Mathematical Finance

An introduction in discrete time

Winter Term 2023/24
CAU Kiel

Sören Christensen

This Version: October 6, 2023

Preface

This material serves as the lecture notes for the course on Mathematical Finance at Kiel University in the winter term 2023-24.

The content is mostly taken from Jan Kallsen's lecture notes (see `http://www.math.uni-kiel.de/finmath/de/personen/kallsen/lec_notes`), which have been extended and reorganized by Mathias Vetter and me. It basically covers all the material needed for the course. We will not follow a single textbook, although there are of course many suitable references; for example Föllmer and Schied (2011); Pliska (1997); De Saporta and Zili (2021); Calogero (2023), or also Bäuerle and Rieder (2017); Irle (2003), to name two books in German.

A characteristic of this course is that the prior knowledge in both mathematics and finance is very heterogeneous in the classroom. We try to take this into account by splitting the group for parts of the lecture. Therefore, most chapters of these lecture notes contain appendices for *QF* (Quantitative Finance major, labeled with A) and *MF* (Mathematics and Financial Mathematics major, labeled with B).

Let us furthermore give a small remark concerning the tables and figures that illustrate various concepts and results. If numbers appear in green, this indicates that they have or can be computed from the other (black) numbers, which are given in the first place. In other words: black numbers are model input, green ones follow from them using the definitions and tools in the text.

Contents

Chapter 1

What is this course about?

In this chapter we explain the real world problems considered in this course as well as some basic ideas for their mathematical treatment.

Fundamental Notions

Most of this course is about pricing of derivatives on the financial market. Here, a *derivative* is a financial contract whose value at an expiration date T is determined by the price of some underlying financial assets.

We consider some examples.

1. A *European call option* gives the owner the right to buy a fixed quantity of a fixed asset at a fixed time (*maturity*) and for a fixed price (*strike*). There is no obligation to buy the asset.
   The *European put option* corresponds to the call but the owner has the right to sell rather than buy the asset. Otherwise, the above comments hold for the put as well.

2. The owner of an *American call* resp. *American put* can exercise her option any time before maturity. She does not necessarily have to wait till maturity. In contrast to the European option, this involves a true choice.

3. In a *forward contract* one agrees to buy a fixed quantity of a fixed asset at a fixed date (*maturity*) and a fixed price (*forward price*). In contrast to a call option this involves an obligation.

4. Forwards are usually contracted directly by two parties (*over the counter*). The exchange-traded version is called *futures contract*. A future is different in the sense that payments are constantly made as long as the contract is running. On the other hand, one can enter and terminate the future free of charge at any time.

These contracts may be signed in order to reduce the own exposure to price fluctuations. However, for the counterparty they involve a potentially larger exposure to risks. Consequently, there are two key issues in mathematical finance which we constantly address in this course:

- What is a *fair* price for an option?
- Can the risk caused e.g. by selling derivatives be reduced or *hedged* by trading the underlying securities skillfully?

Arbitrage Pricing Technique and the Principle of No-Arbitrage

The main concept to answer the questions raised above is the notion of *arbitrage*. An arbitrage is a riskless profit by a transaction on financial markets. An *arbitrageur* is a trader looking for such riskless profit (by simultaneously trading on various markets).

Example 1.1. *Consider, for example, a share traded for $100 in New York and for 90€ in Frankfurt, where we assume the exchange rate to be 0.91€ per US-dollar. Now, the following trading strategy is an arbitrage:*

- *Buy 1000 shares in Frankfurt,*
- *sell 1000 shares in NY,*
- *change $ to €.*

Without transaction costs, the riskless profit is

$$1000 \times (100 \times 0.91 - 90)€ = 1000€.$$

Such arbitrage opportunities can only pertain for a very small amount of time (share price in NY will go down, price in Frankfurt will go up). Therefore, the following principle sounds reasonable:

Principle of No-Arbitrage (PNA): On a reasonable financial market, it should not be possible to guarantee a profit without exposure to risk.

Acting according to the PNA has an obvious consequence: The pricing of a new derivative must not bring arbitrage opportunities into a market. This sounds obvious, but it is the main guiding principle for all derivative pricing. It is connected to one of the main principles in economic theory – the *principle of rational markets*.

Example 1.2 (Price of a forward contract)**.** *Consider a forward contract with maturity T on an asset with price S_0 at time 0 and dividend payment D at time t_0 and interest rate r. The question is how to agree at time 0 on a forward price F for which the asset will change hands at time T?*

The answer is that

$$F = (S_0 - De^{-rt_0})e^{rT}$$

is the only rational price based on the PNA. The argument is as follows:

- *Assume first by contradiction that $F > (S_0 - De^{-rt_0})e^{rT}$. Then consider the following strategy:*

- *at $t = 0$, borrow S_0 and buy the asset,*
- *at $t = t_0$, use the dividends for partial repayment,*
- *at $t = T$, sell the asset for F and repay debt.*

This leads to the arbitrage

$$G = F - (S_0 e^{rt_0} - D)e^{r(T-t_0)} = F - (S_0 - De^{-rt_0})e^{rT} > 0.$$

- *Similarly, if $F < (S_0 - De^{-rt_0})e^{rT}$. Then consider the following strategy:*
  - *at $t = 0$, borrow an asset (with the obligation to pay dividend in t_0) and sell asset in S_0, put S_0 into the bank account.*
  - *at $t = t_0$ pay dividend,*
  - *at $t = T$, buy the asset for F.*

Then return

$$G = (S_0 e^{rt_0} - D)e^{r(T-t_0)} - F = (S_0 - D^{-rt_0})e^{rT} - F > 0.$$

(Note that this strategy involves short selling *(i.e. borrowing assets) which is not always possible in practice.)*

Example 1.3 (Law of One Price)**.** *We start with a heuristic argument for the* Law of One Price *based on the PNA: If two different combinations of assets have the same value $V = W$ at a future time T, then they must have equal value $V_0 = W_0$ at present time:*
The argument is similar as before:

- *Assume first by contradiction that $V_0 > W_0$. Then consider the following strategy (Buy low, sell high):*
  - *at $t = 0$, short sell the first combination, i.e. borrow the first combination and sell at price V_0; furthermore, buy second combination for W_0 and put $V_0 - W_0$ into the bank account (with interest rate r)*
  - *at $t = T$, sell second combination for W, buy first combination for $V_T = W_T$ and return $G = (V_0 - W_0)e^{rT}$ riskless profit*

  This leads to the arbitrage

  $$G = (V_0 - W_0)e^{rT}.$$

- *Similarly, $V_0 < W_0$ leads to arbitrage (just interchange the roles of the objects).*

Example 1.4 (Call-Put Parity)**.** *The Law of One Price seems very obvious, but it has interesting consequences. We now illustrate this with the* Call-Put Parity*:*
Let us consider Call and Put options on an asset with identical strike K and maturity T. How are the prices C_0 and P_0 related?

If the (random) asset price in T is called S^1_T, then

- *the call at T is worth $C = (S^1_T - K)^+ := \max(S^1_T - K, 0)$*[1]*,*
- *the put at T is worth $P = (K - S^1_T)^+ = \max(K - S^1_T, 0)$.*

Therefore, the combination of one asset and one put is worth at time T

$$V := S^1_T + (K - S^1_T)^+ = \max(S^1_T, K).$$

On the other hand, the combination of one call and Ke^{-rT} money in the bank account at time $t = 0$ is worth at time T

$$W := (S^1_T - K)^+ + Ke^{-rT}e^{rT} = \max(S^1_T, K).$$

Hence, $V = W$ and the law of one price implies

$$S^1_0 + P_0 = V_0 = W_0 = C_0 + Ke^{-rT}.$$

This identity is called Call-Put Parity. *So, if one of the prices C_0, P_0 is known, the other can immediately be found.*

Example 1.5. *To find prices of, say, European options explicitly using the PNA, more structure has to be assumed. This leads to probabilistic models. The details will be introduced in the following chapter. Here, we however discuss some basic ideas in an obviously oversimplified situation:*
We consider a call option with strike $K = 130$ on a risky asset with today's price $S^1_0 = 100$ and maturity $T = 1$. We now specify a market model by assuming that the risky asset will either go up to 150 or down to 90 with probabilities p and $1 - p$, resp., see Figure 1.1. Hence, the option is worth either $(150-130)^+ = 20$ or $(90-130)^+ = 0$ at time $T = 1$. Let us assume that in our market, there is also a riskless asset (bank account) with interest rate $r = 0$. We now try to replicate the option's payoff by investing φ_0 into the bond and buying φ_1 risky assets at $T = 0$. Then, at time $T = 1$ the value of the portfolio is

$$\varphi_0 + \varphi_1 S^1_1 = \begin{cases} \varphi_0 + \varphi_1 150, & S^1_1 = 150, \\ \varphi_0 + \varphi_1 90, & S^1_1 = 90. \end{cases}$$

To replicate the portfolio, we have to solve the system of linear equations

$$\varphi_0 + \varphi_1 150 = 20,\ \varphi_0 + \varphi_1 90 = 0,$$

yielding $\varphi_0 = -30$, $\varphi_1 = 1/3$. That is, we borrow 30 Euro from the bank and buy 1/3 of a stock. Then, our portfolio has the same payoff profile as the call option we are interested in. As the portfolio costs

$$\varphi_0 \cdot 1 + \varphi_1 \cdot 100 = 10/3$$

[1] If the market price of the asset at maturity is below the strike, it does not make sense to exercise the option, which is hence worthless. If, on the other hand, the market price settles above the strike, the value of the option is the difference of the two. Indeed, even if the option holder is not interested in the asset, he can realize the profit by selling it immediately at the market price. Hence, the value of a call option written on one share of security S^1 with maturity T and strike K amounts to

$$(S^1_T - K)^+ = \max(S^1_T - K, 0) \tag{1.1}$$

at time T.

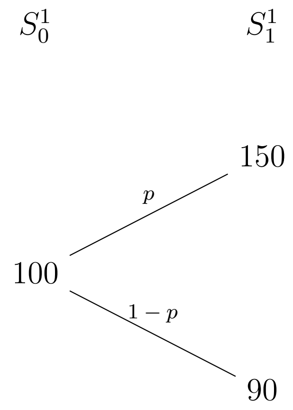

Figure 1.1: Asset price S^1

at $T = 0$, the law of one price yields that we should also accept this as the price for the option.

It is perhaps surprising to observe that the fair price does not depend on the transition probability p. Can you explain this?

Part I

Mathematical Basics

Chapter 2

Preliminaries: Some notions from probability theory

In this chapter we provide the mathematical basics needed in this course.

