S^i_T\right)^+$.

An average strike put can be used, for example, to secure the risk from selling at time T a quantity of an asset which was bought at successive times over the period $\mathbb{T}$.

(iii) The payoff of a *barrier option* depends on whether the price of the underlying asset reaches a certain level before maturity. Most barrier options are either *knock-out* or *knock-in options*. A knock-out barrier option has a zero payoff once the price of the underlying asset reaches a predetermined barrier B. For instance, the so-called *up-and-out call* with strike price K has the payoff

$$C^{\mathrm{call}}_{\mathrm{uo}} = \begin{cases} \left(S^i_T - K\right)^+ & \text{if } \max_{0\le t\le T} S^i_t < B, \\ 0 & \text{else.} \end{cases}$$

Conversely, a knock-in option pays off only if the barrier B is reached. For instance, a *down-and-in put* pays off

$$C^{\mathrm{put}}_{\mathrm{di}} = \begin{cases} \left(K - S^i_T\right)^+ & \text{if } \min_{0\le t\le T} S^i_t < B, \\ 0 & \text{else.} \end{cases}$$

Down-and-out and *up-and-in options* are also traded.

(iv) Using a *lookback option*, one can trade the underlying asset at the maximal or minimal price that occured during the life of the option. A *lookback call* has the payoff

$$S^i_T - \min_{0\le t\le T} S^i_t$$

and a *lookback put*

$$\max_{0\le t\le T} S^i_t - S^i_T.$$

Definition 4.17. A European contingent claim C is called *attainable* or *replicable* if there exists a self-financing strategy $\bar{\theta}$ whose terminal portfolio coincides with C, i.e.

$$C = \bar{\theta}_T \cdot \bar{S}_T \quad \mathbb{P}\text{-a.s.}$$

Such a trading strategy $\bar{\theta}$ is called a *replicating strategy* for C.

The discounted value of a contingent claim C when using S^0 as a numéraire is given by

$$H := \frac{C}{S^0_T},$$

which is called the *discounted European claim* or just *discounted claim* associated with C. Note that a contingent claim C is attainable if and only if the discount claim $H = C/S_T^0$ is of the form

$$H = \frac{\bar{\theta}_T \cdot \bar{S}_T}{S_T^0} = \bar{\theta}_T \cdot \bar{X}_T = V_T = V_0 + \sum_{t=1}^{T} \theta_t \cdot \big(X_t - X_{t-1}\big), \tag{4.5}$$

where V denotes the value process of the replicating strategy $\bar{\theta} = (\theta^0, \theta)$ (cf. Proposition 4.10). In this case, we will also say that the discounted claim H is attainable with replicating strategy $\bar{\theta}$.

From now on, we will assume that our market model is arbitrage-free or, equivalently, that

$$\mathcal{P} \neq \emptyset.$$

Theorem 4.18. *Let H be an attainable discounted claim. Then*

$$\mathbb{E}_{\mathbb{Q}}[H] < \infty \qquad \textit{for all } \mathbb{Q} \in \mathcal{P}.$$

Moreover, for each $\mathbb{Q} \in \mathcal{P}$ the value process V of any replicating strategy satisfies

$$V_t = \mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_t\big] \quad \mathbb{P}\text{-a.s.}, \; t = 0, \dots, T.$$

In particular, V is a non-negative $\mathbb{Q}$-martingale.

Proof. From (4.5) we see that $V_T = H \geq 0$. Then, by Proposition 4.13 the value process V is a $\mathbb{Q}$-martingale for any $\mathbb{Q} \in \mathcal{P}$, so $V_t = \mathbb{E}_{\mathbb{Q}}\big[V_T \,|\, \mathcal{F}_t\big] = \mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_t\big]$. □

Remark 4.19. The last result has two remarkable implications. First, note that $\mathbb{E}_{\mathbb{Q}}\left[H \,|\, \mathcal{F}_t\right]$ is independent of the replicating strategy, so all replicating strategies have the same value process. Further, for any $t = 0, \dots, T$, since $V_t = \bar{\theta}_t \cdot \bar{X}_t$ is independent of the equivalent martingale measure $\mathbb{Q}$, V_t is a version of $\mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_t\big]$ for all $\mathbb{Q} \in \mathcal{P}$.

Pricing a contingent claim. Let us now turn to the problem of *pricing* a contingent claim. Consider an *attainable* discounted claim H with replicating strategy $\bar{\theta}$. Then the (discounted) initial investment

$$\bar{\theta}_1 \cdot \bar{X}_0 = V_0 = \mathbb{E}_{\mathbb{Q}}[H]$$

needed for the replication of H can be interpreted as the *unique (discounted) 'fair price' of H*.

In fact, any different price for H would create an arbitrage opportunity. For instance, if the price $\tilde{\pi}$ of H would be larger than $V_0 = \mathbb{E}_{\mathbb{Q}}[H]$, then at time $t = 0$ an investor could sell H for $\tilde{\pi}$ and buy the portfolio $\bar{\theta}_1$ for V_0. Then, at time $t = 1$ he could buy $\bar{\theta}_2$ for $\bar{\theta}_1 \cdot \bar{X}_1$ and so on. At time $t = T$ the terminal portfolio value $V_T = \bar{\theta}_T \cdot \bar{X}_T$ suffices for settling the claim H at maturity T. This yields a sure profit of $\tilde{\pi} - V_0 > 0$ or, in other words, an arbitrage.

It also becomes clear from these considerations what the seller of an attainable option H needs to do in order to eliminate his risk, in other words to *hedge* the option H. All he needs to do is to buy the replication strategy θ for $\pi = V_0$, which now serves as his *hedging strategy*. Then, at expiration time T the seller will hold a portfolio with value $\bar{\theta}_T \cdot \bar{X}_T = H$, which he can use to settle the claim H.

Example (Put-call parity). The so-called put-call parity is a relation between the prices of a European call option $C^{\text{call}} = \big(S_T^i - K\big)^+$ and the corresponding European put options $C^{\text{put}} = \big(K - S_T^i\big)^+$. More precisely, notice that

$$C^{\text{call}} - C^{\text{put}} = \big(S_T^i - K\big)^+ - \big(K - S_T^i\big)^+ = S_T^i - K,$$

and the right-hand side equals the pay-off of a forward contract (cf. Section 0) with price $S_0^i - \frac{K}{(1+r)^T}$ (note that the contingent claim $C = S_T^i$ can be trivially replicated just by holding one unit of the risky asset which requires an initial investment S_0^i). Hence, the price $\pi(C^{\text{put}})$ for C^{put} can be obtained from the price $\pi(C^{\text{call}})$ for the call C^{call} (or vice versa), namely

$$\pi(C^{\text{put}}) = \pi(C^{\text{call}}) - \Big(S_0^i - \frac{K}{(1+r)^T}\Big).$$

Complete markets. Theorem 4.18 provides not only the price of an attainable claim (i.e. its value at time $t = 0$), but also its value at any time $t \in \{0, \dots, T\}$, which is given by $V_t = \mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_t\big]$, which equals the value of a replicating strategy at time t. We have now discussed pricing market models that are complete in the following sense.

Definition 4.20. An arbitrage-free market model is called *complete* if every European contingent claim is attainable.

Theorem 4.21 (Second Fundamental Theorem of Asset Pricing)**.** *An arbitrage-free market model is complete if and only if there exists exactly one equivalent martingale measure, i.e.* $|\mathcal{P}| = 1$.

Proof. "$\Rightarrow$": If the model is complete, then every discounted claim H is attainable. Let $H := \mathbb{1}_A$ for $A \in \mathcal{F}$. Then there exists a self-financing replicating strategy with initial investment

$$V_0 = \mathbb{E}_{\mathbb{Q}}[H] = \mathbb{Q}(A)$$

for any $\mathbb{Q} \in \mathcal{P}$ (see Theorem 4.18). Recall that V_0 is independent of the particular choice of $\mathbb{Q}$ (cf. Remark 4.19), so this determines $\mathbb{Q}(A)$ uniquely. Since $A \in \mathcal{F}$ is arbitrary, there can only be one equivalent martingale measure.

"$\Leftarrow$": See [6, Theorem 5.39]. ☐

Finally, we record the following property of complete markets which in the context of stochastic analysis in continuous time is usually called the *martingale representation property*.

Proposition 4.22. *Suppose the model is complete with unique equivalent martingale measure $\mathbb{Q}$. Then every $\mathbb{Q}$-martingale M can be represented as a "stochastic integral" of a d-dimensional previsible process θ,*

$$M_t = M_0 + \sum_{k=1}^{t} \theta_k \cdot (X_k - X_{k-1}), \qquad t = 0, \dots, T.$$

Proof. We decompose the terminal value M_T of a $\mathbb{Q}$-martingale M into its positive and negative parts, that is $M_T = M_T^+ - M_T^-$. Then we regard M^+ and M^- as two discounted claims, which are attainable since the model is assumed to be complete. Hence, there exist two d-dimensional previsible processes θ^+ and θ^- such that

$$M_T^{\pm} = V_0^{\pm} + \sum_{k=1}^{T} \theta_k^{\pm} \cdot (X_k - X_{k-1}),$$

for two non-negative constants V_0^+ and V_0^- (cf. (4.5)). By Theorem 4.18 the associated value processes

$$V_t^{\pm} = V_0^{\pm} + \sum_{k=1}^{t} \theta_k^{\pm} \cdot (X_k - X_{k-1}), \qquad t = 0, \dots, T,$$

are $\mathbb{Q}$-martingales and we have that

$$M_t = \mathbb{E}_{\mathbb{Q}}\big[M_T \,|\, \mathcal{F}_t\big] = \mathbb{E}_{\mathbb{Q}}\big[M_T^+ - M_T^- \,|\, \mathcal{F}_t\big] = V_t^+ - V_t^-.$$

Thus, the desired representation for M holds if we define θ to be the d-dimensional previsible process $\theta := \theta^+ - \theta^-$. □

In the next section we will study the prototype of a complete market model, the Cox-Ross-Rubinstein binomial model. However, in discrete time only a very limited class of models turn out to be complete. For incomplete models pricing is more difficult (see e.g. [6, Theorem 5.30]).

5. The Cox-Ross-Rubinstein binomial model

In this section we study the binomial model, a particularly simple model, introduced by Cox, Ross and Rubinstein in [4]. It involves a riskless bond

$$S_t^0 := (1+r)^t, \qquad t = 0, \dots, T,$$

with $r > -1$ and one risky asset S^1 of the form

$$S_t^1 = S_0^1 \prod_{k=1}^{t} (1 + R_k),$$

where the initial value $S_0^1 > 0$ is a given constant and $(R_t)_{t \in \{0,\dots T\}}$ is a family of random variables taking only two possible values $a, b \in \mathbb{R}$ with $-1 < a < b$. Thus, the stock price jumps from S^1_{t-1} either to the higher value $S^1_t = S^1_{t-1}(1+b)$ or to the lower value $S^1_t = S^1_{t-1}(1+a)$. The random variable

$$R_t = \frac{S^1_t - S^1_{t-1}}{S^1_{t-1}}$$

describes the return in the t-th trading period, $t = 1, \dots, T$. In this context, we are going to derive explicit formulas for the arbitrage-free prices and replicating strategies of various contingent claims.

We now construct the model on the sample space

$$\Omega := \{-1,1\}^T = \big\{\omega = (y_1, \dots, y_T) \,|\, y_i \in \{-1,1\}\big\}.$$

Denote by

$$Y_t(\omega) := y_t \qquad \text{for } \omega = (y_1, \dots, y_T)$$

the projection on the t-th coordinate. Further, let

$$R_t(\omega) := a\,\frac{1 - Y_t(\omega)}{2} + b\,\frac{1 + Y_t(\omega)}{2} = \begin{cases} a & \text{if } Y_t(\omega) = -1, \\ b & \text{if } Y_t(\omega) = 1. \end{cases}$$

Now the price process of the risky asset at time T is modelled by

$$S^1_t = S^1_0 \prod_{k=1}^{t} (1 + R_k),$$

where the initial value $S^1_0 > 0$ is a given constant. The discounted value process is given by

$$X_t = \frac{S^1_t}{S^0_t} = S^1_0 \prod_{k=1}^{t} \frac{1 + R_k}{1 + r}.$$

As a filtration we take

$$\mathcal{F}_t := \sigma(S^1_0, \dots, S^1_t) = \sigma(X_0, \dots, X_t), \qquad t = 0, \dots, T.$$

Then, note that $\mathcal{F}_0 = \{\emptyset, \Omega\}$,

$$\mathcal{F}_t = \sigma(Y_1, \dots, Y_t) = \sigma(R_1, \dots, R_t), \qquad t = 1, \dots, T.$$

and $\mathcal{F} = \mathcal{F}_T$ coincides with the power set of Ω. Now we fix any probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{P}\big[\{\omega\}\big] > 0, \qquad \forall \omega \in \Omega,$$

or, in other words,

$$\mathbb{P}\big[R_1 = c_1, \dots, R_T = c_T\big] > 0, \qquad \forall (c_1, \dots, c_T) \in \{a,b\}^T.$$

Definition 5.1. This model is called *binomial model* or *CRR model* (for Cox, Ross, Rubinstein).

Theorem 5.2. *The CRR model is arbitrage-free if and only if $a < r < b$. In this case, there exists a unique equivalent martingale measure $\mathbb{Q}$, i.e. $\mathcal{P} = \{\mathbb{Q}\}$, and $\mathbb{Q}$ is characterised by the fact that the random variables $R_1, \dots, R_T$ are independent under $\mathbb{Q}$ with common distribution*

$$\mathbb{Q}[R_t = b] = p^* := \frac{r-a}{b-a}, \qquad t = 0, \dots, T.$$

Proof. First note that a measure $\mathbb{Q} \in \mathcal{P}$ if and only if X is a martingale under $\mathbb{Q}$, i.e.

$$X_t = \mathbb{E}_{\mathbb{Q}}\big[X_{t+1} \,|\, \mathcal{F}_t\big] = X_t \; \mathbb{E}_{\mathbb{Q}}\left[\frac{1+R_{t+1}}{1+r} \,\Big|\, \mathcal{F}_t\right] \qquad \mathbb{Q}\text{-a.s.}$$

for all $t \leq T-1$, which is equivalent to

$$r = \mathbb{E}_{\mathbb{Q}}\big[R_{t+1} \,|\, \mathcal{F}_t\big] = b\, \mathbb{Q}\big[R_{t+1} = b \,|\, \mathcal{F}_t\big] + a\, \big(1 - \mathbb{Q}\big[R_{t+1} = b \,|\, \mathcal{F}_t\big]\big).$$

This can be rewritten as

$$\mathbb{Q}\big[R_{t+1} = b \,|\, \mathcal{F}_t\big] = \frac{r-a}{b-a} = p^*.$$

But since p^* is a deterministic constant, it can be easily seen that this holds if and only if the random variables $R_1, \dots, R_T$ are i.i.d. with $\mathbb{Q}[R_t = b] = p^*$. In particular, there can be at most one martingale measure for X.

If the model is arbitrage-free, then there exists $\mathbb{Q} \in \mathcal{P}$. Since $\mathbb{Q} \approx \mathbb{P}$ we must have $\mathbb{Q}[R_t = b] = p^* \in (0,1)$, so $a < r < b$.

Conversely, if $a < r < b$ then we can define a measure $\mathbb{Q} \approx \mathbb{P}$ on $(\Omega, \mathcal{F})$ by setting

$$\mathbb{Q}\big[\{\omega\}\big] := (p^*)^k\, (1-p^*)^{T-k} > 0,$$

where k denotes the number of components of $\omega = (y_1, \dots, y_T)$ that are equal to $+1$. Then, under $\mathbb{Q}$, $Y_1, \dots, Y_T$ and hence $R_1, \dots, R_T$ are independent random variables with common distribution $\mathbb{Q}[Y_t = 1] = \mathbb{Q}[R_t = b] = p^*$, so $\mathbb{Q} \in \mathcal{P}$ and thus there is no arbitrage opportunity. $\square$

From now on we only consider CRR models that are arbitrage-free, so we assume that $a < r < b$ and denote by $\mathbb{Q}$ the unique equivalent martingale measure.

Now we turn to the problem of pricing and hedging a given contingent claim C. Let $H = C/(1+r)^T$ be the discounted claim, which can be written as $H = h(S_0^1, S_1^1, \dots, S_T^1)$ for some suitable function h.

Proposition 5.3. *The value process*

$$V_t = \mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_t\big], \qquad t = 0, \dots, T,$$

of a replicating strategy for H is of the form

$$V_t(\omega) = v_t\big(S_0^1, S_1^1(\omega), \dots, S_t^1(\omega)\big),$$

where the function v_t is given by

$$v_t(x_0, \dots, x_t) = \mathbb{E}_{\mathbb{Q}}\left[h\Big(x_0, \dots, x_t, x_t \cdot \frac{S_1^1}{S_0^1}, \dots, x_t \cdot \frac{S_{T-t}^1}{S_0^1}\Big)\right]. \qquad (5.1)$$

Proof. Since the equivalent martingale measure is unique, the model is complete and the every contingent claim is attainable, and by Theorem 4.18 the value process of any replicating strategy is given by $V_t = \mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_t\big]$. Thus,

$$V_t = \mathbb{E}_{\mathbb{Q}}\left[h\Big(S_0^1, \ldots, S_t^1, S_t^1 \cdot \frac{S_{t+1}^1}{S_t^1}, \ldots, S_t^1 \cdot \frac{S_T^1}{S_t^1}\Big) \,\Big|\, \mathcal{F}_t\right].$$

Recall that $S_0^1, S_1^1, \ldots S_t^1$ are $\mathcal{F}_t$-measurable and note that S_{t+s}^1/S_t^1 is independent of $\mathcal{F}_t$ and has under $\mathbb{Q}$ the same distribution as

$$\frac{S_s^1}{S_0^1} = \prod_{k=1}^{s}(1 + R_k).$$

The claim follows now from Proposition 2.16. □

Since the value process V is characterised by the recursion

$$V_T = H, \qquad V_t = \mathbb{E}_{\mathbb{Q}}\big[V_{t+1} \,|\, \mathcal{F}_t\big],$$

we obtain the following recursion formula for the function v_t in Proposition 5.3,

$$v_T\big(x_0, \ldots, x_T\big) = h(x_0, \ldots, x_T),$$

and for $t < T$,

$$\begin{aligned} &v_t\big(x_0, \ldots, x_t\big) \\ &\quad = p^*\, v_{t+1}\big(x_0, x_1, \ldots, x_t, x_t(1+b)\big) + (1-p^*)\, v_{t+1}\big(x_0, x_1, \ldots, x_t, x_t(1+a)\big). \end{aligned} \tag{5.2}$$

Indeed, for $t < T$,

$$\begin{aligned} v_t(S_0^1, \ldots, S_t^1) &= \mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_t\big] = \mathbb{E}_{\mathbb{Q}}\Big[\mathbb{E}_{\mathbb{Q}}\big[H \,|\, \mathcal{F}_{t+1}\big] \,\big|\, \mathcal{F}_t\Big] \\ &= \mathbb{E}_{\mathbb{Q}}\big[v_{t+1}(S_0^1, \ldots, S_{t+1}^1) \,|\, \mathcal{F}_t\big] = \mathbb{E}_{\mathbb{Q}}\left[v_{t+1}\Big(S_0^1, \ldots, S_t^1, S_t^1 \cdot \frac{S_{t+1}^1}{S_t^1}\Big) \,\Big|\, \mathcal{F}_t\right] \\ &= p^*\, v_{t+1}\big(S_0^1, S_1^1, \ldots, S_t^1, S_t^1(1+b)\big) + (1-p^*)\, v_{t+1}\big(S_0^1, S_1^1, \ldots, S_t^1, S_t^1(1+a)\big). \end{aligned}$$

Example 5.4. Suppose that $H = h(S_T^1)$ only depends on the terminal value S_T^1 of the stock price, then V_t depends only on the value S_t^1 of the stock at time t , i.e. $V_t(\omega) = v_t(S_t^1(\omega))$ and the formula (5.1) reduces to

$$\begin{aligned} v_t(x_t) &= \mathbb{E}_{\mathbb{Q}}\left[h\Big(x_t \frac{S_{T-t}^1}{S_0^1}\Big)\right] \\ &= \sum_{k=0}^{T-t} h\Big(x_t\,(1+a)^{T-t-k}\,(1+b)^k\Big)\binom{T-t}{k}(p^*)^k\,(1-p^*)^{T-t-k}. \end{aligned}$$

In particular, the unique arbitrage-free price of H is given by

$$\pi(H) = v_0(S_0^1) = \sum_{k=0}^{T} h\Big(S_0^1\,(1+a)^{T-k}\,(1+b)^k\Big)\binom{T}{k}(p^*)^k\,(1-p^*)^{T-k}.$$

For instance, by choosing $h(x) = (x-K)^+/(1+r)^T$ or $h(x) = (K-x)^+/(1+r)^T$ we get explicit formulas for the arbitrage-free prices of a European call or European put, respectively.

Next we derive a hedging strategy for a discounted claim $H = h(X_0, \ldots, X_T)$. By hedging strategy we mean a self-financing trading strategy the seller of an option can use in order to secure his position at maturity time T. For instance, if the option is attainable, any replicating strategy can serve as a hedging strategy.

Proposition 5.5. *The hedging strategy is given by*

$$\theta_t(\omega) = \Delta_t(S_0^1, S_1^1(\omega), \ldots, S_{t-1}^1(\omega)),$$

where

$$\begin{aligned}\Delta_t&\big(x_0, x_1, \ldots, x_{t-1}\big)\\ &= (1+r)^t\,\frac{v_t\big(x_0, x_1, \ldots, x_{t-1}, x_{t-1}(1+b)\big) - v_t\big(x_0, x_1, \ldots, x_{t-1}, x_{t-1}(1+a)\big)}{x_{t-1}\big(b-a\big)}.\end{aligned}$$

The term Δ_t may be regarded as a discrete derivative of the value function v_t with respect to the possible stock price changes. In financial language, a hedging strategy based on a derivative of the value process is often called a *Delta hedge*.

Proof. By Proposition 4.10 we have that for each $\omega = (y_1, \ldots, y_T) \in \{-1, 1\}^T$ any self-financing strategy $\bar{\theta}$ must satisfy

$$\theta_t(\omega) \cdot \big(X_t(\omega) - X_{t-1}(\omega)\big) = V_t(\omega) - V_{t-1}(\omega). \tag{5.3}$$

In this equation the random variables θ_t, X_{t-1} and V_{t-1} depend only on the first $t-1$ components of ω. For a fixed t we now define

$$\omega^{\pm} := (y_1, \ldots, y_{t-1}, \pm 1, y_{t+1}, \ldots, y_T).$$

Plugging ω^+ and ω^- into (5.3) gives

$$\begin{aligned}\theta_t(\omega)\left(X_{t-1}(\omega)\,\frac{1+b}{1+r} - X_{t-1}(\omega)\right) &= V_t(\omega^+) - V_{t-1}(\omega)\\ \theta_t(\omega)\left(X_{t-1}(\omega)\,\frac{1+a}{1+r} - X_{t-1}(\omega)\right) &= V_t(\omega^-) - V_{t-1}(\omega).\end{aligned}$$

Taking the difference and solving for $\theta_t(\omega)$ gives

$$\theta_t(\omega) = \big(1+r\big)\,\frac{V_t(\omega^+) - V_t(\omega^-)}{(b-a)X_{t-1}(\omega)} = \big(1+r\big)^t\,\frac{V_t(\omega^+) - V_t(\omega^-)}{(b-a)S_{t-1}^1(\omega)},$$

and the claim follows. ☐

Remark 5.6. Proposition 5.5 only provides the number of shares of the risky asset to be held in each trading period. However, the process θ^0 describing the investment in the bond can be directly deduced as the following discussion shows.

If $\bar{\theta}$ is a self-financing trading strategy, then $(\bar{\theta}_{t+1} - \bar{\theta}_t) \cdot \bar{X}_t = 0$ for all $t = 1, \ldots, T-1$. In particular, the numéraire component satisfies

$$\theta^0_{t+1} - \theta^0_t = -(\theta_{t+1} - \theta_t) \cdot X_t, \qquad t = 1, \ldots, T-1, \tag{5.4}$$

and

$$\theta^0_1 = V_0 - \theta_1 \cdot X_0. \tag{5.5}$$

Thus, the entire process θ^0 is determined by the initial investment V_0 and the d-dimensional process θ. Consequently, if a value $V_0 \in \mathbb{R}$ and any d-dimensional previsible process θ are given, we can define the process θ^0 via (5.4) and (5.5) to obtain a self-financing strategy $\bar{\theta} := (\theta^0, \theta)$ with initial capital V_0, and this construction is unique.