We have already stated some interesting results based on the PNA in the previous chapter. These results, like the Call-Put Parity or the Law of One Price hold always, no matter how the underlying asset price dynamics is. Formally, they are *model-free*. For example, if one of the prices C_0, P_0 is known in the setting of Example 1.4, the other can immediately be found. Therefore, the question is how to find, say, C_0. This cannot been done based on such general arguments. We need a mathematical model for how the asset prices evolve over time as illustrated in Example 1.5 in an oversimplified setting. The foundation for considering more realistic models are the notions from probability theory, which we introduce in this chapter. As the prior knowledge is very heterogeneous, we split this chapter between the groups.

Appendix QF

2.A Basics

This course requires a decent background in stochastics which cannot be provided here. We only recall a few indispensable notions and the corresponding notation and refer to standard textbooks on probability for more details.

Random experiments are modeled in terms of *probability spaces* $(\Omega, \mathfrak{P}(\Omega), P)$. The *sample space* Ω represents the set of all possible outcomes of the experiment, e.g. $\Omega = \{1,2,3,4,5,6\}$ if you throw a dice. We consider sample spaces with finitely many elements in this course. These outcomes happen with certain probabilities, which are quantified by a *probability measure* P. For mathematical reasons, probabilities are not assigned to elements $\omega \in \Omega$ (i.e. the outcomes $1,\dots,6$ in our example) but to subsets $A \subset \Omega$, which are called *events*. The *power set* $\mathfrak{P}(\Omega)$ is the set of all subsets $A \subset \Omega$, i.e. the set of all events in stochastic language. In our example, we have

$$\mathfrak{P}(\Omega) = \Big\{\varnothing, \{1\}, \dots, \{6\}, \{1,2\}, \{1,3\}, \dots, \{1,2,3,4,5,6\}\Big\}.$$

A *probability measure* P assigns probabilities to any of these events. More precisely, it is a mapping $P : \mathfrak{P}(\Omega) \to [0,1]$, which is *normalized* and *additive*. This means

1. $P(\Omega) = 1$,
2. $P(A \cup B) = P(A) + P(B)$ for any disjoint sets $A, B \subset \Omega$ (i.e. $A \cap B = \varnothing$).

As a side remark, additivity is replaced by the slightly stronger requirement of *σ-additivity* for infinite sample spaces.

Very often, probability measures are defined in terms of a *probability mass function*, i.e. a function $\varrho : \Omega \to [0,1]$ with $\sum_{\omega\in\Omega} \varrho(\omega) = 1$. The number $\varrho(\omega)$ stands for the probability of the single outcome ω. The probability measure corresponding to ϱ is given by

$$P(A) := \sum_{\omega \in A} \varrho(\omega).$$

In our example we would assume all outcomes to be equally likely, i.e. $\varrho(\omega) = 1/6$ for $\omega = 1,\dots,6$. This leads to

$$P(A) = \sum_{\omega\in A} \frac{1}{6} = \frac{|A|}{6}$$

for any $A \subset \Omega$. $|A|$ denotes the cardinality of a set A, i.e. the number of elements in A.

2.A.1 Random variables

Often we are not so much interested in the particular outcome of the random experiment but of a quantitative aspect of it. This is formalized in terms of a *random variable*, which is a function $X : \Omega \to \mathbb{R}$ assigning a real number $X(\omega)$ to any particular outcome ω. Instead of $\mathbb{R}$ we sometimes consider $\mathbb{R}^n$, in which case X is a vector-valued random variable. The *expected value* $E(X)$ of a random variable is its mean if the values are weighted by the probabilities of the outcomes, specifically

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

This can also be written as

$$E(X) = \sum_{x\in\mathbb{R}} xP(X = x),$$

where $P(X = x)$ is an abbreviation for $P(\{\omega \in \Omega : X(\omega) = x\})$ and the sum actually extends only to the finitely many values of X. We also recall the *variance*

$$\mathrm{Var}(X) := E\Big((X - E(X))^2\Big) = E(X^2) - (E(X))^2.$$

If we consider $X(\omega) = \omega$ in our example of throwing a dice, we have of course

$$E(X) = \frac{1 + \cdots + 6}{6} = \frac{7}{2} = 3.5$$

and

$$\mathrm{Var}(X) = \frac{1^2 + \cdots + 6^2}{6} - (E(X))^2 = \frac{35}{12} = 2.91\overline{6}.$$

An important random variable is the *indicator* of a set $A \subset \Omega$, which is defined as

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

Here, $A^C := \Omega \setminus A$ denotes the *complement* of a set A. The expectation of an indicator is the probability of the set: $E(1_A) = P(A)$ for $A \subset \Omega$. Moreover, we sometimes write

$$\int XdP := \int X(\omega)P(d\omega) := E(X) = \sum_{\omega\in\Omega} X(\omega)P(\{\omega\})$$

and

$$\int_A XdP := \int_A X(\omega)P(d\omega) := E(X1_A) = \sum_{\omega\in A} X(\omega)P(\{\omega\}),$$

which is motivated by the fact that the expected value in general – not necessarily finite – probability spaces is defined in terms of a (Lebesgue) integral.

2.A.2 Conditional probabilities and expectations

Suppose that we receive partial information about the outcome of a random experiment, namely that the outcome of our experiment belongs the event $B \subset \Omega$. We may e.g. know

ω	$P(\{\omega\})$	$P(\{\omega\}\vert B)$
1	1/6	0
2	1/6	1/3
3	1/6	0
4	1/6	1/3
5	1/6	0
6	1/6	1/3

Table 2.1: Conditional probabilities given the event $B := \{2, 4, 6\}$

that throwing a dice produced an even number, which means that the event $B := \{2, 4, 6\}$ has happened.

This partial knowledge changes our assessment. We know that outcomes in B^C are impossible whereas those in B are more likely than without the additional information. More precisely, the *conditional probability* of an event A *given* B is defined as

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

unless $P(B) = 0$. It is easy to see that $P(A|B)$ is again a probability measure if we consider it as a function of A with B being fixed. Cf. Table 2.1 for an illustration in the above example.

Two sets A, B are called (stochastically) *independent* if $P(A|B) = P(A)$, i.e. knowing that B has happened does not make A more or less probable. Independence is equivalent to $P(A \cap B) = P(A)P(B)$ which makes sense if $P(B) = 0$ as well. Moreover, one can see that independence is a symmetric concept. Random variables X, Y are called (stochastically) *independent* if $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any two sets A, B. This implies that $E(XY) = E(X)E(Y)$.

If probabilities change, expectations of random variables X change as well. The *conditional expectation of* X *given* B is just the ordinary expected value if we replace the original probabilities by conditional ones, i.e.

$$E(X|B) := \sum_{\omega \in B} X(\omega)P(\{\omega\}|B) = \frac{\sum_{\omega \in B} X(\omega)P(\{\omega\})}{P(B)} = \frac{E(X1_B)}{P(B)}.$$

In our example, the expectation of the outcome $X(\omega) = \omega$ changes from $E(X) = 3.5$ to

$$E(X|B) = \frac{2 + 4 + 6}{3} = 4$$

if we know the result to be even.

2.A.3 σ-fields

In this course we often consider subsets of the set $\mathfrak{P}(\Omega)$ of all events, which are called *algebras*. Specifically, $\mathscr{F} \subset \mathfrak{P}(\Omega)$ is called an *algebra* if

1. $\Omega \in \mathscr{F}$,

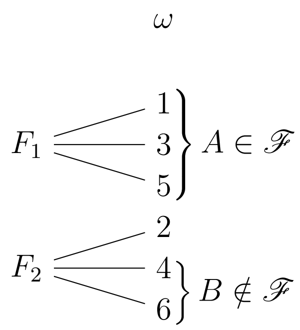

Figure 2.1: The partition generating a σ-field

2. $A^C \in \mathscr{F}$ for any $A \in \mathscr{F}$ (where $A^C := \Omega \setminus A$),
3. $A \cup B \in \mathscr{F}$ for any $A, B \in \mathscr{F}$.

This means that unions, intersections, complements etc. of sets in $\mathscr{F}$ always lead to sets in $\mathscr{F}$. For infinite sample spaces one replaces Axiom 3 by a slightly stronger requirement, which leads to the notion of a *σ-algebra* or *σ-field*. In our setup of finite Ω this amounts to the same thing. Nevertheless, we use from now on the terminology *σ-field* instead of algebra because it is more common in the literature. It is not hard to show the following result.

Lemma 2.A.1. *For any σ-field $\mathscr{F}$ there is a partition $F_1, \dots, F_n$ of Ω (i.e. $F_1, \dots, F_n \subset \Omega$ with $F_i \cap F_j = \varnothing$ for $i \neq j$ and $F_1 \cup \dots \cup F_n = \Omega$) such that $\mathscr{F}$ consists of all unions of the F_i, i.e.*

$$\mathscr{F} = \Big\{\varnothing, F_1, \dots, F_n, F_1 \cup F_2, F_1 \cup F_3, \dots, F_1 \cup \dots \cup F_n\Big\}.$$

We call $F_1, \dots, F_n$ the partition that generates *$\mathscr{F}$ and we write $\mathscr{F} = \sigma(F_1, \dots, F_n)$.*

Consequently, we can identify a σ-field with a partition of the sample space.

A σ-field stands for partial information. Suppose that we do not know the exact outcome ω of a random experiment. We only know which of the sets $F_1, \dots, F_n$ the outcome ω belongs to, e.g. we only know whether an even or odd number has appeared on our dice. This partial information is represented by the σ-field which is generated from $F_1 := \{2, 4, 6\}$ and $F_2 := \{1, 3, 5\}$, cf. Figure 2.1. So we can tell whether $A := \{1, 3, 5\}$ has happened but not whether the outcome is in $B := \{4, 6\}$ or not.

The smallest σ-field $\mathscr{F} = \{\varnothing, \Omega\}$ is generated by the partition that consists only of Ω. This *trivial σ-field* stands for absence of any information. The other extreme is the power set $\mathscr{F} = \mathfrak{P}(\Omega)$, which is generated by the finest partition $\{\omega_1\}, \dots, \{\omega_n\}$ of $\Omega = \{\omega_1, \dots, \omega_n\}$. It stands for complete information.

We also need a notion for the fact that a random variable depends only on the information given by a σ-field $\mathscr{F}$.

Definition 2.A.2. A random variable X is called *$\mathscr{F}$-measurable* if X is constant on the sets of the partition that generates $\mathscr{F}$. In other words, it is of the form

$$X(\omega) = \sum_{i=1}^{n} x_i 1_{F_i}(\omega), \tag{2.1}$$

	ω	$X(\omega)$	$Y(\omega)$
	1	1	0
F_1	3	1	0
	5	1	0
	2	0	0
F_2	4	0	1
	6	0	1

Figure 2.2: An $\mathscr{F}$-measurable X and a not $\mathscr{F}$-measurable random variable Y

ω	$X(\omega)$	$\sigma(X)$
1	-1	$F_1 = \{X = -1\}$
2	-1	
3	0	$F_2 = \{X = 0\}$
4	0	
5	1	$F_3 = \{X = 1\}$
6	1	

Figure 2.3: The σ-field resp. partition generated by a random variable X

where $F_1, \dots F_n$ is the partition that generates $\mathscr{F}$ and x_i is the value of X on F_i.