6. Brownian motion

The binomial model for a share is a discrete-time model, and as such it is a poor approximation to the reality of a market, where trading happens in an almost continuous fashion. We might try to make the binomial model describe such a market better by thinking of the time period as being very short, such as one second, or even one microsecond; if we did this, there would be a very large number of moves of the share in an hour. Recall that under the equivalent martingale measure the share price in the binomial model, or more precisely its logarithm, is a random walk (its steps are independent identically distributed random variables), and in view of the Central Limit Theorem, it would not be surprising if there existed some (distributional) limit of the binomial random walk as the time periods became ever shorter. It would also be expected that the Gaussian distribution should feature largely in that limit process, and indeed it does. This chapter introduces the basic ideas about a continuous-time process called *Brownian motion*, in terms of which the most common continuous-time model of a share is defined. Using this model, various derivative prices can be computed in closed form; the celebrated Black-Scholes formula for the price of a European call option is the prime example. We finish these motivating remarks with a very short overview about the history of Brownian motion.

1827: **Robert Brown** observes the jittery motion of a grain of pollen in water

1900: **Louis Bachelier** discusses in his Ph.D.-thesis the use of Brownian motion as a model for share prices.

1905: **Albert Einstein** formulates a diffusion equation for the motion of particles in a fluid. A particle in water undergoes an enormous number of bombardments by the fast-moving molecules in the fluid, roughly of the order of 10^{13} collisions per second (at room temperature). So the particle performs a random walk on a very short scale. The increments of this random walk should have mean zero and the variance should be propertional to the number of collisions, i.e. proportional to the elapsed time. Let X_t denote the position of the particle at time t and x its initial position. In view of the huge number of collisions and the weak strength of every single push, the central limit theorem would suggest that it is reasonable to assume that $X_t \sim \mathcal{N}(x, \sigma^2 t)$ for some $\sigma > 0$. Furthermore, the evolution of the motion of the particle on disjoint time intervals should be independent.

1923: **Norbert Wiener** provides a mathematical model for Brownian motion.

1965: **Paul Samuelson** suggest a geometric Brownian motion as a model for share prices, more precisely,

$$S_t = S_0 \exp(\sigma B_t + \mu t),$$

where B is a Brownian motion, $\mu \in \mathbb{R}$ a drift and $\sigma > 0$ a volatility parameter.

We now turn to the precise definition.

Definition 6.1. A stochastic process $(B_t)_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *Brownian motion* or *Wiener process* if

(a) $B_0 = 0$, $\mathbb{P}$-a.s.

(b) For any $\omega \in \Omega$, the map $t \mapsto B_t(\omega)$ is continuous.

(c) For any $n \in \mathbb{N}$ and any $0 = t_0 < t_1 < \cdots < t_n$, the increments $B_{t_1}, B_{t_2} - B_{t_1}, ..., B_{t_n} - B_{t_{n-1}}$ are independent and each increment $B_{t_i} - B_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})$, so it is a Gaussian random variable with mean zero and variance $t_i - t_{i-1}$.

Remark 6.2. (i) Brownian motion is a *Markov process* with the transition probility density given by

$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}} \exp\Big(-\frac{(y-x)^2}{2t}\Big), \qquad t > 0,\ x, y \in \mathbb{R}, \tag{6.1}$$

so for any $n \in \mathbb{N}$ and any $0 < t_1 < \cdots < t_n$ the distribution of $(B_{t_1}, \ldots, B_{t_n})$ is given by

$$\begin{aligned} &\mathbb{P}\big[B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n\big] \\ &= \int_{A_1}\int_{A_2} \cdots \int_{A_n} p_{t_1}(0, x_1)\, p_{t_2 - t_1}(x_1, x_2) \cdots p_{t_n - t_{n-1}}(x_{n-1}, x_n)\, \mathrm{d}x_n \cdots \mathrm{d}x_1 \end{aligned}$$

for all $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$.

(ii) In the definition, $B_0 = 0$ is not essential; for general $x \in \mathbb{R}$ we call $(x + B_t)_{t\geq 0}$ a Brownian motion started at x.

(iii) In the definition, the condition (b) stating that B has continuous sample paths is in fact an additional requirement and does not follow from (a) and (c). Indeed, let B be a Brownian motion and set

$$\bar{B}_t(\omega) := B_t(\omega)\, \mathbb{1}_{\mathbb{R}\setminus\mathbb{Q}}(B_t(\omega)), \qquad t > 0.$$

Then, for every t, $\bar{B}_t$ has the same distribution as B_t, so $\bar{B}$ satisfies conditions (a) and (c) but $\bar{B}$ is obviously not continuous. One can even show that it is not continuous at any point.

Alternatively, we could describe Brownian motion as follows.

Lemma 6.3. *Brownian motion is the Gaussian process*[4] $(B_t)_{t\geq 0}$ *with values in* $\mathbb{R}$ *such that*

(a) $B_0 = 0$, *$\mathbb{P}$-a.s.*
(b) For $\mathbb{P}$-a.e. ω, the map $t \mapsto B_t(\omega)$ is continuous.
(c) $\mathbb{E}[B_t] = 0$ *and* $\mathbb{E}[B_t B_s] = t \wedge s$ *for all* $s, t \geq 0$.

Proof. Let B be Brownian motion as defined in Definiton 6.1. Then properties (a) and (b) are obviously satisfied. To show that (c) holds, we may assume without loss of generality that $t > s$. Then

$$\mathbb{E}\big[B_t\, B_s\big] = \mathbb{E}\big[(B_t - B_s)B_s + B_s^2\big] = \mathbb{E}[B_t - B_s]\ \mathbb{E}[B_s] + \mathbb{E}\big[B_s^2\big] = 0 + s = s \wedge t,$$

where we used that $B_t - B_s$ and B_s are independent and centred and B_s has variance s.

To prove the converse, i.e. that any process with the properties given in the statement is a Brownian motion, we can just use the fact that the law of a Gaussian process is uniquely determined by its mean and covariance (see e.g. [3, Section 3]). Thus the process has the same law as Brownian motion and has only continuous paths (by (b)), so it is Brownian motion. □

Once we have Brownian motion in one dimension, we can trivially define Brownian motion in d dimensions.

Definition 6.4. A d-dimensional Brownian motion is a stochastic process $(B_t)_{t\geq 0}$ with values in $\mathbb{R}^d$, such that if $B = (B^1, \dots, B^d)$, then the components B^i are mutually independent Brownian motions in $\mathbb{R}$.

The question remains whether such a process actually exists.

Theorem 6.5. *Brownian motion exists.*

We will not give a formal proof here; some nice short constructions of Brownian motion can be found, for instance, in [3, Section 6] or [8, Section 7]. However, the maybe most natural approach would require a good amount of preparation, so we only sketch the main idea here.

Let $(Y_i)_{i\in\mathbb{N}}$ be i.i.d. with $\mathbb{E}[Y_i] = 0$ and variance $E[Y_i^2] = 1$. Consider the random walk X defined by $X_0 := 0$ and $X_k := \sum_{i=1}^k Y_i$, $k \geq 1$ and let

$$X_t^{(n)} := \frac{1}{n} X_{\lfloor n^2 t\rfloor} + \frac{tn^2 - \lfloor tn^2 \rfloor}{n}\left(X_{\lfloor n^2 t\rfloor + 1} - X_{\lfloor n^2 t\rfloor}\right), \qquad t \geq 0,$$

[4] A stochastic process $(X_t)_{t\geq 0}$ is called a *Gaussian process* if for any $n \in \mathbb{N}$ and any $0 < t_1 < \cdots < t_n$ the vector $(X_{t_1}, \dots, X_{t_n})$ is normally distributed.

that is, $X^{(n)}_{t_k} = X_k/n$ for $t_k = k/n^2$, $k \geq 0$, and on each interval $[t_k, t_{k+1}]$, $X^{(n)}$ interpolates linearly between X_k/n and X_{k+1}/n. Then, Donsker's invariance principle (Donsker, 1952) states that

$$\big(X^{(n)}\big)_{t\geq 0} \underset{n\to\infty}{\Rightarrow} \big(B_t\big)_{t\geq 0}$$

where B is a Brownian motion (and $\Rightarrow$ denotes convergence in distribution). If we would just consider the sequence $X^{(n)}_1$, i.e. the case $t = 1$, this is exactly the Central Limit Theorem, since $B_1 \sim \mathcal{N}(0,1)$. But Donsker's theorem is in fact a much stronger result as it provides such a convergence *simultaneously for all* $t \geq 0$. Therefore the theorem is also called *Functional Central Limit Theorem*. Equvalently, this result could be also formulated as follows. The rescaled random walk $X^{(n)}$ and the Brownian motion B may be regarded as random variables taking values in the *path space* $C([0,\infty),\mathbb{R})$. Then Donsker's theorem says that the distribution $\mathbb{P}\circ(X^{(n)}_{\cdot})^{-1}$ of the entire path of the rescaled random walk, which is a measure on $C([0,\infty),\mathbb{R})$, converges weakly to the distribution of Brownian motion, which also called the *Wiener measure*.

As a consequence, on an intuitive level Brownian motion looks locally like a random walk on a very large time scale, so the paths are very rough. We will make this now a bit more precise. For any $t > 0$ consider a partition Π of $[0,t]$ of the form

$$\Pi = \big\{t_0, t_1, \ldots, t_k\big\} \quad \text{with } 0 \leq t_0 < t_1 < \ldots < t_k = t,\, k \in \mathbb{N},$$

and set $|\Pi| := \max_{t_i\, \in \Pi} |t_i - t_{i-1}|$ (with $t_{-1} := 0$).

Definition 6.6. Let $p \geq 1$ and let $(\Pi_n)_{n\in\mathbb{N}}$ be an increasing sequence of partitions of $[0,t]$ (i.e. $\Pi_n \subseteq \Pi_{n+1}$ for all $n \in \mathbb{N}$) such that $|\Pi_n| \to 0$ as $n \to \infty$. For any $f \in C([0,t])$,

$$V_t^p(f) := \lim_{n\to\infty} V^p_{t,\Pi_n}(f),$$

where

$$V^p_{t,\Pi_n}(f) := \sum_{t_i\in\Pi_n} \big|f(t_i) - f(t_{i-1})\big|^p,$$

is called the p-*variation* of f.

There are two important special cases.

(i) In the case $p = 1$, note that $V^1_{t,\Pi_n}(f) \leq V^1_{t,\Pi_{n+1}}(f)$ for all n, so that $V^1_t(f) = \sup_n V^1_{t,\Pi_n}(f)$. Further, we observe that $V^1_t(f)$ is the length of the curve defined by f on $[0,t]$. If $V^1_t(f) < \infty$, then f is called a *function of bounded variation* (or *function of finite variation*) on $[0,t]$.

(ii) In the case $p = 2$, $V^2(f)$ is called *quadratic variation of* f.

Remark 6.7. If $p < p'$ and $V^p_t(f) < \infty$ then $V^{p'}_t(f) = 0$. (Exercise!)

Example 6.8. (i) If f is differentiable, then f is of bounded variation on any finite interval (Exercise).

(ii) If f is increasing (or decreasing), then f is of bounded variation on any finite interval (Exercise).