This is illustrated in Figure 2.2 where $\mathscr{F}$ is the σ-field from above. X is $\mathscr{F}$-measurable because it is determined by the information whether the outcome is even or odd. This is not the case for Y which is not constant on F_2 and hence not $\mathscr{F}$-measurable.

$\mathscr{F}$-measurability means that a random variable may not be "as random" as an arbitrary random variable. In the extreme case of the trivial σ-field $\mathscr{F} = \{\varnothing, \Omega\}$, any $\mathscr{F}$-measurable random variable is *deterministic*, i.e. constant. On the other hand, any random variable is $\mathscr{F}$-measurable for the power set $\mathscr{F} = \mathfrak{P}(\Omega)$.

We can also turn things around and wonder what information – in the sense of a σ-field – is delivered by a random variable. This σ-field is denoted as $\sigma(X)$ and it is generated by the partition $\{X = x_1\}, \dots, \{X = x_n\}$ of X, where $x_1, \dots, x_n$ are the values of X and we use the shorthand notation $\{X = x_1\} := \{\omega \in \Omega : X(\omega) = x_1\}$ etc. $\sigma(X)$ is called the *σ-field generated by* X. It is the smallest σ-field such that X is $\mathscr{F}$-measurable. For an illustration cf. Figure 2.3.

2.A.4 Conditional expectation relative to a σ-field

Similarly as in Section 2.A.2 our assessment of probabilities and expectations change if we receive partial information. Here, our information comes from a σ-field $\mathscr{F}$. In other words, it will be revealed which of the sets $F_1, \dots, F_n$ of the generating partition the outcome ω

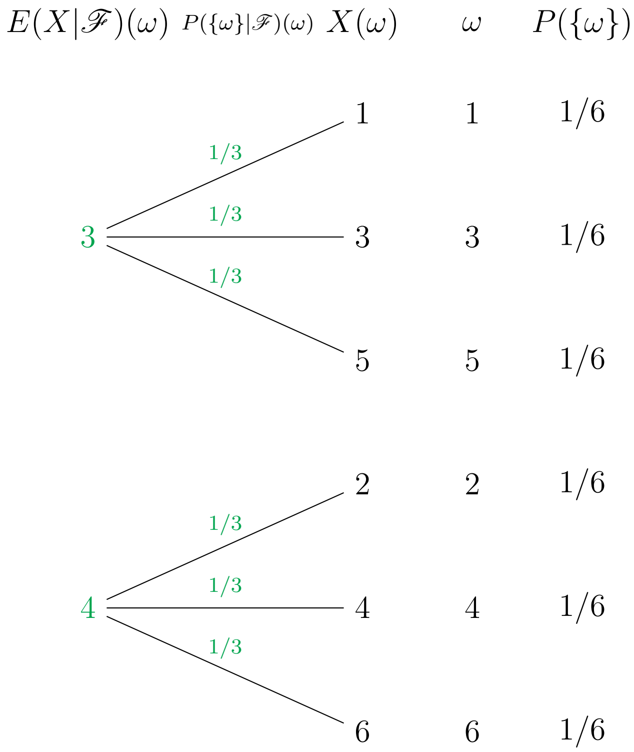

Figure 2.4: Conditional probabilities and conditional expectation

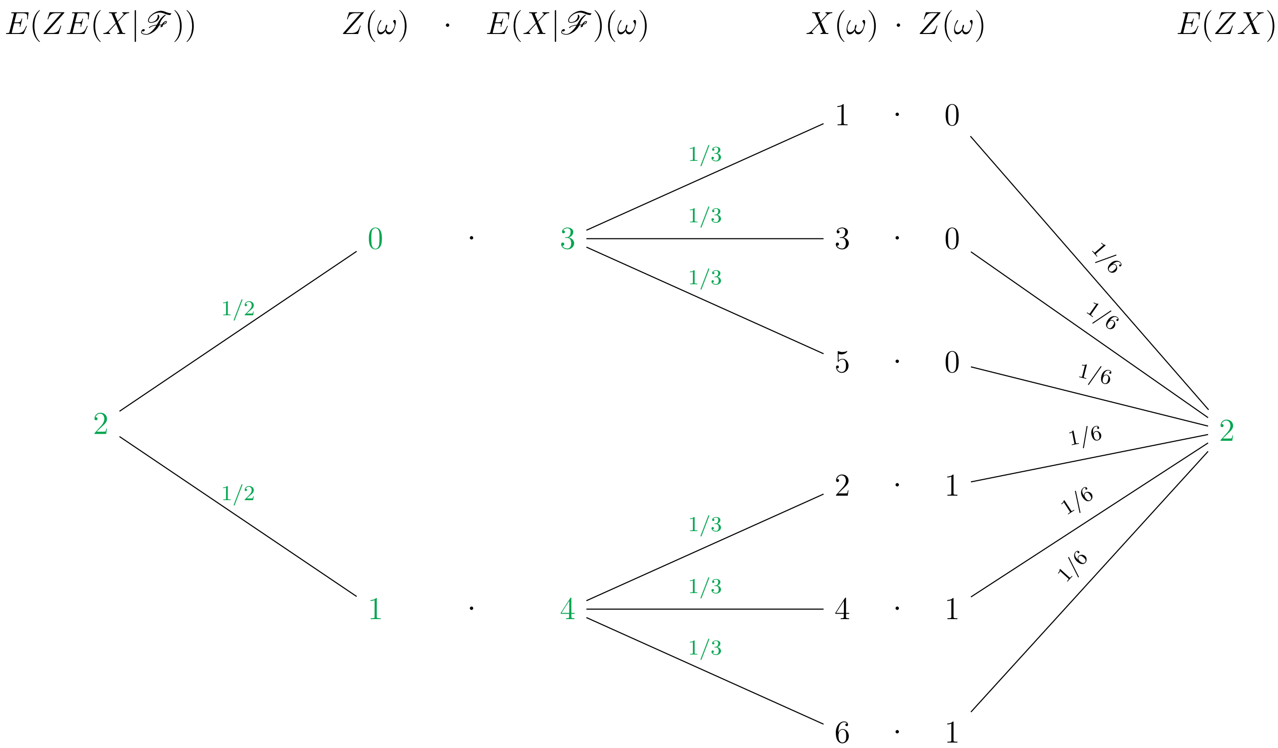

Figure 2.5: Illustration of Lemma 2.A.3

belongs to. The *conditional probability* of an event $A \subset \Omega$ *given* $\mathscr{F}$ is defined as

$$P(A|\mathscr{F})(\omega) := P(A|F_i) = \frac{P(A \cap F_i)}{P(F_i)} \text{ if } \omega \in F_i,$$

i.e. it is the conditional probability of Section 2.A.2 given the particular event F_i that has happened. Similarly, we define the *conditional expectation* of a random variable X *given* $\mathscr{F}$ as

$$E(X|\mathscr{F})(\omega) := E(X|F_i) = \sum_{\omega \in F_i} X(\omega) P(\{\omega\}|F_i) \text{ if } \omega \in F_i, \tag{2.2}$$

Note that conditional probabilities and expectations given $\mathscr{F}$ are random variables because they depend on ω. But they are not "as random" as an arbitrary random variable because they are $\mathscr{F}$-measurable, i.e. they depend not directly on ω but only on which of the sets $F_1, \dots, F_n$ has happened.

Here is an important property of conditional expectation:

Lemma 2.A.3. *The random variable $E(X|\mathscr{F})$ is $\mathscr{F}$-measurable. Moreover, it satisfies*

$$E\Big(ZE(X|\mathscr{F})\Big) = E(ZX)$$

for any $\mathscr{F}$-measurable random variable Z.

The previous lemma is illustrated in Figure 2.5. In general infinite probability spaces, the definition in (2.2) does not make sense. Then the properties in Lemma 2.A.3 are used to define conditional expectations.

The conditional expectation $E(X|\mathscr{F})$ can be interpreted as a best prediction or approximation of X given the partial information $\mathscr{F}$. If $\mathscr{F}$ is trivial (no information), then $E(X|\mathscr{F}) = E(X)$. If, on the other hand, $\mathscr{F} = \mathfrak{P}(\Omega)$ (full information), we have $E(X|\mathscr{F}) = X$. For general $\mathscr{F}$, the conditional expectation $E(X|\mathscr{F})$ somehow interpolates between X and $E(X)$.

Let us state some useful rules.

Lemma 2.A.4.

1. *If X is $\mathscr{F}$-measurable, then $E(X|\mathscr{F}) = X$. This holds in particular for $\mathscr{F} = \mathfrak{P}(\Omega)$.*
2. *If X and $\mathscr{F}$ are independent (which means $P(\{X \in B\} \cap F) = P(X \in B)P(F)$ for any sets B and $F \in \mathscr{F}$), then $E(X|\mathscr{F}) = E(X)$. This holds in particular for $\mathscr{F} = \{\varnothing, \Omega\}$.*
3. *$E(E(X|\mathscr{F})) = E(X)$*
4. *$E(E(X|\mathscr{F})|\mathscr{G}) = E(X|\mathscr{G}) = E(E(X|\mathscr{G})|\mathscr{F})$ if $\mathscr{G}$ is a* sub-σ-field *of $\mathscr{F}$ (i.e. $\mathscr{G}$ is a σ-field with $\mathscr{G} \subset \mathscr{F}$).*
5. *$E(X|\mathscr{F})$ is linear in X, i.e. $E(X+Y|\mathscr{F}) = E(X|\mathscr{F}) + E(Y|\mathscr{F})$ and $E(cX|\mathscr{F}) = cE(X|\mathscr{F})$.*
6. *$E(X|\mathscr{F})$ is increasing in X, i.e. $E(X|\mathscr{F}) \le E(Y|\mathscr{F})$ if $X \le Y$.*
7. *If $X_n \to X$ for $n \to \infty$, then $E(X_n|\mathscr{F}) \to E(X|\mathscr{F})$.*