Proposition 6.9. *Let B be a Brownian motion and $(\Pi_n)_{n\in\mathbb{N}}$ be an increasing sequence of partitions of $[0,t]$ with $|\Pi_n| \to 0$ as $n \to \infty$. Then*

$$\lim_{n\to\infty} V^2_{t,\Pi_n}(B) = t \qquad \text{in } L^2(\mathbb{P}),$$

that is

$$\lim_{n\to\infty} \mathbb{E}\big[\big(V^2_{t,\Pi_n}(B) - t\big)^2\big] = 0.$$

Proof. Let n initially be be fixed and let $\Pi_n = \{t_0, t_1, \dots, t_k\}$ be a partition of $[0,t]$. For abbreviation we set $\Delta B_j := B_{t_j} - B_{t_{j-1}}$ and $\Delta t_j := t_j - t_{j-1}$. Recall that $\Delta B_j \sim \mathcal{N}(0, \Delta t_j)$ so that $\mathbb{E}\big[(\Delta B_j)^2\big] = \Delta t_j$ and $\mathbb{E}\big[(\Delta B_j)^4\big] = 3(\Delta t_j)^2$. Further, note that $\sum_{i\le k} \Delta t_i = t$. Thus,

$$\begin{aligned}
\mathbb{E}\big[\big(V^2_{t,\Pi_n}(B) - t\big)^2\big] &= \mathbb{E}\Big[\Big(\sum_{i\le k} \big(\Delta B_i\big)^2 - t\Big)^2\Big] \\
&= \mathbb{E}\Big[\sum_{i\le k}\sum_{j\le k} \big(\Delta B_i\big)^2\big(\Delta B_j\big)^2 - 2t\sum_{i\le k}\big(\Delta B_i\big)^2 + t^2\Big] \\
&= \sum_{i\le k} \underbrace{\mathbb{E}\big[(\Delta B_i)^4\big]}_{=3(\Delta t_j)^2} + \sum_{\substack{i,j\le k\\ i\ne j}} \underbrace{\mathbb{E}\big[(\Delta B_i)^2\big]}_{=\Delta t_i}\underbrace{\mathbb{E}\big[(\Delta B_j)^2\big]}_{=\Delta t_j} - 2t\sum_{i\le k}\underbrace{\mathbb{E}\big[(\Delta B_i)^2\big]}_{=\Delta t_i} + t^2 \\
&\le 3\sum_{i\le k}\big(\Delta t_i\big)^2 + t^2 - 2t^2 + t^2 \\
&\le 3\,|\Pi_n| \sum_{i\le k} \Delta t_i = 3\,|\Pi_n|\, t \to 0, \qquad \text{as } n\to\infty,
\end{aligned}$$

which gives the claim. ☐

By combining Proposition 6.9 with Remark 6.7 one can show that $\mathbb{P}$-almost every Brownian path is locally of infinite variation. To finish our discussion of regularity of Brownian paths, we briefly mention their modulus of continuity. Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is locally *Hölder continuous of order α* for $\alpha \in [0,1]$ if, for every $L > 0$,

$$\sup\left\{\frac{|f(t) - f(s)|}{|t-s|^\alpha},\ |t|,|s| \le L,\ t \ne s\right\} < \infty.$$

If $\alpha = 1$, then f is locally Lipschitz-continuous.

Proposition 6.10. *Let B be a Brownian motion. Then, almost surely,*

(i) for all $\alpha < 1/2$, B is locally Hölder continuous of order α,

(ii) for all $\alpha \ge 1/2$, B is nowhere Hölder continuous of order α. In particular, B is nowhere differentiable.

Proof. See e.g. [8, Theorem 7.7.2]. ☐

Proposition 6.11. *Let B be a Brownian motion. Then each of the following processes are also Brownian motions.*

(i) $B_t^1 := -B_t$,
(ii) $B_t^2 := cB_{t/c^2}$ *for any* $c > 0$ *(scale invariance),*
(iii) $B_t^3 := t\,B_{1/t}$ *for* $t > 0$ *and* $B_0^3 := 0$ *(time-inversion),*
(iv) $B_t^4 := B_{T+t} - B_T$ *for any* $T \geq 0$ *fixed.*

Proof. We leave the proofs for (i), (ii) and (iv) to the reader as an exercise. To see (iii) we note first that B^3 is a Gaussian process with mean $\mathbb{E}[B_t^3] = 0$ for all $t \geq 0$ and covariance

$$\operatorname{cov}(B_s^3, B_t^3) = \mathbb{E}\big[B_s^3 B_t^3\big] = s\,t\;\mathbb{E}\big[B_{1/s}B_{1/t}\big] = s\,t\,\tfrac{1}{t} = s, \qquad 0 < s \leq t.$$

Hence, again using the fact that the law of a Gaussian process is determined by its mean and covariance, we get that for any $0 \leq t_1 < \cdots < t_n$ the law of $(B_{t_1}^3, \ldots, B_{t_n}^3)$ is the same as the law of $(B_{t_1}, \ldots, B_{t_n})$. Further, the paths $t \mapsto B_t^3$ are almost sure continuous on $(0, \infty)$. It remains to show the continuity at $t = 0$. Let $\mathbb{Q}_+ := \mathbb{Q} \cap (0, \infty)$, $(A_t)_{t \in \mathbb{Q}_+}$ be a collection of sets in $\mathcal{B}(\mathbb{R})$ and $\{s_n, n \geq 1\}$ be a numbering of the elements in $\mathbb{Q}_+$. Then, by the monotone continuity of the measure $\mathbb{P}$,

$$\begin{aligned}\mathbb{P}\Bigg[\bigcap_{t \in \mathbb{Q}_+} \{B_t^3 \in A_t\}\Bigg] &= \lim_{N \to \infty} \mathbb{P}\Bigg[\bigcap_{n=1}^{N} \{B_{s_n}^3 \in A_{s_n}\}\Bigg] = \lim_{N \to \infty} \mathbb{P}\Bigg[\bigcap_{n=1}^{N} \{B_{s_n} \in A_{s_n}\}\Bigg] \\ &= \mathbb{P}\Bigg[\bigcap_{t \in \mathbb{Q}_+} \{B_t \in A_t\}\Bigg].\end{aligned}$$

Hence, also the distribution of $(B_t^3, t \in \mathbb{Q}_+)$ is the same as the distribution of $(B_t, t \in \mathbb{Q}_+)$. In particular,

$$\lim_{t \downarrow 0 R, t \in \mathbb{Q}_+} B_t^3 = 0, \qquad \mathbb{P}\text{-a.s.}$$

But $\mathbb{Q}_+$ is dense in $(0, \infty)$ and B^3 is almost surely continuous on $(0, \infty)$, so that

$$0 = \lim_{\substack{t \downarrow 0 \\ t \in \mathbb{Q}_+}} B_t^3 = \lim_{t \downarrow 0} B_t^3, \qquad \mathbb{P}\text{-a.s.}$$

Thus, B^3 is also continuous at $t = 0$. □

Corollary 6.12 (Law of large numbers)**.** *Let B be a Brownian motion. Then,*

$$\lim_{t \to \infty} \frac{B_t}{t} = 0, \qquad \mathbb{P}\text{-}a.s.$$

Proof. Let B^3 be as in Proposition 6.11, then

$$\lim_{t \to \infty} \frac{B_t}{t} = \lim_{t \to \infty} B_{1/t}^3 = B_0^3 = 0, \qquad \mathbb{P}\text{-a.s.}$$

□

Proposition 6.13. *Let B be a Brownian motion. Then,*

$$\mathbb{P}\left[\sup_{t\geq 0} B_t = \infty\right] = 1.$$

Proof. Set $Z := \sup_{t\geq 0} B_t$ and $\tilde{B}_t := c^{-1} B_{c^2 t}$ for any $c > 0$. Then by scaling invariance

$$Z = \sup_{t\geq 0} B_t = \sup_{t\geq 0} B_{c^2 t} = c \sup_{t\geq 0} \tilde{B}_t \overset{(d)}{=} c \sup_{t\geq 0} B_t = c\, Z.$$

In particular, $\mathbb{P}[Z \leq z] = \mathbb{P}[cZ \leq z]$ for all $z > 0$, so the distribution function $F(z) = \mathbb{P}[Z \leq z]$ of Z is constant on $(0, \infty)$, which shows that $Z \in \{0, +\infty\}$ a.s.

Recall that $B'_t = B_{1+t} - B_1$ is another Brownian motion, so $Z' = \sup_{t\geq 0} B'_t$ has the same law as Z. In particular $Z' \in \{0, \infty\}$ a.s. It suffices to show that $\mathbb{P}[Z = 0] = 0$. Note that on the event $\{Z = 0\}$ we have $Z' \neq +\infty$ and therefore $Z' = 0$. Furthermore, $\{Z = 0\} \subseteq \{B_1 \leq 0\}$. Hence, also using the fact that Brownian motion has independent increments we get

$$\begin{aligned}
\mathbb{P}[Z = 0] &= \mathbb{P}[Z = 0,\, Z' = 0] \leq \mathbb{P}\big[B_1 \leq 0,\, \sup_{t\geq 0} B_{1+t} - B_1 = 0\big] \\
&= \mathbb{P}[B_1 \leq 0]\, \mathbb{P}\big[\sup_{t\geq 0} B_{1+t} - B_1 = 0\big] = \frac{1}{2}\, \mathbb{P}[Z = 0],
\end{aligned}$$

which implies $\mathbb{P}[Z = 0] = 0$. □

7. Martingales in continuous time

We start with some definitions.

Definition 7.1. (i) A function $f : [0, \infty) \to \mathbb{R}$ is said to be *cadlag* if it is right-continuous with left limits, that is to say, for all $t \geq 0$,

$$f(s) \to f(t) \qquad \text{as } s \to t \text{ with } s > t$$

and, for all $t > 0$, there exists $f(t-) \in \mathbb{R}$ such that

$$f(s) \to f(t-) \qquad \text{as } s \to t \text{ with } s < t.$$

The term is a French acronym for *continu à droite, limité à gauche*.

(ii) A *continuous stochastic process* is a family of random variables $(X_t)_{t\geq 0}$ such that, for all $\omega \in \Omega$, the *path* (or *trajectory*) $t \mapsto X_t(\omega) : [0, \infty) \to \mathbb{R}$ is continuous. Similarly, a *cadlag stochastic process* is a family of random variables $(X_t)_{t\geq 0}$ such that, for all $\omega \in \Omega$, the path $t \mapsto X_t(\omega) : [0, \infty) \to \mathbb{R}$ is cadlag.

A continuous stochastic process $(X_t)_{t\geq 0}$ can then be considered as a random variable X in $C([0, \infty), \mathbb{R})$ given by

$$X(\omega) = (t \mapsto X_t(\omega) : t \geq 0).$$

Similarly, a cadlag stochastic process can be thought of as a random variable in in the space of cadlag functions.

We assume in this and the next section that our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is equipped with a *continuous-time filtration*, that is, a family of σ-algebras $(\mathcal{F}_t)_{t\geq 0}$ such that

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, \qquad s \leq t.$$

Definition 7.2. A stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is a *martingale* if

(i) X is adapted, that is X_t is $\mathcal{F}_t$-measurable for every t.
(ii) X is integrable, i.e. $\mathbb{E}[|X_t|] < \infty$ for every t.
(iii) The *martingale property* holds, i.e. for all $0 \leq s \leq t$,

$$\mathbb{E}[X_t | \mathcal{F}_s] = X_s, \qquad \mathbb{P}\text{-a.s.}$$

If equality is replaced in this condition by $\leq$ or $\geq$, we obtain notions of *supermartingale* and *submartingale*, respectively.

Next we discuss some important examples for martingales which will be constructed from Brownian motion and the Poisson process, respectively.

Brownian motion. One of the most fundamental examples for a continuous martingale is Brownian motion. Indeed, let $(B_t)_{t\geq 0}$ be a Brownian motion with its natural filtration given by $\mathcal{F}_t = \mathcal{F}_t^B = \sigma(B_s, s \leq t)$. Then B is clearly integrable and adapted to $(\mathcal{F}_t)_{t\geq 0}$, and for $0 \leq s \leq t$,

$$\mathbb{E}\big[B_t \,|\, \mathcal{F}_s\big] = B_s + \mathbb{E}\big[B_t - B_s \,|\, \mathcal{F}_s\big] = B_s + \mathbb{E}\big[B_t - B_s\big] = B_s,$$

where we used that the increment $B_t - B_s$ is independent of $\mathcal{F}_s$. So B is a martingale, but there are more martingales related to Brownian motion.

Example 7.3. Let $(B_t)_{t\geq 0}$ be a Brownian motion. Then the processes $(B_t^2 - t)_{t\geq 0}$ and $(\exp(\sigma B_t - \frac{1}{2}\sigma^2 t))_{t\geq 0}$ for $\sigma > 0$ are both martingales w.r.t. $\mathcal{F}_t = \mathcal{F}_t^B$.

Proof. The process $(B_t^2 - t)_{t\geq 0}$ is clearly adapted and integrable. For $0 \leq s \leq t$,

$$\begin{aligned}
\mathbb{E}\big[B_t^2 \,|\, \mathcal{F}_s\big] &= \mathbb{E}\big[\big(B_t - B_s + B_s\big)^2 \,|\, \mathcal{F}_s\big] \\
&= \mathbb{E}\big[\big(B_t - B_s\big)^2 \,|\, \mathcal{F}_s\big] + 2B_s\, \mathbb{E}\big[B_t - B_s \,|\, \mathcal{F}_s\big] + B_s^2 \\
&= \mathbb{E}\big[\big(B_t - B_s\big)^2\big] + B_s^2 = t - s + B_s^2,
\end{aligned}$$

where we used again that $B_t - B_s$ is independent of $\mathcal{F}_s$ and $B_t - B_s \sim \mathcal{N}(0, t-s)$. Thus, $\mathbb{E}\big[B_t^2 - t \,|\, \mathcal{F}_s\big] = B_s^2 - s$ and the martingale property holds.

The verification that $(\exp(\sigma B_t - \frac{1}{2}\sigma^2 t))_{t\geq 0}$ is a martingale is left as an exercise. $\square$

Poisson process. One of the simplest relevant examples for a non-continuous cadlag stochastic process is the Poisson process.

Definition 7.4 (Poisson process)**.** A stochastic process $(N_t)_{t\geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *Poisson process with intensity* λ if

(a) $N_0 = 0$ and for any $\omega \in \Omega$, the map $t \mapsto N_t(\omega)$ is cadlag.

(b) For any $n \in \mathbb{N}$ and any $0 = t_0 < t_1 < \cdots < t_n$, the increments $N_{t_1}, N_{t_2} - N_{t_1}, ..., N_{t_n} - N_{t_{n-1}}$ are independent.

(c) For every $0 \leq s < t$ the increment $N_t - N_s \sim \mathrm{Pois}(\lambda(t-s))$, i.e.

$$\mathbb{P}[N_t - N_s = k] = e^{-\lambda(t-s)} \frac{\big(\lambda(t-s)\big)^k}{k!}, \qquad k \in \{0, 1, 2, \ldots\}.$$

Remark 7.5. One possible way to construct a Poisson process is as follows. Let $W_1, W_2, \ldots$ be i.i.d. $\exp(\lambda)$ distributed random variables. Set $T_n := \sum_{i=1}^n W_i$. Then,

$$N_t := \sup\big\{n : T_n \leq t\big\}, \qquad t \geq 0,$$

is a Poisson process with intensity λ.

Example 7.6. Let $(N_t)_{t\geq 0}$ be a Poisson process with intensity λ. Then, the processes $(N_t - \lambda t)_{t\geq 0}$ and $\big((N_t - \lambda t)^2 - \lambda t\big)_{t\geq 0}$ are martingales w.r.t. the filtration $\mathcal{F}_t = \mathcal{F}_t^N$.

Proof. Both processes are clearly adapted and integrable. To see the martingale property, recall that the increments $N_t - N_s$ are independent of $\mathcal{F}_s$ and Poisson-distributed with Parameter $\lambda(t-s)$. Hence,

$$\begin{aligned}
\mathbb{E}\big[N_t - \lambda t \,\big|\, \mathcal{F}_s\big] &= \mathbb{E}\big[N_t - N_s - \lambda(t-s) \,\big|\, \mathcal{F}_s\big] + N_s - \lambda s \\
&= \mathbb{E}\big[N_t - N_s - \lambda(t-s)\big] + N_s - \lambda s \\
&= N_s - \lambda s.
\end{aligned}$$

Hence, $X_t := N_t - \lambda t$ is a martingale. Similarly as above,

$$\begin{aligned}
&\mathbb{E}\big[(N_t - \lambda t)^2 - \lambda t \,\big|\mathcal{F}_s\big] \\
&\quad= \mathbb{E}\big[(X_t - X_s)^2 \,\big|\mathcal{F}_s\big] + 2\mathbb{E}\big[X_s(X_t - X_s) \,\big|\mathcal{F}_s\big] + \mathbb{E}\big[X_s^2 - \lambda t \,\big|\mathcal{F}_s\big] \\
&\quad= \mathbb{E}\big[(X_t - X_s)^2\big] + 2X_s\mathbb{E}\big[X_t - X_s \,\big|\, \mathcal{F}_s\big] + X_s^2 - \lambda t \\
&\quad= \underbrace{\mathbb{E}\big[(X_t - X_s)^2\big]}_{=\mathrm{Var}(N_t - N_s)} + X_s^2 - \lambda t \\
&\quad= \lambda(t-s) + X_s^2 - \lambda t = X_s^2 - \lambda s,
\end{aligned}$$

where we used the martingale property of X in the third step. □

Our next goal is to derive Doob's maximal inequality, the martingale convergence theorem and Doob's stopping theorem for martingales in continuous time. For that purpose we write, for $n \geq 0$,

$$\mathbb{D}_n = \{k2^{-n} : k \in \mathbb{Z}^+\}.$$

Define, for a cadlag stochastic process X,

$$X^* = \sup_{t\geq 0} |X_t|, \qquad X^{(n)*} = \sup_{t\in\mathbb{D}_n} |X_t|.$$

The cadlag property implies that

$$X^{(n)*} \uparrow X^* \qquad \text{as } n \to \infty$$

while, if $(X_t)_{t\geq 0}$ is a cadlag martingale, then $(X_t)_{t\in\mathbb{D}_n}$ is a discrete-time martingale, for the filtration $(\mathcal{F}_t)_{t\in\mathbb{D}_n}$, and similarly for supermartingales and submartingales. Thus, on applying Doob's inequalities to $(X_t)_{t\in\mathbb{D}_n}$ and passing to the limit, which is justified by the monotone convergence theorem, we obtain the following results.

Theorem 7.7 (Doob's maximal and L^p inequalities). *Let X be a cadlag martingale or non-negative submartingale. Then, for all $\lambda \geq 0$,*

$$\lambda\,\mathbb{P}(X^* \geq \lambda) \leq \sup_{t\geq 0} \mathbb{E}[|X_t|],$$

and for all $p > 1$,

$$\mathbb{E}\big[(X^*)^p\big] \leq \Big(\frac{p}{p-1}\Big)^p \sup_{t\geq 0} \mathbb{E}\big[|X_t|^p\big].$$

Similarly, the cadlag property implies that every upcrossing of a non-trivial interval by $(X_t)_{t\geq 0}$ corresponds, eventually as $n \to \infty$, to an upcrossing by $(X_t)_{t\in\mathbb{D}_n}$. This leads to the following estimate.

Theorem 7.8 (Doob's upcrossing inequality). *Let X be a cadlag supermartingale and let $a, b \in \mathbb{R}$ with $a < b$. Then*

$$(b-a)\,\mathbb{E}[U_\infty(X, [a,b])] \leq \sup_{t\geq 0} \mathbb{E}\big[(X_t - a)^-\big]$$

where $U_\infty(X,[a,b])$ is the total number of disjoint upcrossings of $[a,b]$ by X.

Then, arguing as in the discrete-time case, we obtain continuous-time versions of each martingale convergence theorem. We say that $(X_t)_{t\geq 0}$ is L^1-bounded if $\sup_{t\geq 0} \mathbb{E}[|X_t|] < \infty$.

Theorem 7.9 (Almost sure martingale convergence theorem). *Let X be an L^1-bounded cadlag supermartingale. Then there exists an integrable random variable X_∞ such that $X_t \to X_\infty$ almost surely as $t \to \infty$.*

We say that a random variable

$$\tau : \Omega \to [0, \infty]$$

is a *stopping time* if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. Given a cadlag stochastic process X, we define X_τ and the stopped process X^τ by

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega), \qquad X_t^\tau(\omega) = X_{\tau(\omega)\wedge t}(\omega)$$

where we leave $X_\tau(\omega)$ undefined if $\tau(\omega) = \infty$ and $X_t(\omega)$ fails to converge as $t \to \infty$.

Theorem 7.10 (Optional stopping theorem). *Let X be a cadlag martingale.*

(i) *For all bounded stopping times σ and τ with $\sigma \leq \tau$, X_σ and X_τ are integrable and $\mathbb{E}[X_\tau] = \mathbb{E}[X_\sigma]$.*
(ii) *For all stopping times τ, the stopped process X^τ is a martingale.*
(iii) *For all stopping times τ such that X^τ is bounded and $\mathbb{P}[\tau < \infty] = 1$, we have*

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

Sketch of proof. The idea is to apply the corresponding results in discrete time on the martingale $(X_t)_{t\in\mathbb{D}_n}$ and the stopping times $\sigma_n := 2^{-n}\lceil 2^n\sigma\rceil$ and $\tau_n := 2^{-n}\lceil 2^n\tau\rceil$ (here $\lceil x\rceil$ denotes the smallest integer larger or equal to $x\in\mathbb{R}$). Then, $\sigma_n, \tau_n \in \mathbb{D}_n$ and $\sigma_n \downarrow \sigma$ and $\tau_n \downarrow \tau$ as $n\to\infty$. Since X is cadlag and therefore right-continuous, we have $X_{\tau_n} \to X_\tau$ and $X_{\sigma_n} \to X_\sigma$, and statements (i)-(iii) follow from their analogues in discrete time by the dominated convergence theorem, the application of which needs to be justified. This can be done by showing that $(X_{\tau_n})_n$ and $(X_{\sigma_n})_n$ are *uniformly integrable*, which we omit here. See, for instance, [8, Theorem 4.3.8] for details. □

As for the discrete-time version, it suffices that X^τ is bounded by some intergrable random variable for Theorem 7.10-(iii) to hold (cf. Remark 3.20-(iii) above). Similar versions of Theorem 7.10 holds if X is a submartingale or supermartingale, respectively. A typical application of the optional stopping theorem is the following statement on the hitting times of Brownian motion.

Corollary 7.11. *Let $(B_t)_{t\geq 0}$ be a Brownian motion. For $a,b>0$ and $x\in\mathbb{R}$ we define*

$$\tau_x := \inf\{t\geq 0 : B_t = x\}, \quad \tau_{a,b} := \min\big(\tau_{-a},\tau_b\big).$$

Then τ_x and $\tau_{a,b}$ are stopping times with $\mathbb{P}[\tau_x<\infty]=1$, $\mathbb{P}[\tau_b<\tau_{-a}]=\frac{a}{b+a}$ and $\mathbb{E}[\tau_{a,b}]=ab$.

Proof. Exercise. □

8. Continuous semimartingales and quadratic variation

In this section we introduce semimartingales and quadratic variation processes which play a central role in stochastic calculus.

8.1. **Local martingales.**

Definition 8.1. A cadlag adapted process X is a *local martingale* if there exists a sequence of stopping times $(\tau_n)_n$ such that $\tau_n\uparrow\infty$ as $n\to\infty$ and X^{τ_n} is a martingale for every n. The sequence $(\tau_n)_n$ is called a *localising sequence*.

Remark 8.2. Every martingale is a local martingale (take simply $\tau_n = n$ and use that stopped martingales are martingales). On the other hand, there exists local martingales which are not martingales (see [7, Exercise 3.36 in Chapter 3] for an example).

We write $\mathcal{M}$ for the set of all continuous martingales and $\mathcal{M}_{\text{loc}}$ for the set of all continuous local martingales.

Lemma 8.3. *Let $X\in\mathcal{M}_{\text{loc}}$.*

(i) If X is non-negative, then X is a supermartingale.
(ii) If X is bounded, then $X\in\mathcal{M}$.

Proof. (i) Clearly, X is an adapted process. For every $t \geq 0$ we have by Fatou's lemma,

$$\mathbb{E}[X_t] = \mathbb{E}\big[\lim_{n\to\infty} X_{t\wedge\tau_n}\big] \leq \liminf_{n\to\infty} \mathbb{E}[X_{t\wedge\tau_n}] = \mathbb{E}[X_0] < \infty,$$

and for $0 \leq s < t$,

$$\mathbb{E}\big[X_t \,|\, \mathcal{F}_s\big] = \mathbb{E}\big[\lim_{n\to\infty} X_{t\wedge\tau_n} \,|\, \mathcal{F}_s\big] \leq \liminf_{n\to\infty} \mathbb{E}\big[X_{t\wedge\tau_n} \,|\, \mathcal{F}_s\big] = \liminf_{n\to\infty} X_{s\wedge\tau_n} = X_s,$$

where we used that X^{τ_n} is a martingale in the third step.

(ii) Let $C \in (0,\infty)$ be such that $|X| \leq C$. Then $C - X \geq 0$ and $C + X \geq 0$, and statement (i) implies that $C - X$ and $C + X$ are both supermartingales. In particular, $\pm X$ are both supermartingales, so X is a martingale. □

Recall the definition of the 1-variation in Definition 6.6.

Definition 8.4. We say that a continuous stochastic process X is of bounded variation if $V_t^1(X(\omega)) < \infty$ for $\mathbb{P}$-a.e. $\omega \in \Omega$ and all $t > 0$. We write $\mathcal{A}$ for the set of all continuous adapted stochastic processes X of bounded variation.

Definition 8.5. A process X is called a continuous *semimartingale* if there exist $M \in \mathcal{M}_{\mathrm{loc}}$ and $A \in \mathcal{A}$ such that

$$X = M + A. \tag{8.1}$$

The decomposition in (8.1) is almost surely unique (up to the starting points of M and A) as the following result shows.