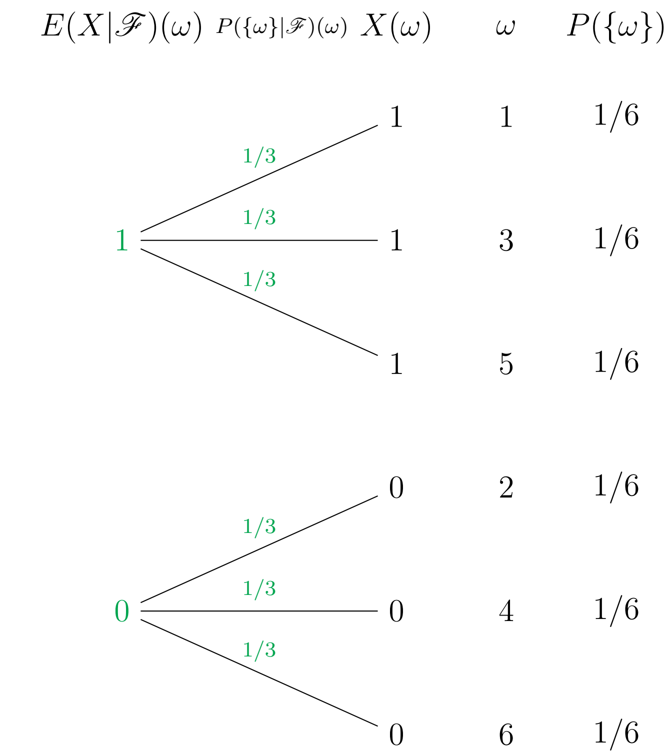

Figure 2.6: $\mathscr{F}$-measurable X

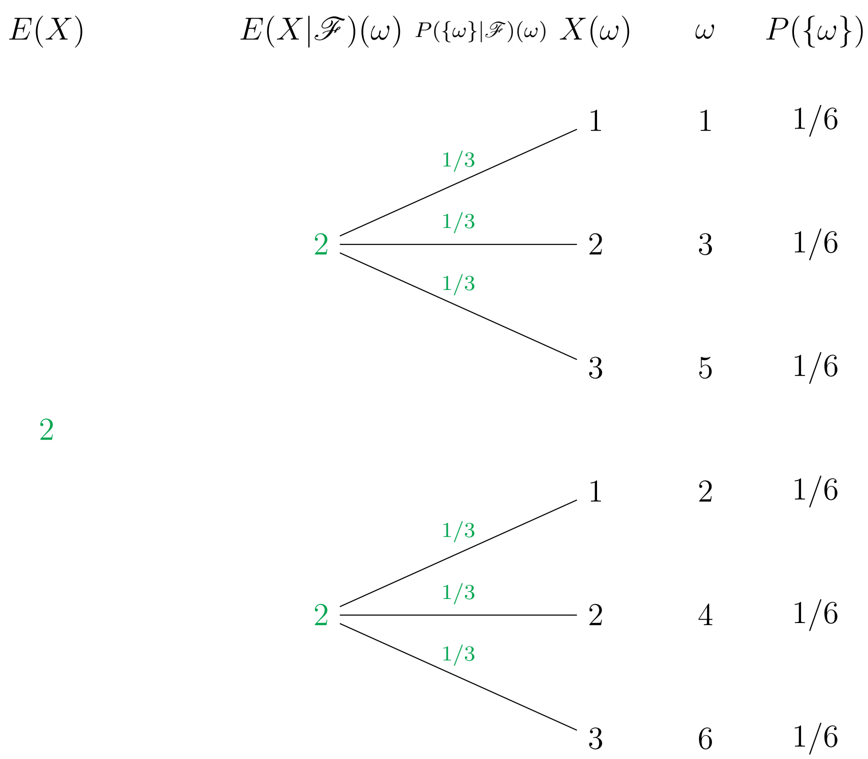

Figure 2.7: Independence of X and $\mathscr{F}$

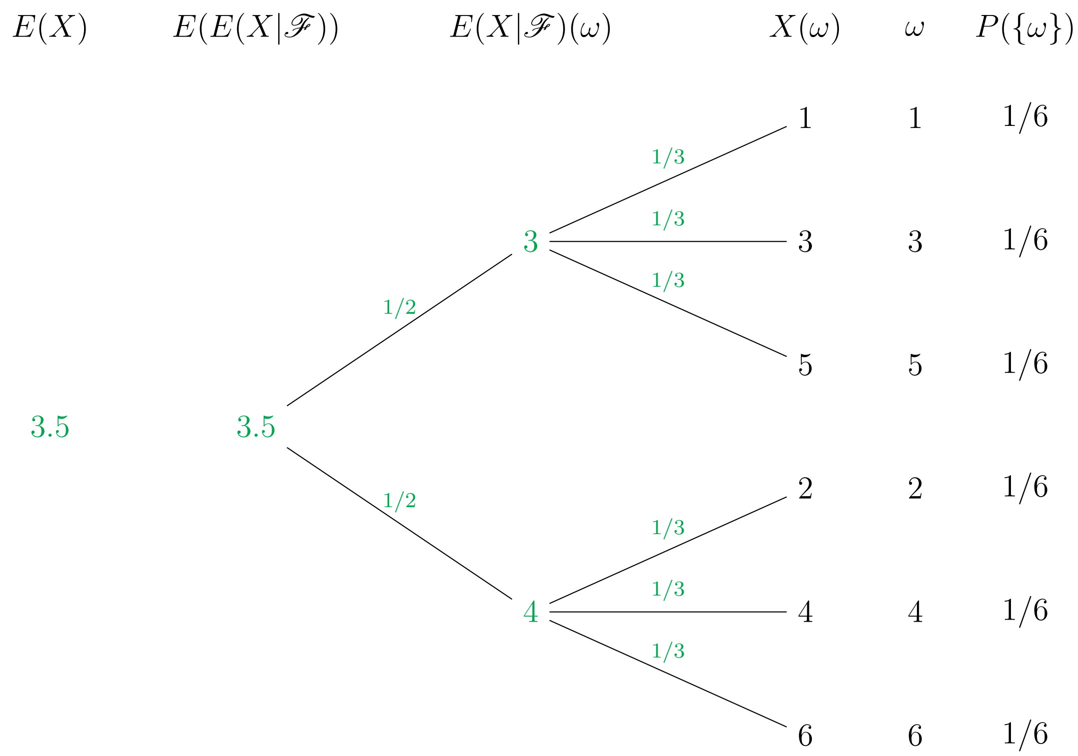

Figure 2.8: Expected values of X and $E(X|\mathscr{F})$

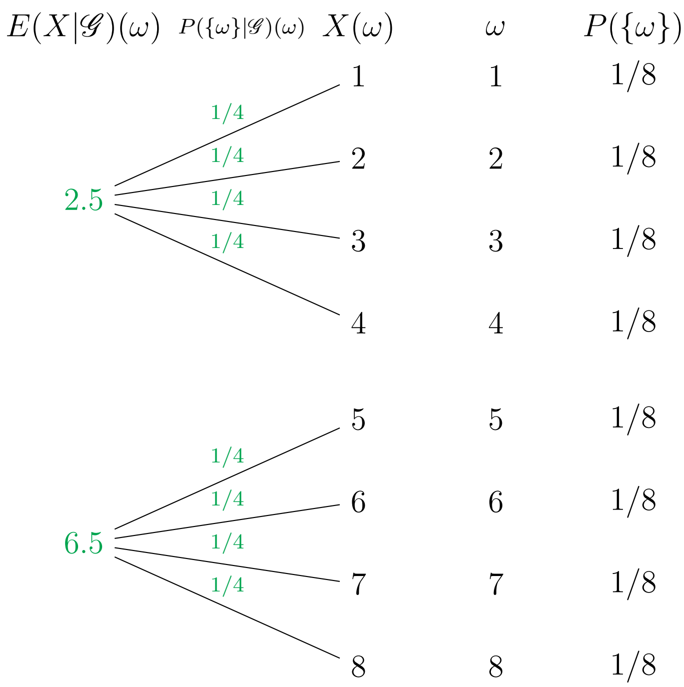

Figure 2.9: Iterated conditional expectations: Computation of $E(X|\mathscr{G})$

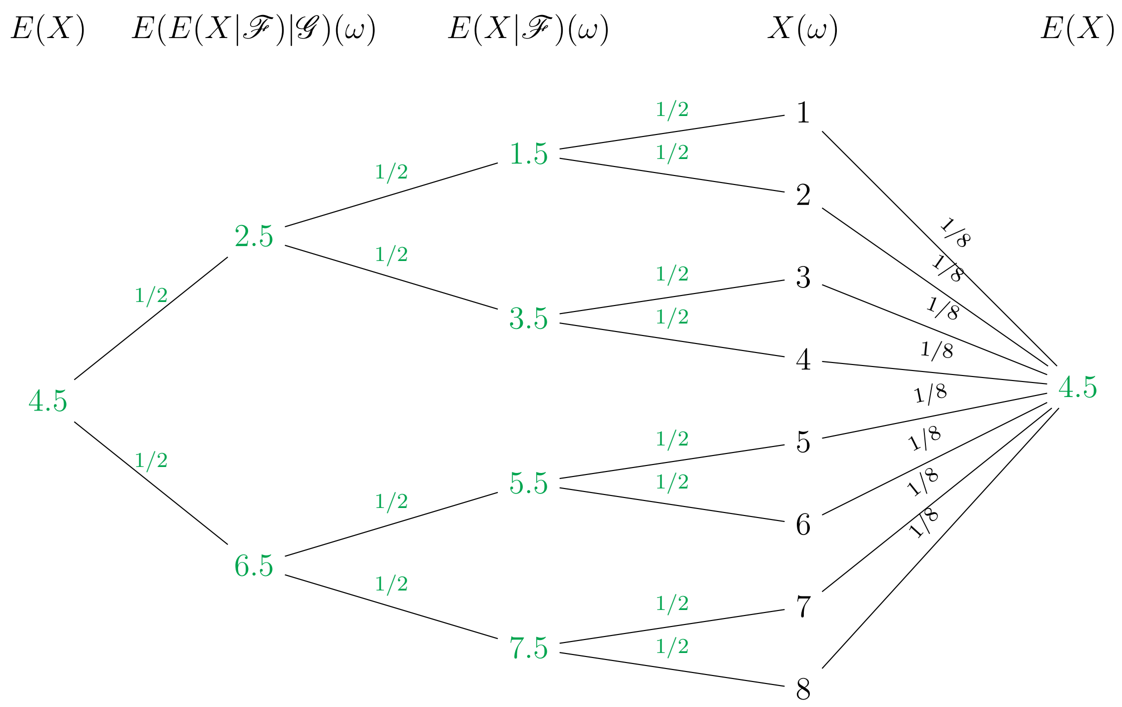

Figure 2.10: Iterated conditional expectations: Computation of $E(E(X|\mathscr{F})|\mathscr{G})$

ω	$P(\{\omega\})$	$Q(\{\omega\})$	$\frac{dQ}{dP}(\omega)$	$\frac{dP}{dQ}(\omega)$
1	1/6	1/10	3/5	5/3
2	1/6	1/10	3/5	5/3
3	1/6	1/10	3/5	5/3
4	1/6	1/10	3/5	5/3
5	1/6	1/10	3/5	5/3
6	1/6	1/2	3	1/3

Table 2.2: Equivalent probability measures and the corresponding densities

8. *If Z is $\mathscr{F}$-measurable, then $E(ZX|\mathscr{F}) = ZE(X|\mathscr{F})$.*

Let us illustrate some of these rules rather than proving them. The situation of Statement 1 happens in Figure 2.6, where X is 0 for even ω and 1 for odd ones. In Figure 2.7, the random variable X is independent of $\mathscr{F}$ as required in Statement 2. Here, $X(\omega)$ is $\omega/2$ rounded to integer values. Statement 3 is illustrated in Figure 2.8. Here, $X(\omega) = \omega$ and $\mathscr{F}$ is the σ-field generated by the sets where X is odd resp. even. Figure 2.9 and Figure 2.10 illustrate $E(E(X|\mathscr{F})|\mathscr{G}) = E(X|\mathscr{G})$ for two nested σ-fields $\mathscr{F}$ and $\mathscr{G}$. The similar statement $E(E(X|\mathscr{G})|\mathscr{F}) = E(X|\mathscr{G})$ follows from Statement 1 because $Y := E(X|\mathscr{G})$ is $\mathscr{F}$-measurable and hence $E(Y|\mathscr{F}) = Y$. Statements 5–7 parallel analogous rules for expected values. The first and the last statement means that $\mathscr{F}$-measurable random variables behave as constants if they appear inside a conditional expectation.