Theorem 8.6. *Let $X \in \mathcal{M}_{\mathrm{loc}}$ be of bounded variation. Then X is $\mathbb{P}$-a.s. constant.*

Proof. Without loss of generality let $X_0 = 0$. Further, let $V_t := V_t^1(X)$ be the 1-variation of X on $[0,t]$ for any $t > 0$, i.e.

$$V_t = \sup_n \sum_{t_i \in \Pi_n} \big|X_{t_i} - X_{t_{i-1}}\big|,$$

for any increasing sequence $(\Pi_n)_n$ of partitions of $[0,t]$ with $|\Pi_n| \to 0$ as $n \to \infty$. Fix $K > 0$ and define the stopping time

$$\sigma := \inf\{s : V_s \geq K\}.$$

Since $|X_t| \leq V_t$ for all $t \geq 0$, we have that $|X_t| \leq K$ for all $t \leq \sigma$. In particular, $M_t := X_{t\wedge\sigma}$ defines a bounded martingale. Note that for any $0 \leq s < t$,

$$\mathbb{E}\big[(M_t - M_s)^2 \,|\, \mathcal{F}_s\big] = \mathbb{E}\big[M_t^2 \,|\, \mathcal{F}_s\big] - 2M_s\,\mathbb{E}\big[M_t \,|\, \mathcal{F}_s\big] + M_s^2 = \mathbb{E}\big[M_t^2 - M_s^2 \,|\, \mathcal{F}_s\big].$$

Let $\Pi_n = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ be a partition of $[0, t]$ with $|\Pi_n| \to 0$. Then,

$$\begin{aligned}
\mathbb{E}\left[M_t^2\right] &= \mathbb{E}\left[\sum_{k=1}^{n} M_{t_k}^2 - M_{t_{k-1}}^2\right] = \mathbb{E}\left[\sum_{k=1}^{n} \left(M_{t_k} - M_{t_{k-1}}\right)^2\right] \\
&\le \mathbb{E}\left[\left(\max_{1\le k\le n} \left|M_{t_k} - M_{t_{k-1}}\right|\right) \sum_{k=1}^{n} \left|M_{t_k} - M_{t_{k-1}}\right|\right] \\
&\le \mathbb{E}\left[\left(\max_{1\le k\le n} \left|M_{t_k} - M_{t_{k-1}}\right|\right) V_{t\wedge\sigma}\right] \le K\,\mathbb{E}\left[\max_{1\le k\le n} \left|M_{t_k} - M_{t_{k-1}}\right|\right]. \qquad (8.2)
\end{aligned}$$

Recall that X has continuous paths. In particular, $s \mapsto M_s$ is uniformly continuous on $[0, t]$. Hence, $\mathbb{P}$-a.s.,

$$\max_{1\le k\le n} \left|M_{t_k} - M_{t_{k-1}}\right| \to 0 \qquad \text{as } |\Pi_n| \to 0.$$

On the other hand,

$$\max_{1\le k\le n} \left|M_{t_k} - M_{t_{k-1}}\right| = \max_{1\le k\le n} \left|X_{t_k\wedge\sigma} - X_{t_{k-1}\wedge\sigma}\right| \le 2K,$$

so by the dominated convergence theorem we also have that

$$\mathbb{E}\left[\max_{1\le k\le n} \left|M_{t_k} - M_{t_{k-1}}\right|\right] \to 0 \qquad \text{as } |\Pi_n| \to 0.$$

Thus, it follows from (8.2) that $\mathbb{E}[M_t^2] = 0$ and therefore $M_t = 0$, $\mathbb{P}$-a.s., which in turn implies that $\mathbb{P}[M_t = 0 \text{ for all } t \in \mathbb{Q}_+] = 1$. Since M has continuous paths, we get that $\mathbb{P}[M_t = 0 \text{ for all } t \ge 0] = 1$. In other words, $\mathbb{P}[X_t = 0 \text{ for all } t \in [0, \sigma]] = 1$. Since $\sigma \uparrow \infty$ as $K \uparrow \infty$, this gives $\mathbb{P}[X_t = 0 \text{ for all } t \ge 0] = 1$. □

8.2. **Quadratic variation and covariation.** We start with the continuous-time version of the Doob decomposition in Theorem 3.24, which holds under some additional technical assumptions on the underlying filtration, the so-called 'usual condition', which we do not discuss here.

Theorem 8.7 (Doob-Meyer decomposition). *Let X be a non-negative continuous submartingale. Then, there exist $M \in \mathcal{M}$ and an increasing process $A \in \mathcal{A}$ with $M_0 = A_0 = 0$ such that*

$$X_t = X_0 + M_t + A_t, \qquad t \ge 0.$$

This decomposition is unique in the sense that for any other $M' \in \mathcal{M}$ and $A' \in \mathcal{A}$ increasing with $M'_0 = A'_0 = 0$ we have $\mathbb{P}[M_t = M'_t, \forall t \ge 0] = \mathbb{P}[A_t = A'_t, \forall t \ge 0] = 1$.

Proof. In order to show uniqueness, assume that $X_t = X_0 + M_t + A_t = X_0 + M'_t + A'_t$. Then, $M_t - M'_t = A'_t - A_t$ is a martingale of bounded variation starting at zero, and uniqueness follows from Theorem 8.6.

The proof of existence is much more difficult, see for instance [7, Theorem 4.10 in Chapter 1]. □

Remark 8.8. The assumption that X is non-negative can be replaced by a certain integrability condition ('uniform integrability'). In particular, continuous submartingales and supermartingales, which are non-negative or satisfy those integrability conditions, are continuous semimartingales. This follows immediately from Theorem 8.7.

Now we introduce the class $\mathcal{M}^2$ of continuous, square-integrable martingales starting at zero,

$$\mathcal{M}^2 := \{M \in \mathcal{M} : \mathbb{E}[M_t^2] < \infty \text{ for every } t \geq 0 \text{ and } M_0 = 0\}.$$

Theorem 8.9. *Let $M, N \in \mathcal{M}^2$.*

(i) *There exists a unique increasing process $\langle M \rangle = (\langle M \rangle_t)_{t \geq 0} \in \mathcal{A}$ with $\langle M \rangle_0 = 0$ such that*

$$M^2 - \langle M \rangle \in \mathcal{M}.$$

(ii) *There exists a unique process $\langle M, N \rangle = (\langle M, N \rangle_t)_{t \geq 0} \in \mathcal{A}$ with $\langle M, N \rangle_0 = 0$ such that*

$$M \cdot N - \langle M, N \rangle \in \mathcal{M}.$$

Proof. (i) Let $M \in \mathcal{M}^2$. Then M^2 is a non-negative submartingale. By the Doob-Meyer decomposition, there exists an increasing process $A \in \mathcal{A}$ with $A_0 = 0$ such that $M^2 = M' + A$ for some $M' \in \mathcal{M}$. We now define $\langle M \rangle := A$ to obtain that $M^2 - \langle M \rangle = M' \in \mathcal{M}$.

(ii) This follows from the polarisation identity

$$M \cdot N = \frac{1}{4}\Big(\big(M + N\big)^2 - \big(M - N\big)^2\Big). \tag{8.3}$$

and by applying statement (i) on the martingales $M + N$ and $M - N$. □

Definition 8.10. Let $M, N \in \mathcal{M}^2$.

(i) The process $\langle M \rangle$ is called the *quadratic variation of M*.
(ii) The process $\langle M, N \rangle$ is called the *covariation of M and N*.

Note that from the polarisation identity (8.3) it follows that

$$\langle M, N \rangle = \frac{1}{4}\big(\langle M + N \rangle - \langle M - N \rangle\big). \tag{8.4}$$

This definition of the quadratic variation is compatible with the one below Definition 6.6 as the following result shows.

Theorem 8.11. *Let $M, N \in \mathcal{M}^2$ and for any $t > 0$ let $(\Pi_n)_n$ be an increasing sequence of partitions of $[0, t]$ with $|\Pi_n| \to 0$ as $n \to \infty$. Then,*

$$\lim_{n \to \infty} \sum_{t_i \in \Pi_n} \big(M_{t_i} - M_{t_{i-1}}\big)^2 = \langle M \rangle_t \qquad \textit{in probability},$$

i.e., for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big[\Big| \sum_{t_i\in\Pi_n} \big(M_{t_i} - M_{t_{i-1}}\big)^2 - \langle M\rangle_t\Big| > \varepsilon\Big] \to 0.$$

Moreover,

$$\lim_{n\to\infty} \sum_{t_i\in\Pi_n} (M_{t_i} - M_{t_{i-1}})(N_{t_i} - N_{t_{i-1}}) = \langle M, N\rangle_t \quad \textit{in probability}.$$

Proof. See [7, Theorem 5.8 in Chapter 1]. □

Example 8.12. Let $(B_t)_{t\geq 0}$ be a Brownian motion. We have seen in Example 7.3 that $B_t^2 - t$ is a martingale. Hence, $\langle B\rangle_t = t$ for all $t \geq 0$. This is also confirmed by Proposition 6.9.

We now record some properties of the quadratic variation.

Lemma 8.13. *Let $M, N \in \mathcal{M}^2$ and τ be a stopping time.*

(i) $\langle\cdot,\cdot\rangle$ is symmetric, bilinear and positive definite.
(ii) $\langle M, N\rangle^\tau = \langle M^\tau, N^\tau\rangle$.
(iii) $\langle M\rangle = 0 \Leftrightarrow M = 0$.
(iv) $\langle M, N\rangle \leq \langle M\rangle^{1/2}\langle N\rangle^{1/2}$.
(v) Let A be a continuous process with $\langle A\rangle = 0$. Then $\langle M + A\rangle = \langle M\rangle$.

Proof. (i) is an easy exercise.

(ii) By (8.4) it suffices to show $\langle M\rangle^\tau = \langle M^\tau\rangle$. But since $M^2 - \langle M\rangle \in \mathcal{M}$, this follows from the fact that

$$(M^\tau)^2 - \langle M\rangle^\tau = \big(M^2 - \langle M\rangle\big)^\tau \in \mathcal{M}$$

and the uniqueness of the quadratic variation.

(iii) Suppose that $\langle M\rangle = 0$. Then M^2 is a martingale, in particular for any $t \geq 0$,

$$\mathbb{E}[M_t^2] = \mathbb{E}[M_0^2] = 0,$$

since $M_0 = 0$. By Doob's L^2-maximal inequality

$$\mathbb{E}\big[\sup_{t\geq 0} M_t^2\big] \leq 4 \sup_{t\geq 0} \mathbb{E}[M_t^2] = 0.$$

Hence, $\mathbb{P}[\sup_{t\geq 0} M_t^2 = 0] = 1$. This shows that if $\langle M\rangle = 0$ then $M = 0$. The reverse implication is trivial.

(iv) We use Theorem 8.11. For any increasing sequence $(\Pi_n)_n$ of partitions of $[0, t]$ with $|\Pi_n| \to 0$ as $n \to \infty$, we have by the Cauchy Schwarz inequality,

$$\begin{aligned}\langle M, N\rangle &= \lim_{n\to\infty} \sum_{t_i\in\Pi_n} (M_{t_i} - M_{t_{i-1}})(N_{t_i} - N_{t_{i-1}})\\ &\leq \lim_{n\to\infty} \Big(\sum_{t_i\in\Pi_n} \big(M_{t_i} - M_{t_{i-1}}\big)^2\Big)^{\frac{1}{2}} \Big(\sum_{t_i\in\Pi_n} \big(N_{t_i} - N_{t_{i-1}}\big)^2\Big)^{\frac{1}{2}} = \langle M\rangle^{\frac{1}{2}}\langle N\rangle^{\frac{1}{2}}.\end{aligned}$$

(v) Note that

$$\langle M+A\rangle = \langle M\rangle + 2\langle M,A\rangle + \langle A\rangle = \langle M\rangle + 2\langle M,A\rangle,$$

and from (iv) we get $|\langle M,A\rangle| \le \langle M\rangle^{1/2}\langle A\rangle^{1/2} = 0$. □

Example 8.14. Let X be a continuous, adapted process such that $X_0 = 0$ and $\mathbb{E}[X_t^2] < \infty$ for all $t \ge 0$. Suppose that X has independent increments with mean zero. Then,

(i) $X \in \mathcal{M}^2$.
(ii) $\langle X\rangle_t = \operatorname{var}(X_t) = \mathbb{E}[X_t^2]$ for all $t \ge 0$.

Proof. Exercise. □

9. The Cameron-Martin theorem

In this section we will investigate how Brownian motion is transformed under a change of measure. Such results will turn out to be fundamental for pricing in continuous time models for which, as in the discrete time case, equivalent martingale measure will play a crucial role. It will be convenient to specify the underlying probability space, similarly as we did in Section 5. Let $\Omega = C([0,T])$ the *path space* of continuous function on $[0,T]$ and denote by B the *coordinate process*, that is $B_t(\omega) = \omega_t$ for $t \in [0,T]$ and $\omega \in \Omega$. We endow Ω with the σ-algebra $\mathcal{F} = \sigma(B_t, 0 \le t \le T)$, which can be shown to coincide with the Borel σ-algebra on $C([0,T])$ (with respect to the topology induced by the uniform convergence on $[0,T]$). Finally, let $\mathbb{P}$ be the probability measure on $(\Omega,\mathcal{F})$ under which the coordinate process B is a Brownian motion. This measure is known as *Wiener measure* and its existence follows along with the existence of Brownian motion from Donsker's invariance principle, for instance (cf. the discussion below Theorem 6.5).

Consider now a *Brownian motion with drift at speed* c, that is

$$B_t + c\,t, \qquad 0 \le t \le T,$$

for any $c \in \mathbb{R}$. Recall that the transition density of the Brownian motion B is given by

$$p_t(x,y) := p_t(x-y) := \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right), \qquad t > 0,\ x,y \in \mathbb{R}.$$

We now compute the finite dimensional distributions of the process $(B_t + ct)_{t\in[0,T]}$. For any $0 = t_0 < t_1 < \cdots < t_n = T$ and any $A_1,\ldots,A_n \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned}
&\mathbb{P}\left[B_{t_1} + ct_1 \in A_1,\ldots,B_{t_n} + ct_n \in A_n\right] \\
&\quad = \int_{A_1 - ct_1} \cdots \int_{A_n - ct_n} p_{t_1}(x_1)\, p_{t_2-t_1}(x_2 - x_1) \cdots p_{t_n - t_{n-1}}(x_n - x_{n-1})\, \mathrm{d}x_n \cdots \mathrm{d}x_1 \\
&\quad = \int_{A_1} \cdots \int_{A_n} p_{t_1}(y_1 - ct_1)\, p_{t_2-t_1}(y_2 - y_1 - c(t_2 - t_1)) \cdots \\
&\qquad \times\ p_{t_n - t_{n-1}}(y_n - y_{n-1} - c(t_n - t_{n-1}))\, \mathrm{d}y_n \cdots \mathrm{d}y_1,
\end{aligned}$$

where we used the substitution $y_i = x_i + ct_i$. Since

$$\begin{aligned} &p_{t_i - t_{i-1}}\big(y_i - y_{i-1} - c(t_i - t_{i-1})\big) \\ &\qquad = p_{t_i - t_{i-1}}(y_i - y_{i-1}) \exp\big(c\,(y_i - y_{i-1}) - \tfrac{c^2}{2}(t_i - t_{i-1})\big) \end{aligned}$$

and $\sum_{i=1}^{n} t_i - t_{i-1} = t_n = T$ and $\sum_{i=1}^{n} y_i - y_{i-1} = y_n$ with $y_0 := 0$, this becomes

$$\begin{aligned} &\mathbb{P}\left[B_{t_1} + ct_1 \in A_1, \dots, B_{t_n} + ct_n \in A_n\right] \\ &\qquad = \int_{A_1} \cdots \int_{A_n} p_{t_1}(y_1)\, p_{t_2 - t_1}(y_2 - y_1) \cdots \\ &\qquad\qquad \times\, p_{t_n - t_{n-1}}(y_n - y_{n-1}) \exp\big(cy_n - \tfrac{c^2}{2}t_n\big)\, \mathrm{d}y_n \cdots \mathrm{d}y_1 \\ &\qquad = \mathbb{E}\left[\mathbb{1}_{\{B_{t_1} \in A_1, \dots, B_{t_n} \in A_n\}} \exp\big(cB_{t_n} - \tfrac{c^2}{2}t_n\big)\right] \end{aligned}$$

with $t_n = T$. We have just shown that any *cylindrical functional* F, that is a functional $F : C([0,T]) \to \mathbb{R}$ of the form

$$F(\omega) = \begin{cases} 1 & \text{if } \omega_{t_1} \in A_1, \dots, \omega_{t_n} \in A_n, \\ 0 & \text{else,} \end{cases}$$

satisfies

$$\mathbb{E}\big[F\big(B_t + ct : 0 \le t \le T\big)\big] = \mathbb{E}\Big[F\big(B_t : 0 \le t \le T\big) \exp\big(cB_T - \tfrac{c^2}{2}T\big)\Big].$$

By linearity and approximation arguments this can be extended to all bounded and measurable $F : C([0,T]) \to \mathbb{R}$. Choosing $F = \mathbb{1}_A$ for any $A \in \mathcal{F}$ we get

$$\mathbb{P}^{(c)}[A] := \mathbb{P}\Big[\big(B_t + ct : 0 \le t \le T\big) \in A\Big] = \int_A \exp\big(cB_T - \tfrac{c^2}{2}T\big)\, \mathrm{d}\mathbb{P}\,.$$

Thus, the measures $\mathbb{P}^{(c)}$ and $\mathbb{P}$ are equivalent with Radon-Nikodym density given by

$$\frac{d\mathbb{P}^{(c)}}{d\mathbb{P}} = \exp\big(cB_T - \tfrac{c^2}{2}T\big).$$

To summarize, under the Wiener measure $\mathbb{P}$ the paths in $C(0,T])$ have the distribution of a Brownian motion while under the measure $\mathbb{P}^{(c)}$ the paths in $C(0,T])$ have the distribution of a Brownian motion with drift c. We have arrived at

Theorem 9.1 (Cameron-Martin theorem). *For any $c \in \mathbb{R}$, $T > 0$ and any bounded and measurable $F : C([0,T]) \to \mathbb{R}$,*

$$\begin{aligned} \mathbb{E}\big[F\big(B_t + ct : 0 \le t \le T\big)\big] &= \mathbb{E}\Big[F\big(B_t : 0 \le t \le T\big) \exp\big(cB_T - \tfrac{c^2}{2}T\big)\Big] \\ &= \mathbb{E}_{\mathbb{P}^{(c)}}\Big[F\big(B_t : 0 \le t \le T\big)\Big]. \end{aligned}$$

The Cameron-Martin theorem is a special case of the more general Girsanov theorem, which is classical material in lectures on stochastic calculus. It describes the effect on continuous local martingales of an absolutely continuous change of the probability measure. In Theorem 9.1 we are restricted to the case where the continuous local martingale is just Brownian motion.

10. The Black-Scholes model

In 1965 Paul Samuelson proposed the following market model in continuous time. There is a riskless bond

$$S_t^0 = e^{rt}, \qquad 0 \le t \le T,$$

with interest rate $r \ge 0$ and one risky asset with price process given

$$S_t = S_0 \exp(\sigma B_t + \mu t), \qquad 0 \le t \le T,$$

where B denotes a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mu \in \mathbb{R}$ a drift, $\sigma > 0$ a volatility parameter and $S_0 > 0$ the initial price of the asset.

Fisher Black and Myron Scholes (1973) and Robert Merton (1973) added the crucial replication argument which leads to a complete pricing and hedging theory. Merton and Scholes received the 1997 Nobel Memorial Prize in Economic Sciences for their work (sadly, Black was ineligible for the prize because of his death in 1995).

10.1. **Black-Scholes via change of measure.** Our goal is to determine the price π_C at time $t = 0$ of a contingent claim C dependent on the entire path $(S_t)_{t\in[0,T]}$ with maturity at time $t = T$. In analogy to the general results obtained in Section 4 for the discrete time-setting, we suppose that

$$\pi_C = \mathbb{E}_{\mathbb{Q}}\left[\frac{C}{e^{rT}}\right],$$

where $\mathbb{Q}$ is an equivalent martingale measure, that is $\mathbb{Q} \approx \mathbb{P}$ and the discounted price process

$$X_t = e^{-rt} S_t = S_0 \exp\big(\sigma B_t + (\mu - r)t\big), \qquad 0 \le t \le T,$$

is a $\mathbb{Q}$-martingale with respect to the natural filtration $(\mathcal{F}_t)_{t\ge 0}$ generated by $(S_t)_{t\ge 0}$. How can we find such a measure $\mathbb{Q}$? First, recall that for a Brownian motion W the process $\exp(\lambda W_t - \frac{\lambda^2}{2} t)$ is a martingale for every $\lambda \in \mathbb{R}$ (see exercises). We define the measure $\mathbb{Q}$ by $\mathrm{d}\mathbb{Q} := \exp(cB_T - \frac{c^2}{2}T)\,\mathrm{d}\mathbb{P}$, which is short for

$$\mathbb{Q}[A] := \int_A \exp\big(cB_T - \tfrac{c^2}{2}T\big)\,\mathrm{d}\mathbb{P} = \mathbb{E}\Big[\exp\big(cB_T - \tfrac{c^2}{2}T\big)\mathbb{1}_A\Big],$$

for some $c \in \mathbb{R}$ to be chosen later. In particular, $\mathbb{Q}$ is equivalent to $\mathbb{P}$ since the density $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp(cB_T - \frac{c^2}{2}T) > 0$ $\mathbb{P}$-a.s. (see Theorem 1.23 (ii)).

By the Cameron-Martin theorem we have for any bounded and measurable functional $F : C([0,T]) \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\big[F\big(B_t : 0 \le t \le T\big)\big] &= \mathbb{E}\big[F\big(W_t : 0 \le t \le T\big)\exp\big(cB_T - \tfrac{c^2}{2}T\big)\big]\\ &= \mathbb{E}_{\mathbb{Q}}\big[F\big(W_t : 0 \le t \le T\big)\big],\end{aligned}$$

with $W_t := B_t - ct$, $t \in [0,T]$. (We apply here Theorem 9.1 on the functional $\tilde{F}(\omega) = F((\omega_t - ct)_{t\le T}$.) In particular, W is a Brownian motion under the measure

$\mathbb{Q}$. Now we choose c such that

$$\sigma c + \mu - r = -\frac{\sigma^2}{2} \quad \Longleftrightarrow \quad c = \frac{r - \frac{\sigma^2}{2} - \mu}{\sigma}.$$

Hence,

$$\begin{aligned} X_t = e^{-rt} S_t = S_0 \exp\big(\sigma B_t + (\mu - r)t\big) &= S_0 \exp\big(\sigma W_t + (\sigma c + \mu - r)t\big) \\ = S_0 \exp\big(\sigma W_t - \tfrac{\sigma^2}{2} t\big), \end{aligned}$$

which is a $\mathbb{Q}$-martingale. Thus, $\mathbb{Q}$ is an equivalent martingale measure, which can also be shown to be unique.

To summarize, under $\mathbb{Q}$ the price process $(S_t)_{t\in[0,T]}$ is of the form

$$S_t = S_0 \exp\Big(\sigma W_t + \big(r - \tfrac{1}{2}\sigma^2\big)t\Big),$$

where W is a $\mathbb{Q}$-Brownian motion. Note that for pricing of a contingent claim only the behaviour of the price process under the equivalent martingale measure is relevant.

Consider now, as an example, a European option of the form $C = f(S_T)$ with expiry $T > 0$ for any bounded, continuous function $f : [0, \infty) \to [0, \infty)$. Then the Black-Scholes price $\pi_C = e^{-rT}\, \mathbb{E}_{\mathbb{Q}}\big[f(S_T)\big]$ of C is given by

$$\pi_C = e^{-rT}\, \mathbb{E}_{\mathbb{Q}}\Big[f\Big(S_0 \exp\Big(\sigma W_T + \big(r - \tfrac{1}{2}\sigma^2\big)T\Big)\Big)\Big]. \tag{10.1}$$

Since $W_T \sim \mathcal{N}(0, T)$, so $W_T = \sqrt{T} Y$ with $Y \sim \mathcal{N}(0, 1)$, it follows that $\pi_C = v(T, S_0)$, where

$$v(t, x) := e^{-rt} \int_{-\infty}^{\infty} f\Big(x \exp\big(\sigma\sqrt{t} y + (r - \tfrac{1}{2}\sigma^2)t\big)\Big) \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \tag{10.2}$$

for $t \in [0, T]$, $x \in \mathbb{R}$. Of course, for this argument to work, the payoff function f does not have to be continuous or bounded. It suffices that $\mathbb{E}_{\mathbb{Q}}\big[f(S_T)\big] < \infty$, which is already guaranteed if f has polynomial volume growth, that is there exist $c > 0$ and $p \geq 0$ such that $f(x) \leq c\,\big(1 + x\big)^p$ for all $x \geq 0$. In particular, we can use formula (10.1) to compute the price of a call option with $f(x) = \big(x - K\big)^+$.