2.A.5 Absolute continuity and equivalence

In statistics and mathematical finance we often need to consider several probability measures at the same time. E.g. the probability measures P and Q in Table 2.2 correspond to an ordinary fair dice resp. a loaded dice, where the outcome 6 happens much more often. The tools of statistics are used if we do not know beforehand whether P or Q (or yet another probability measure) corresponds to our experiment. In mathematical finance, however, we usually assume the "real" probabilities to be known. Here, we consider alternative probability measures as a means to simplify calculations.

The relation between two probability measures P, Q can be described in terms of the *density* $\frac{dQ}{dP}$, which is the random variable whose value is simply the ratio of Q- and P-probabilities:

$$\frac{dQ}{dP}(\omega) := \frac{Q(\{\omega\})}{P(\{\omega\})}, \tag{2.3}$$

In other words, it is the ratio of the corresponding probability mass functions, cf. Table 2.2 for an example. Of course, (2.3) only makes sense if we do not divide by zero. More generally, densities are defined in the case of *absolute continuity*.

Definition 2.A.5.

1. A set $N \subset \Omega$ is called *null set* or, more precisely, *P-null set* if its probability is zero, i.e. $P(N) = 0$.
2. A probability measure Q is called *absolutely continuous* relative to P, written $Q \ll P$, if all P-null sets are Q-null sets as well.
3. Probability measures P, Q are called *equivalent*, written $P \sim Q$, if both $Q \ll P$ and $P \ll Q$, i.e. if both probability measures have the same null sets.

Equivalent probability measures differ with respect to concrete probabilities but not as to whether a given event may happen at all. If Q is absolutely continuous with respect to P, we define the *density* $\frac{dQ}{dP}$ as in (2.3). For ω with $P(\{\omega\}) = 0$, where (2.3) does not make sense, we take an arbitrary value, e.g. 0. We can then compute Q-probabilities from P-probabilities via

$$Q(A) = \int_A \frac{dQ}{dP} dP \left(= E_P\left(1_A \frac{dQ}{dP}\right)\right)$$

and, more generally, Q-expectations via

$$E_Q(X) = E_P\left(X \frac{dQ}{dP}\right).$$

Indeed, we have

$$\begin{aligned} E_Q(X) &= \sum_{\omega\in\Omega} X(\omega) Q(\{\omega\}) \\ &= \sum_{\omega\in\Omega} X(\omega) \frac{dQ}{dP}(\omega) P(\{\omega\}) \\ &= E_P\left(X \frac{dQ}{dP}\right) \end{aligned}$$

and hence

$$
\begin{aligned}
Q(A) &= E_Q(1_A) \\
&= E_P\Big(1_A \frac{dQ}{dP}\Big) \\
&= \int_A \frac{dQ}{dP} dP.
\end{aligned}
$$

Appendix MF

This chapter introduces a few probabilistic terms which are usually not discussed in introductory stochastics lectures.

2.B Some basics

2.B.1 Probability spaces, random variables, etc.

To follow this course, we assume basic knowledge in measure-theoretic probability. A good reference is (Jacod and Protter, 2004, Ch. 1-23).

2.B.2 Absolute continuity and equivalence

An essential device of financial mathematics is to not only consider the probability measure at hand. Instead, we analyze various measures under which certain expected values appear. We are however mainly interested in measures under which the sets with positive probability are the same as those under the original probability measure. Such *equivalent* changes of measure also play an important role in statistics.

Let μ, ν be measures on a measurable space $(\Omega, \mathscr{F})$. We will later almost exclusively talk about probability measures.

Definition 2.B.1. ν is called *absolutely continuous* relative to μ if all μ-null sets are also ν-null sets. We write $\nu \ll \mu$.[1]

Here a μ-null set is any subset of a set $N \in \mathscr{F}$ with $\mu(N) = 0$. We therefore do not necessarily require null sets to be measurable. At times, this is helpful for technical reasons.

Definition 2.B.2. If $\mu \ll \nu$ and $\nu \ll \mu$ we call μ and ν *equivalent* and write $\mu \sim \nu$.

The *Radon-Nikodym theorem* states that, if absolute continuity applies the dominated measure already has a density relative to the dominating measure.

Theorem 2.B.3 (Radon-Nikodým). *Let μ and ν be σ-finite measures (i.e. there are sets $A_1, A_2, \dots \in \mathscr{F}$ with $\cup_{i=1}^{\infty} A_i = \Omega$ and $\mu(A_n) < \infty$ for all n).*[2] *Then the following statements are equivalent:*

[1] If Ω is finite, $\nu \ll \mu$ simply means that $\mu(\{\omega\}) = 0$ implies $\nu(\{\omega\}) = 0$.

[2] In the case of a finite base space or for a probability measure this is automatically true.

1. *ν has a μ-density f, i.e. there is a non-negative measurable function $f : \Omega \to \mathbb{R}$ with $\nu(A) = \int_A f d\mu$ for all $A \in \mathscr{F}$.*[3]

2. *$\nu \ll \mu$.*

The density $\frac{d\nu}{d\mu} := f$ is μ-almost everywhere unique.[4]

Remark. For all (non-negative and measurable) or integrable functions $g : \Omega \to \mathbb{R}$ with respect to ν we have

$$\int g d\nu = \int g \frac{d\nu}{d\mu} d\mu.$$

Most proofs in the literature are based on Hilbert space theory (Riesz representation) or martingale theory (martingale convergence theorem). We give an elementary proof here. The following argument is needed:

Lemma. *Let μ be finite and $\mathcal{P} \subseteq \mathscr{F}$ be as follows:*

- *For all $A \in \mathscr{F}$ with $\mu(A) > 0$, there exists $B \subseteq A$ with $B \in \mathcal{P}$ and $\mu(B) > 0$.*
- *$\mathcal{P}$ is closed under countable disjoint unions.*
- *If $B \in \mathcal{P}$, $N \in \mathscr{F}$ with $\mu(N) = 0$, then $B \cup N \in \mathcal{P}$.*

Then, $\Omega \in \mathcal{P}$.

Proof. Let $\mathbb{A}$ be the collection of all sets

$$\{B_i : i \in I\} \subseteq \mathcal{P}$$

where I is some arbitrary index set and the B_i are disjoint and satisfy $\mu(B_i) > 0$ for all $i \in I$. We partially order $\mathbb{A}$ using the usual set inclusion. Now, every chain (i.e. every totally ordered subset) has an upper bound in $\mathbb{A}$, the union. By Zorn's Lemma, $\mathbb{A}$ has a maximal element $\{B_i : i \in I\}$. Since $\mu(B_i) > 0$ and $\mu(\Omega) < \infty$, I is countable and therefore $\bigcup_{i \in I} B_i \in \mathcal{P}$. The maximality ensures

$$\mu(\Omega \setminus \bigcup_{i \in I} B_i) = 0$$

as otherwise we could by assumption find $B \subseteq \Omega \setminus \bigcup_{i \in I} B_i$ such that $\mu(B) > 0$, $B \in \mathcal{P}$, hence

$$\{B_i : i \in I\} \cup \{B\} \in \mathbb{A},$$

a contradiction to maximality. By the last assumption, we obtain

$$\Omega = \bigcup_{i \in I} B_i \cup (\Omega \setminus \bigcup_{i \in I} B_i) \in \mathcal{P}.$$

□

[3]If Ω is finite this therefore means

$$\nu(A) = \sum_{\omega \in A} f(\omega) \mu(\{\omega\}). \tag{2.4}$$

[4]For a finite Ω it is $f(\omega) = \frac{\nu(\{\omega\})}{\mu(\{\omega\})}$ if not $\mu(\{\omega\}) = 0$, and the equation (2.4) is obvious.

Proof of Radon-Nikodym: The implication 1. $\Longrightarrow$ 2. is well-known (or easy to prove using algebraic induction). Now, we come to the other implication and assume $\nu \ll \mu$. We just prove the (particularly relevant) special case that μ, ν are finite measures. The general case can easily be obtained from that one by standard arguments.

1. Write
$$\mathcal{H} := \{f : f \text{ is measurable, } 0 \leq f, \nu(A) \geq \int_A f d\mu \text{ for all } A \in \mathscr{F}\}.$$
(The idea is to later find the density as the maximal function in $\mathcal{H}$.) Note that $0 \in \mathcal{H}$, hence $\mathcal{H} \neq \varnothing$. For all $f_1, f_2 \in \mathcal{H}$ it furthermore holds that for all $A \in \mathscr{F}$
$$\begin{aligned}\int_A \max\{f_1, f_2\} d\mu &= \int_{A \cap \{f_1 \geq f_2\}} f_1 d\mu + \int_{A \cap \{f_1 < f_2\}} f_2 d\mu \\ &\leq \nu(A \cap \{f_1 \geq f_2\}) + \nu(A \cap \{f_1 < f_2\}) \\ &= \nu(A),\end{aligned}$$
hence $\max\{f_1, f_2\} \in H$.

2. Write $M = \sup\{\int f d\mu : f \in \mathcal{H}\}$. Then $M < \infty$ as for all $f \in \mathcal{H}$ it holds that $\int f d\mu \leq \nu(\Omega) < \infty$. Now, find a sequence $(f_n)_n$ in $\mathcal{H}$ such that $\int f_n d\mu > M - 1/n$ for all n. Due to 1. we can – by considering $\max\{f_1, ..., f_n\}$ – assume w.l.o.g. that $f_1 \leq f_2 \leq ....$ Then $f_0 := \lim_{n\to\infty} f_n$ exists and for all $A \in \mathscr{F}$ it holds by monotone convergence
$$\int_A f_0 d\mu = \lim_{n\to\infty} \int_A f_n d\mu \leq \nu(A),$$
hence $f_0 \in \mathcal{H}$ and $\int f d\mu \leq \int f_0 d\mu = M$ for all $f \in \mathcal{H}$.

3. To prove that f_0 is the density as wanted, we assume by contradiction that there exists $\overline{A} \in \mathscr{F}$ with $\nu(\overline{A}) > \int_{\overline{A}} f_0 d\mu$. As $\nu(\overline{A}) > 0$, by assumption it also holds that $\mu(\overline{A}) > 0$. Let $\varepsilon > 0$ such that $\int_{\overline{A}} (f_0 + \varepsilon) d\mu < \nu(\overline{A})$. Now, denote the set of all A such that
$$\int_A (f_0 + \varepsilon 1_{\overline{A}}) d\mu > \nu(A)$$
by $\mathcal{P}$. Then, $\overline{A} \notin \mathcal{P}$. Obviously, $\mathcal{P}$ is closed under countable disjoint unions. Furthermore, by the absolute continuity, $\mathcal{P}$ is stable under unions with μ-null sets.

4. We now claim that there exists a measurable set $A \subseteq \overline{A}$ such that $f_0 + \varepsilon 1_A \in \mathcal{H}$. If not, then for all measurable sets $A \subseteq \overline{A}$ with positive μ-measure, there exists $B \subseteq A$ with
$$\int_B (f_0 + \varepsilon 1_B) d\mu > \nu(B).$$
Hence, the previous lemma is applicable, yielding $\overline{A} \in \mathcal{P}$, which is a contradiction. Hence, we can find $A \subseteq \overline{A}$ such that $f_0 + \varepsilon 1_A \in \mathcal{H}$.

5. It holds that
$$\int (f_0 + \varepsilon 1_A) d\mu = M + \varepsilon \mu(A) > M,$$
a contradiction to the definition of M.

□

2.B.3 Conditional expectation

In simple probability experiments we usually deal with only two different states of information. *Before the experiment* the outcome is still mostly unclear, and we can merely state probabilities for the different possible events. *After the experiment*, however, the actual outcome has been fully determined. This also reflects in random variables X. After the experiment we know X exactly, when before we settle e.g. with the expected value $E(X)$ as the "expected" mean value.

However, if experiments take place over a longer period of time, this approach seems inappropriate. As time passes, our picture of what the outcome will be becomes more precise. We would therefore like to consider probabilities and expected values *on the basis of the information we have at that point in time.* In order to achieve this, we have to first define the vague term of current information. It is one of the extraordinarily productive ideas of probability theory to do this by using σ-algebras. In measure theory, they originally only show up in the definition of measures for which the power set proved to be too big, the Lebesgue measure for example.

Now, in what way does a σ-algebra $\mathscr{C}$ stand for the extent of information that is available at a particular point in time? This happens as $\mathscr{C}$ contains exactly those events for which we can–at the current moment–say with certainty whether or not they occur.

Let us look at an example. We roll a dice three times and call the results X_1, X_2, X_3. After the first roll, all events that only refer to this first roll have already been decided, e.g. the event $\{X_1 \text{ is even}\}$. The σ-algebra $\mathscr{C}$ that fits this state of information is therefore the one that is generated by the random variable X_1, i.e. $\mathscr{C} = \sigma(X_1) = \{X_1^{-1}(B) : B \in \mathscr{B}\}$.[5] But expectations regarding the not yet determined results and random variable can have changed after the first roll as well. For example, for the total value we have $E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 10.5$; however, after the first roll we expect on average $X_1 + E(X_2) + E(X_3) = X_1 + 7$, because the random variable X_1 is no longer random for us.

We call this expected value with respect to the information $\mathscr{C}$ *conditional expectation given* $\mathscr{C}$ and write $E(X|\mathscr{C})$. By defining $P(A|\mathscr{C}) = E(1_A|\mathscr{C})$ we can view *conditional probabilities* as a special case of conditional expectations. For the function $A \mapsto P(A|\mathscr{C})$ to be also σ-additive, as is expected of a probability measure, we need to overcome a few obstacles in measure theory which we are not going to discuss further at this point. Therefore, we restrain ourselves to conditional expectations.

Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\mathscr{C}$ a sub σ-algebra of $\mathscr{F}$ (i.e. $\mathscr{C} \subset \mathscr{F}$). Let also X be an $\overline{\mathbb{R}}$-valued random variable[6]. We will first examine the case that Ω only contains finite or countably many elements.

Lemma 2.B.4. *Let Ω be finite or countable. We then have a partition[7] $(C_i)_{i\in I}$ of Ω, so that $\mathscr{C} = \{\cup_{i\in J} C_i : J \subset I\}$.*

Proof. For $\omega \in \Omega$ let $C(\omega) = \bigcap\{C \in \mathscr{C} : \omega \in C\}$. For $\omega' \in \Omega \backslash C(\omega)$ there is a $C(\omega, \omega') \in \mathscr{C}$

[5] If Ω is finite, $\sigma(X_1)$ contains unions of sets of the type $X^{-1}(x) = \{X = x\} := \{\omega \in \Omega : X(\omega) = x\}$, wherein $x \in \mathbb{R}$.

[6] $\overline{\mathbb{R}} := [-\infty, \infty]$

[7] *Partition* means $\cup_{i\in I} C_i = \Omega$ and all C_i are pairwise disjoint, i.e. $C_i \cap C_j = \varnothing$ for $i \neq j$.

such that $\omega \in C(\omega, \omega')$ and $\omega' \notin C(\omega, \omega')$. It follows that $C(\omega) = \cap\{C(\omega, \omega') : \omega' \in \Omega \setminus C(\omega)\} \in \mathscr{C}$, because Ω is countable. Thus, $C(\omega)$ is the smallest element of $\mathscr{C}$ which contains ω. We number the at most countably many sets $C(\omega)$, $\omega \in \Omega$ as $C_0, C_1, \ldots$, I being the set of used indices.

We first prove that these sets are disjoint, and then it is clear that they form a partition. So, note that $\eta \in C(\omega)$ implies $C(\eta) \subset C(\omega)$. If now ω is not an element of $C(\eta)$ then $C(\omega)\backslash C(\eta)$ is an element of $\mathscr{C}$ which contains ω and that is a contradiction of the minimality of $C(\omega)$. Thus, two sets $C(\eta)$ and $C(\omega)$ are either disjoint or the same.

Finally, define $\widetilde{\mathscr{C}} := \{\cup_{i \in J} C_i : J \subset I\}$. We have $\widetilde{\mathscr{C}} \subset \mathscr{C}$, because $\mathscr{C}$ is a σ-algebra. For $D \in \mathscr{C}$ we also easily have $D = \cup_{\omega \in D} C(\omega) \in \widetilde{\mathscr{C}}$. Together it follows that $\mathscr{C} = \widetilde{\mathscr{C}}$. □

This lemma basically means that there are finite or countably many *atoms* C_i which make up the sets of the σ-algebra. The different outcomes $\omega \in C_i$ of the individual atoms always show up together in events and can not be separated based on the information provided by $\mathscr{C}$. In other words: Based on the information given by $\mathscr{C}$ we know exactly which of the events C_i happens, but we do not know which specific outcome $\omega \in C_i$ has occurred at the end of the experiment. The bigger the σ-algebra $\mathscr{C}$ is, the more we can narrow down to the outcome ω, i.e. the more information we have about the outcome of the experiment.

We can now define the conditional expectation in finite or countable cases.

Definition 2.B.5. Define $\mathscr{C} = \{\cup_{i \in J} C_i : J \subset I\}$, whereas $(C_i)_{i \in I}$ is a finite or countable partition of Ω.[8] Let X also be non-negative or integrable.[9] We define the *conditional expectation of X given $\mathscr{C}$* as

$$E(X|\mathscr{C})(\omega) := \begin{cases} E(X|C_i) & \text{if } \omega \in C_i \text{ with } P(C_i) > 0, \\ 0 & \text{if } \omega \in C_i \text{ with } P(C_i) = 0. \end{cases} \tag{2.5}$$

Here $E(X|C_i)$ is the expectation under the conditional probability measure $P(\cdot|C_i)$ which is defined by $A \mapsto \frac{P(A \cap C_i)}{P(C_i)}$. We have

$$E(X|C_i) := \frac{E(X 1_{C_i})}{P(C_i)}.$$

If we interprete the expectation $E(X)$ as the best prognosis of a random variable X, then the conditional expectation $E(X|\mathscr{C})$ is the best prognosis given the information $\mathscr{C}$. In this case, since we already know in which set C_i the outcome of the experiment is, we calculate the expectation in (2.5) using only the $\omega \in C_i$. The conditional expectation has, among others, following important properties.

Theorem 2.B.6. *The conditional expectation of Definition 2.B.5 is $\mathscr{C}$-measurable, and we have*

$$\int_C E(X|\mathscr{C}) dP = \int_C X dP \tag{2.6}$$

for all $C \in \mathscr{C}$. Also, $E(X|\mathscr{C})$ is non-negative or integrable if that is the case for X.

[8] If Ω is finite, then every σ-algebra has such a form according to the previous lemma. In this case, we only need a finite number of *atoms* C_i.

[9] Integrability follows automatically for finite Ω if the values of the random variables are all finite.

Proof. Let $B \in \mathscr{B}$ with $0 \notin B$. Then we have $\{E(X|\mathscr{C}) \in B\} = \cup_{i \in J} C_i$ for $J := \{i \in I : P(C_i) > 0 \text{ and } E(X|C_i) \in B\}$. Notably, $E(X|\mathscr{C})$ is $\mathscr{C}$-measurable.

Now let $C \in \mathscr{C}$, and suppose first that $C = C_i$ for an $i \in I$. Then we have

$$\int_C E(X|\mathscr{C})dP = \int_{C_i} E(X|C_i)dP = P(C_i)\frac{E(X1_{C_i})}{P(C_i)} = \int_{C_i} XdP = \int_C XdP$$

in the case of $P(C_i) > 0$ and similarly for $P(C_i) = 0$. The claim for a general C now follows from σ-additivity.

$E(X|\mathscr{C})$ is obviously non-negative if that is true for X. If X is integrable we have for $C := \{E(X|\mathscr{C}) \geq 0\} \in \mathscr{C}$:

$$\int E(X|\mathscr{C})^+ dP = \int_C E(X|\mathscr{C})dP = \int_C XdP \leq \int |X|dP < \infty$$

and analogously $E(E(X|\mathscr{C})^-) < \infty$. It follows that $E(X|\mathscr{C})$ is integrable. $\square$

$\mathscr{C}$-measurability in the previous theorem basically means that the conditional expectation given the information of $\mathscr{C}$ is known. It is therefore a very natural property. The integral equation means that X is not different from its conditional expectation viewed through $\mathscr{C}$-measurable sets in the sense that they produce the same integrals. So in a sense it behaves almost like X itself.