10.2. **The Black-Scholes Model as limit of the Binomial Model.** Black-Scholes models also arise as a natural limit of certain binomial models after a suitable scaling, meaning that the number of intermediate trading periods becomes large and their durations becomes small. This should not come as big surprise. In the CRR model the price process is a random walk and in the Black-Scholes model it is governed by a Brownian motion, which can obtained as scaling limit of random walks.

Throughout this section, T will *not* denote the number of trading periods in a fixed discrete-time market model but rather a physical *date*. We divide the interval $[0, T]$ into $N \cdot T$ equidistant time steps $\frac{1}{N}, \frac{2}{N}, \ldots, \frac{NT}{N}$. Then the i-th trading period corresponds to the 'real time interval' $(\frac{i-1}{N}, \frac{i}{N})$. Now consider a family of multi-period CRR-models, indexed by $N \in \mathbb{N}$, with parameters

$$r_N := \frac{r}{N}, \quad a_N := -\frac{\sigma}{\sqrt{N}}, \quad b_N := \frac{\sigma}{\sqrt{N}}, \quad p_N := \frac{1}{2} + \frac{1}{2}\frac{\mu}{\sigma\sqrt{N}},$$

where $r \geq 0$ is the instantaneous interest rate, $\mu \in \mathbb{R}$ a drift and $\sigma > 0$ a volatility parameter. We denote by $(S^0_{i,N})_{i=0,\dots,NT}$ the riskless bond and by $(S^1_{i,N})_{i=0,\dots,NT}$ the risky asset. The initial prices are assumed not to depend on N, i.e. $S^1_{0,N} = S^1_0$ for some constant $S^1_0 > 0$.

The question is whether the prices of contingent claims in the approximating market models converge as N tends to infinity. It will turn out that they do converge towards the Black-Scholes prices derived in the last section.

Theorem 10.1. *Let f be a continuous function on $[0,\infty)$ such that $|f(x)| \leq c(1+x)^q$ $c > 0$ and $q \in [0,2)$. Then the limit of the arbitrage-free prices of $C^{(N)} = f(S^1_{NT,N})$ for $N \to \infty$ is given by the Black-Scholes price $v(T, S^1_0)$, where as before*

$$v(T,x) := e^{-rT}\int_{-\infty}^{\infty} f\Big(x\exp\big(\sigma\sqrt{T}y + (r - \tfrac{1}{2}\sigma^2)T\big)\Big)\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy, \qquad x \in \mathbb{R}.$$

In other words,

$$\lim_{N\to\infty}\mathbb{E}_{\mathbb{Q}_N}\left[\frac{C^{(N)}}{(1+r_N)^{NT}}\right] = \mathbb{E}\Big[e^{-rT}f\Big(S^1_0\exp\big(\sigma\sqrt{T}W + (r-\tfrac{1}{2}\sigma^2)T\big)\Big)\Big],$$

where W is a $\mathcal{N}(0,1)$-distributed random variable.

Proof. See [6, Section 5.6] or [1, Section 7.2]. □

Hedging in the Black-Scholes model. In the Black-Scholes model consider an attainable contingent claim of the form $C = f(S_T)$ with replicating strategy (or *hedging strategy*) $\bar\theta = (\theta^0, \theta)$. Then the (discounted) value process of $\bar\theta$ is given by $V_t = v(T-t, S_t)$ with v defined in (10.2). Since the Black-Scholes model may be regarded as a limit of binomial models in the sense of Theorem 10.1, in view of the hedging strategy for the CRR model derived in Proposition 5.5, one can argue that the hedging strategy is given by

$$\theta_t(\omega) = \Delta\big(T-t, S_t(\omega)\big), \qquad \theta^0_t(\omega) = v(T-t, S_t(\omega)) - \theta_t(\omega)e^{-rt}S_t,$$

where

$$\Delta(t,x) := \frac{\partial}{\partial x}v(t,x), \qquad t \in [0,T],\ x \in \mathbb{R}.$$

In the financial language this is called '*Delta hedging*'.

10.3. **Black-Scholes pricing formula for European Calls and Puts.** We now derive an explicit formula for the Black-Scholes price of the European call option $C^{\mathrm{call}} = \big(S_T - K\big)^+$. For that purpose we simply choose $f(x) = (x-K)^+$ in (10.1) and (10.2), so

$$v(T,x) = e^{-rT}\,\mathbb{E}_{\mathbb{Q}}\Big[\Big(x \exp\Big(\sigma\sqrt{T}W + (r - \tfrac{1}{2}\sigma^2)T\Big) - K\Big)^+\Big],$$

where W is $\mathcal{N}(0,1)$-distributed under $\mathbb{Q}$. Substituting $\tilde{K} = e^{-rT}K/x$ and $\tilde{\sigma} = \sigma\sqrt{T}$ we get

$$\begin{aligned} v(T,x) &= \frac{x}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Big(e^{\tilde{\sigma}y - \frac{1}{2}\tilde{\sigma}^2} - \tilde{K}\Big)^+ e^{-y^2/2}\,\mathrm{d}y \\ &= \frac{x}{\sqrt{2\pi}} \int_{\frac{\log \tilde{K} + \frac{1}{2}\tilde{\sigma}^2}{\tilde{\sigma}}}^{\infty} \Big(e^{\tilde{\sigma}y - \frac{1}{2}\tilde{\sigma}^2} - \tilde{K}\Big) e^{-y^2/2}\,\mathrm{d}y \\ &= \frac{x}{\sqrt{2\pi}} \int_{-d_-}^{\infty} e^{-\frac{(y-\tilde{\sigma})^2}{2}}\,\mathrm{d}y - x\tilde{K}\,\big(1 - \Phi(-d_-)\big) \\ &= x\,\Phi(d_+) - Ke^{-rT}\Phi(d_-), \end{aligned}$$

where Φ denotes the distribution function of the standard normal distribution,

$$d_- = d_-(T,x) := \frac{\log(\frac{x}{K}) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = -\frac{\log \tilde{K} + \frac{1}{2}\tilde{\sigma}^2}{\tilde{\sigma}}$$

and

$$d_+ = d_+(T,x) := d_-(T,x) + \sigma\sqrt{T} = \frac{\log(\frac{x}{K}) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}.$$

To summarize, the Black-Scholes price for a European call option with strike price K is given by $v(T, S_0)$ where

$$v(T,x) = x\,\Phi(d_+) - Ke^{-rT}\Phi(d_-) \tag{10.3}$$

with

$$d_\pm = d_\pm(T,x) := \frac{\log(\frac{x}{K}) + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}},$$

and for any $t \in [0,T]$ the value of the option at time t is given by $v(T-t, S_t)$.

Now we turn to pricing European put options $C^{\mathrm{put}} = \big(K - S_T\big)^+$. Since we have computed already the price of the corresponding call option we can use the so-called *put-call parity*, which refers to the fact that

$$C^{\mathrm{call}} - C^{\mathrm{put}} = \big(S_T - K\big)^+ - \big(K - S_T\big)^+ = S_T - K,$$

and the right-hand side equals the pay-off of a forward contract (cf. Section 0) with price $S_0 - e^{-rT}K$ (note that the contingent claim $C = S_T$ can be trivially replicated just by holding one unit of the risky asset which requires an initial investment S_0).

Hence, the price $\pi(C^{\text{put}})$ for C^{put} can be obtained from the price $\pi(C^{\text{call}})$ for the call C^{call}, namely

$$\begin{aligned}\pi(C^{\text{put}}) &= \pi(C^{\text{call}}) - \big(S_0 - e^{-rT}K\big)\\ &= S_0\,\Phi(d_+) - Ke^{-rT}\Phi(d_-) - S_0 + e^{-rT}K\\ &= e^{-rT}K\Phi(-d_-) - S_0\Phi(-d_+).\end{aligned}$$

We can also determine a hedging strategy $\bar\theta$ for the call option if we use the 'Delta hedging' discussed at the end of Section 10.2, which gives $\theta_t = \Delta(T-t, S_t)$. The *Delta* of the call option $C^{\text{call}} = \big(S_T - K\big)^+$ can be computed by differentiating the value function in (10.3) with respect to x,

$$\Delta(t,x) := \frac{\partial}{\partial x} v(t,x) = \Phi\big(d_+(t,x)\big).$$

In particular, note that $\theta_t \in (0,1)$ a.s. ('long in stock').

The *Gamma* of the call option is given by

$$\Gamma(t,x) := \frac{\partial}{\partial x}\Delta(t,x) = \frac{\partial^2}{\partial x^2} v(t,x) = \varphi\big(d_+(t,x)\big)\frac{1}{x\sigma\sqrt{t}},$$

where $\varphi = \Phi'$ denotes the density of the standard normal distribution. Large Gamma values occur in regions where the Delta changes rapidly, corresponding to the need for frequent readjustments of the Delta hedging portfolio. It follows that $x \mapsto v(t,x)$ is strictly convex.

Another important parameter is the *Theta*

$$\Theta(t,x) := \frac{\partial}{\partial t} v(t,x) = \frac{x\sigma}{2\sqrt{t}}\varphi(d_+(t,x)) + Kre^{-rt}\Phi\big(d_-(t,x)\big).$$

The fact that $\Theta > 0$ corresponds to the observation that arbitrage-free prices of European call options are typically increasing functions of the maturity. Note that the parameters Δ, Γ and Θ are related by the equation

$$r\,v(t,x) = rx\,\Delta(t,x) + \frac{1}{2}\sigma^2 x^2\,\Gamma(t,x) - \Theta(t,x).$$

Thus, the function v solves the following partial differential equation, often called the *Black-Scholes equation*

$$rv = rx\frac{\partial v}{\partial x} + \frac{1}{2}\sigma^2 x^2\frac{\partial^2 v}{\partial x^2} - \frac{\partial v}{\partial t}.$$

Recall that the Black-Scholes price $v(T, S_0)$ was obtained as the expectation of the discounted payoff $e^{-rT}(S_T - K)^+$ under the equivalent martingale measure $\mathbb{Q}$. Thus, at first glance it may come as surprise that the *Rho* of the option

$$\varrho(t,x) := \frac{\partial}{\partial r} v(t,x) = Kte^{-rt}\Phi\big(d_-(t,x)\big)$$

is strictly positive, i.e. the price is increasing in r. Note, however, that the martingale measure $\mathbb{Q}$ depends itself on the interest rate r.

The parameter σ is called the *volatility* of the model and may be regarded as a measure of the fluctuations in the stock price process. The price of a European call option is an increasing function of the volatility as the *Vega* of the option

$$\mathcal{V}(t,x) := \frac{\partial}{\partial\sigma} v(t,x) = x\sqrt{t}\,\varphi\big(d_+(x,t)\big)$$

is strictly positive. The functions $\Delta, \Gamma, \Theta, \varrho$ and $\mathcal{V}$ are usually called the *Greeks* (although 'vega' is not a letter in the Greek alphabet). We refer to [6, Section 5.6] for more details and some nice plots of the Greeks.

Remark 10.2 (Implied volatility). In practice, the prices for European call and put options are known as they can be directly observed in the market, but the volatility parameter σ is unknown. Since the Vega is strictly positive the function $\sigma \mapsto v(t,x)$ is injective, and by inverting this function one can deduce a value for σ from the observed market prices, the so-called *implied volatility*.

References

[1] S. Andres. Stochastic Financial Models. *Lecture notes, University of Cambridge*, 2019.

[2] H. Bauer. *Measure and integration theory*, volume 26 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2001. Translated from the German by Robert B. Burckel.

[3] A. Bovier. Stochastic Processes. *Lecture Notes, University of Bonn*, 2013.

[4] J. Cox, S. Ross, and M. Rubinstein. Option pricing: a simplified approach. *J. Financial Econom.*, 7:229–263, 1979.

[5] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520, 1994.

[6] H. Föllmer and A. Schied. *Stochastic finance*. Walter de Gruyter & Co., Berlin, extended edition, 2011. An introduction in discrete time.

[7] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[8] J. Norris. Advanced Probability. *Lecture notes, University of Cambridge*, 2016.

University of Manchester
Current address: Department of Mathematics, Oxford Road, Manchester M13 9PL
E-mail address: sebastian.andres@manchester.ac.uk