In general probability spaces Definition 2.B.5 is not applicable, because sub-σ-algebras are generally not generated by a partition. We therefore substitute a definition based on the properties of the previous theorem that is still useful in general situations.

Theorem 2.B.7. *If X is non-negative (or integrable) then there is a P-almost surely unique $\mathscr{C}$-measurable non-negative (or integrable) random variable $E(X|\mathscr{C})$ so that*

$$\int_C E(X|\mathscr{C})dP = \int_C XdP \tag{2.7}$$

for all $C \in \mathscr{C}$.

Proof. 1. Step: First, let $X \geq 0$ and almost surely finite. We define a new measure $Q \ll P$ with the density $\frac{dQ}{dP} := X$. For measures $P|_{\mathscr{C}}, Q|_{\mathscr{C}}$ that are limited to the sub-σ-algebra $\mathscr{C}$ we have $Q|_{\mathscr{C}} \ll P|_{\mathscr{C}}$: For every $P|_{\mathscr{C}}$-null set N there is an $A \in \mathscr{C} \subset \mathscr{F}$ with $N \subset A$ and $P|_{\mathscr{C}}(A) = P(A) = 0$, which also implicates $Q|_{\mathscr{C}}(A) = Q(A) = 0$ because $Q \ll P$. According to the Radon-Nikodym theorem there is a $\mathscr{C}$-measurable, non-negative density $Y := \frac{dQ|_{\mathscr{C}}}{dP|_{\mathscr{C}}}$, i.e. $Q|_{\mathscr{C}}(C) = \int_C YdP|_{\mathscr{C}}$ for all $C \in \mathscr{C}$. For Y to have the properties of a conditional expectation we need to show that

$$\int_C YdP = \int_C XdP \tag{2.8}$$

for all $C \in \mathscr{C}$. This follows because

$$\int_C XdP = Q(C) = Q|_{\mathscr{C}}(C) = \int_C YdP|_{\mathscr{C}} \overset{(\star)}{=} \int_C YdP,$$

The last equation $(\star)$ may not be obvious.

We prove it using a proof method sometimes called *algebraic induction*. This means we first show a statement for indicator functions, then for random variables with finitely many values, then for general non-negative ones and lastly for all random variables by disassembling them into positive and negative parts.

For $Z = 1_A$ with $A \in \mathscr{C}$ we have

$$\int Z dP|_{\mathscr{C}} = P|_{\mathscr{C}}(A) = P(A) = \int Z dP$$

as wanted. Because of the linearity of the integral $\int Z dP|_{\mathscr{C}} = \int Z dP$ is also true for linear combinations of those indicators. According to the monotone convergence theorem it is also true for general non-negative $\mathscr{C}$-measurable Z, since every such Z can be written as monotone limit of linear combinations of indicators. To prove $(\star)$ we select $Z = Y 1_C$.

2. Step: We now consider integrable X. We construct their conditional expectation as

$$E(X|\mathscr{C}) := E(X^+|\mathscr{C}) - E(X^-|\mathscr{C}),$$

where the terms on the right hand side are defined by the first step. The right hand side is obviously $\mathscr{C}$-measurable. Integrability follows because of

$$\begin{aligned} E(|E(X^+|\mathscr{C}) - E(X^-|\mathscr{C})|) &\leq E(E(X^+|\mathscr{C}) + E(X^-|\mathscr{C})) \\ &= E(E(X^+|\mathscr{C})) + E(E(X^-|\mathscr{C})) \\ &= E(X^+) + E(X^-) \\ &= E(|X|) < \infty \end{aligned}$$

where we have used that (2.7) holds for the terms defined by the first step. Equation (2.7) then follows in the integrable case from the linearity of the integrals.

3. Step: For the sake of completeness we also examine the less important case of a general non-negative random variable X. Let $A := \{X = \infty\}$. We define $Y := \infty E(1_A|\mathscr{C}) + E(X 1_{A^C}|\mathscr{C})$, whereby the right side is explained by the first step and we use the usual convention of $\infty \cdot 0 = 0$. Obviously, Y is non-negative and $\mathscr{C}$-measurable. Let $C \in \mathscr{C}$. In the case that $P(C \cap A) > 0$ we have

$$\int_C Y dP \geq \int_C \infty 1_A dP = \infty P(C \cap A) = \infty$$

and

$$\int_C X dP \geq \infty P(C \cap A) = \infty,$$

so that (2.8) applies. If $P(C \cap A) = 0$, then (2.8) follows from

$$\int_C Y dP = \infty P(C \cap A) + \int_C X 1_{A^C} dP = \int_C X 1_{A^C} dP = \int_C X dP.$$

4. Step: Lastly, we need to show indistinguishability. Let $Y, \widetilde{Y}$ be two $\mathscr{C}$-measurable random variables that have the properties of conditional expectations. For $n \in \mathbb{N}$ define $C := \{Y > \widetilde{Y} \text{ and } |Y| \vee |\widetilde{Y}| \leq n\}$. We have

$$\int_C (Y - \widetilde{Y}) dP = \int_C Y dP - \int_C \widetilde{Y} dP = \int_C X dP - \int_C X dP = 0,$$

thus $P(C) = 0$, since $Y - \widetilde{Y} > 0$ on C. Continuity from below implicates $P(Y > \widetilde{Y}) = 0$. $P(Y < \widetilde{Y}) = 0$ follows in the same manner and therefore $Y = \widetilde{Y}$ almost surely. □

According to Theorem 2.B.6, $E(X|\mathscr{C})$ coincides with the conditional expectation introduced in Definition 2.B.5 if the conditions for $\mathscr{C}$ are met. This motivates the following definition.

Definition 2.B.8.
1. $E(X|\mathscr{C})$ is called *conditional expectation of X given $\mathscr{C}$.*
2. By setting $E(X|\mathscr{C}) := E(X^+|\mathscr{C}) - E(X^-|\mathscr{C})$ we can define $E(X|\mathscr{C})$ even in the case where either $E(X^+) = \infty$ or $E(X^-) = \infty$.

The following example shows how to calculate conditional expected values directly with the definition.

Example 2.B.9. Let $X_1, X_2, \dots, X_n$ be iid, $U(0,1)$-distributed and $S = \max\{X_1, ..., X_n\}$. The goal now is to determine $E(X_1|S) := E(X_1|\sigma(S))$. To do this, we first heuristically search for a candidate: With probability $1/n$, X_1 is the maximum. With probability $(n-1)/n$ it is one of the other random variables. If $S = s$, then X_1 has values in $[0, s]$. It seems plausible that X_1 then continues to be uniformly-distributed, but now at $[0, s]$, that is, with expected value $s/2$. As a candidate for the conditional expectation value, we get

$$Y := \frac{1}{n}S + \frac{n-1}{n}\frac{S}{2} = \frac{n+1}{2n}S.$$

We will now prove this. First, Y is $\sigma(S)$-measurable and sets of the form $A = \{S \leq s\}$, $s \in \mathscr{R}$, form a $\cap$-stable generating end system of $\sigma(S)$. To be shown, according to the previous remark, is thus

$$E(1_A X_1) = E(1_A Y) \text{ f.a. } A\{S \leq s\},\ s \in \mathscr{R}.$$

We first note that because of

$$P(S \leq t) = t^n,\ t \in [0,1],$$

the density of S is given by

$$f_S(t) = nt^{n-1},\ t \in [0,1].$$

Thus

$$E(1_A Y) = \frac{n+1}{2n}E(1_{\{S \leq s\}}S) = \frac{n+1}{2n}\int_0^s tnt^{n-1}dt = \frac{s^{n+1}}{2}.$$

On the other hand, with the known formula

$$EZ = E\int_0^\infty 1_{\{z<Z\}}dz = \int_0^\infty P(Z > z)dz$$

for nonnegative ZV Z:

$$\begin{aligned}
E(1_A X_1) &= \int_0^1 P(1_A X_1 > x)dx = \int_0^1 P(X_1 > x, S \leq s)dx \\
&= \int_0^s P(X_1 \in (x,s], \max\{X_2,\dots,X_n\} \leq s)dx \\
&= \int_0^s P(X_1 \in (x,s])P(\max\{X_2,\dots,X_n\} \leq s)dx \\
&= \int_0^s (s-x)s^{n-1}dx \\
&= s^{n-1}s^2/2 = \frac{s^{n+1}}{2},
\end{aligned}$$

thus the assertion.

In general, the conditional expectation is therefore determined implicitly by desired properties rather than explicitly by a formula. The following theorem contains important calculation rules for conditional expectations.

Theorem 2.B.10. *Let X be non-negative or integrable. Then we have:*

1. $E(X|\mathscr{C}) = X$ *if X is $\mathscr{C}$-measurable.*
2. $E(X|\mathscr{C}) = E(X)$ *if $\sigma(X)$ and $\mathscr{C}$ are independent.*[10]
3. $E(E(X|\mathscr{C})) = E(X)$.
4. $E(E(X|\mathscr{C})|\mathscr{D}) = E(X|\mathscr{D})$ *if $\mathscr{D}$ is a sub-σ-algebra of $\mathscr{C}$.*
5. *The function $X \mapsto E(X|\mathscr{C})$ is linear and monotone on its domain.*
6. *The* monotone convergence theorem *applies, i.e. for every increasing sequence $(X_n)_{n\in\mathbb{N}}$ of non-negative random variables with the limit $X := \sup_{n\in\mathbb{N}} X_n$ we have*

   $$E(X_n|\mathscr{C}) \uparrow E(X|\mathscr{C})$$

   almost surely.
7. *The* dominated convergence theorem *applies, i.e. for every almost surely converging sequence $(X_n)_{n\in\mathbb{N}}$ of random variables with the limit X and $E(\sup_{n\in\mathbb{N}} |X_n|) < \infty$ we have*

   $$E(X_n|\mathscr{C}) \to E(X|\mathscr{C}) \tag{2.9}$$

   almost surely and in L^1.
8. Jensen's inequality *applies, i.e. for every integrable X and every convex function $f : \mathbb{R} \to \mathbb{R}$ so that $f(X)$ is integrable, we have*

   $$f(E(X|\mathscr{C})) \leq E(f(X)|\mathscr{C}).$$

[10] i.e. if $P(A \cap B) = P(A)P(B)$ for all $A \in \sigma(X)$, $B \in \mathscr{C}$

9. *For $\mathscr{C}$-measurable $Y : \Omega \to \overline{\mathbb{R}}$ we have $E(XY|\mathscr{C}) = E(X|\mathscr{C})Y$ if both sides of the equation are well-defined, i.e. if $X, Y \geq 0$ or X, XY are integrable. We have in particular $E(XY) = E(E(X|\mathscr{C})Y)$.*

Proof.

1. X obviously has the properties demanded in the definition.

2. $\mathscr{C}$-measurability as well as non-negativity or integrability of the candidate are obvious. Let $C \in \mathscr{C}$. Since X is $\sigma(X)$-measurable and 1_C is $\mathscr{C}$-measurable, we have $E(X1_C) = E(X)E(1_C)$ because of the independence of the σ-algebras and therefore

$$\int_C E(X)dP = E(X)E(1_C) = E(X1_C) = \int_C XdP$$

as desired.

3. This is true because

$$E(E(X|\mathscr{C})) = \int_\Omega E(X|\mathscr{C})dP = \int_\Omega XdP = E(X).$$

4. We show that the candidate $E(E(X|\mathscr{C})|\mathscr{D})$ meets the properties of the conditional expectation $E(X|\mathscr{D})$. $\mathscr{D}$-measurability follows from the definition. For $D \in \mathscr{D} \subset \mathscr{C}$ we also have

$$\int_D E(E(X|\mathscr{C})|\mathscr{D})dP = \int_D E(X|\mathscr{C})dP = \int_D XdP$$

as desired.

5. For random variables X, Y, $E(X|\mathscr{C}) + E(Y|\mathscr{C})$ is $\mathscr{C}$-measurable: With

$$\begin{aligned} \int_C (E(X|\mathscr{C}) + E(Y|\mathscr{C}))dP &= \int_C E(X|\mathscr{C})dP + \int_C E(Y|\mathscr{C})dP \\ &= \int_C XdP + \int_C YdP \\ &= \int_C (X+Y)dP \end{aligned}$$

for $C \in \mathscr{C}$ follows that $E(X|\mathscr{C}) + E(Y|\mathscr{C})$ has the properties of the conditional expectation $E(X+Y|\mathscr{C})$. Similarly, we can show $E(cX|\mathscr{C}) = cE(X|\mathscr{C})$ for $c \in \mathbb{R}$, provided that cX is non-negative or integrable.

In the case that $X \leq Y$ we disassemble $Y = X + (Y - X)$. Because of $Y - X \geq 0$ we also have $E(Y - X|\mathscr{C}) \geq 0$. Because of the additivity we have $E(Y|\mathscr{C}) = E(X|\mathscr{C}) + E(Y - X|\mathscr{C})$ and therefore $E(Y|\mathscr{C}) \geq E(X|\mathscr{C})$.

6. Because of the monotony of conditional expectations $(E(X_n|\mathscr{C}))_{n\in\mathbb{N}}$ is an increasing sequence. Its limit $Y := \sup_{n\in\mathbb{N}} E(X_n|\mathscr{C})$ is $\mathscr{C}$-measurable. It remains to show that (2.8) is true for any desired $C \in \mathscr{C}$. According to the usual monotone convergence theorem we have

$$\int_C YdP = \lim_{n\to\infty} \int_C E(X_n|\mathscr{C})dP = \lim_{n\to\infty} \int_C X_n dP = \int_C XdP$$

as desired.

7. Define $Y := \sup_{n\in\mathbb{N}} |X_n|$, $X_n' := \sup_{k\geq n} X_k$, $X_n'' := \inf_{k\geq n} X_k$. Then we have
$$-Y \leq X_n'' \leq X_n \leq X_n' \leq Y, \quad n \in \mathbb{N}.$$
For $n \to \infty$ we have
$$\begin{aligned} Y - X_n' &\uparrow Y - \limsup_{n\to\infty} X_n = Y - X, \\ Y + X_n'' &\uparrow Y + \liminf_{n\to\infty} X_n = Y + X. \end{aligned}$$
It then follows that
$$\begin{aligned} E(Y - X_n'|\mathscr{C}) &\uparrow E(Y - X|\mathscr{C}) \text{ a.s. and in } L^1, \\ E(Y + X_n'|\mathscr{C}) &\uparrow E(Y + X|\mathscr{C}) \text{ a.s. and in } L^1 \end{aligned}$$
where we have applied Statement 6 for the a.s. convergence and the usual dominated convergence theorem for the convergence in L^1. Therefore $E(X_n'|\mathscr{C}), E(X_n''|\mathscr{C}) \to E(X|\mathscr{C})$ almost surely and in L^1. Because of
$$E(X_n'|\mathscr{C}) \leq E(X_n|\mathscr{C}) \leq E(X_n''|\mathscr{C})$$
(2.9) follows.

8. According to theorems of analysis, convex functions can be written as
$$f(x) = \sup_{n\in\mathbb{N}} (a_n x + b_n)$$
with real coefficients a_n, b_n. It follows that
$$\begin{aligned} f(E(X|\mathscr{C})) &= \sup_{n\in\mathbb{N}} (a_n E(X|\mathscr{C}) + b_n) \\ &= \sup_{n\in\mathbb{N}} E(a_n X + b_n|\mathscr{C}) \\ &\leq E\left(\sup_{n\in\mathbb{N}} (a_n X + b_n)\right) \\ &= E(f(X)|\mathscr{C}). \end{aligned}$$

9. We show that the candidate $E(X|\mathscr{C})Y$ has the properties of $E(XY|\mathscr{C})$. $\mathscr{C}$-measurability is clear. For $Y = 1_A$ with $A \in \mathscr{C}$ we have
$$\int_C E(X|\mathscr{C})Y\,dP = \int_{C\cap A} E(X|\mathscr{C})dP = \int_{C\cap A} X\,dP = \int_C XY\,dP, \quad \forall C \in \mathscr{C}$$
as desired. We continue with algebraic induction. The formula for unconditional expectations follows from property 3. □

The conditional expectation can also be interpreted as an orthogonal projection.

Theorem 2.B.11. *Let $X \in L^2(\Omega, \mathscr{F}, P)$. Then $E(X|\mathscr{C})$ is the orthogonal projection*[11] *of X on the linear subspace $L^2(\Omega, \mathscr{C}, P)$.*

[11] $L^2(\Omega, \mathscr{F}, P)$ is a Hilbert space regarding the scalar product $\langle X, Y\rangle := E(XY)$ if we identify almost surely identical random variables with each other.

Proof. According to Jensen's inequality we have $(E(X|\mathscr{C}))^2 \leq E(X^2|\mathscr{C})$ and therefore

$$E\Big(E(X|\mathscr{C})^2\Big) \leq E\Big(E(X^2|\mathscr{C})\Big) = E(X^2) < \infty.$$

Since $E(X|\mathscr{C})$ is $\mathscr{C}$-measurable, we also have $E(X|\mathscr{C}) \in L^2(\Omega, \mathscr{C}, P)$. For $Y \in L^2(\Omega, \mathscr{C}, P)$ we also have, according to theorem 2.B.10(9),

$$\begin{aligned} \langle X - E(X|\mathscr{C}), Y\rangle &= E((X - E(X|\mathscr{C}))Y) \\ &= E(XY) - E(E(X|\mathscr{C})Y) \\ &= E(XY) - E(XY) = 0, \end{aligned}$$

i.e. $X - E(X|\mathscr{C})$ and $L^2(\Omega, \mathscr{C}, P)$ are orthogonal. This proves the theorem. □

The following result can be useful when calculating conditional expectations.

Lemma 2.B.12. *Let Y be a random variable and $g : \overline{\mathbb{R}} \times \mathbb{R} \to \mathbb{R}$ a measurable function so that $g(X, Y)$ is non-negative or integrable. If X is $\mathscr{C}$-measurable and Y independent of $\mathscr{C}$ we have*

$$E(g(X, Y)|\mathscr{C}) = \int g(X, y)P^Y(dy).^{12}$$

Proof. Let g first be non-negative. In connection with Fubini's theorem it is shown that the function $x \mapsto \int g(x, y)P^Y(dy)$ is Borel-measurable. Therefore the random variable $\int g(X, y)P^Y(dy)$ is $\mathscr{C}$-measurable as a composition of two measurable functions. For $C \in \mathscr{C}$ define $Z := 1_C$. Since (X, Z) is independent of Y, we have $P^{(X,Z)} \otimes P^Y = P^{(X,Z,Y)}$. Using the transformation theorem and Fubini's theorem we have

$$\begin{aligned} \int_C \int g(X, y)P^Y(dy)dP &= \int \int g(X, y)ZP^Y(dy)dP \\ &= \int \int g(x, y)zP^Y(dy)P^{(X,Z)}(d(x, z)) \\ &= \int g(x, y)zP^{(X,Z,Y)}(d(x, z, y)) \\ &= \int g(X, Y)ZdP \\ &= \int_C g(X, Y)dP, \end{aligned}$$

which results in our claim.

The proof for integrable $g(X, Y)$ works the same way. The integrability of the candidate results from the previous calculation applied to $|g(X, Y)|$ and $C = \Omega$:

$$E(|\int g(X, y)P^Y(dy)|) \leq E(\int |g(X, y)|P^Y(dy)) = E(|g(X, Y)|) < \infty.$$

□

Notation. $E(X|Y) := E(X|\sigma(Y))$ for measurable functions $Y : (\Omega, \mathscr{F}) \to (\Gamma, \mathscr{G})$ with values in a measurable space $(\Gamma, \mathscr{G})$.

[12] In the finite case that means

$$E(g(X, Y)|\mathscr{C})(\omega) = \sum_{y\in\mathbb{R}} g(X(\omega), y)P(Y = y).$$

Intuitively, the property $\mathscr{C}$*-measurability* of conditional expectation means that $E(X|\mathscr{C})$ is determined by the information in $\mathscr{C}$. Now, if the σ-algebra $\mathscr{C}$ is generated by a random variable Y, we should expect that $E(X|\mathscr{C})$ can be written as a function of Y. The following remark shows that this is indeed the case.

Lemma 2.B.13. *Let* $Y : (\Omega, \mathscr{F}) \to (\Gamma, \mathscr{G})$ *with values in a measurable space* $(\Gamma, \mathscr{G})$. *Then* X *is* $\sigma(Y)$*-measurable if and only if there is a measurable function* $g : (\Gamma, \mathscr{G}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ *with* $X = g \circ Y$.

Proof. $\Rightarrow$: In the case of $X = 1_C$, $\sigma(Y)$-measurability means that $C \in \sigma(Y) = Y^{-1}(\mathscr{G})$, and thus $C = Y^{-1}(G)$ for a $G \in \mathscr{G}$. Therefore we have $X = 1_G(Y) = g \circ Y$ for $g := 1_G$. We continue with algebraic induction.

$\Leftarrow$: This holds because the composition of measurable functions is measurable. $\square$

Chapter 3

Discrete stochastic calculus

In this chapter we introduce dynamic stochastic systems.

In mathematical finance, we treat asset price movements as *stochastic processes*, i.e. as random functions of time. The same is true for the varying number of assets in a portfolio and the resulting wealth. In this chapter we introduce the corresponding mathematical notions which are applicable outside of mathematical finance as well. We confine ourselves at this point to a discrete set $0, 1, 2, 3, \dots$ of periods (e.g. days, minutes, seconds) because the continuous case requires considerably more involved mathematics. We also assume that Ω is finite so that (conditional) expectations of a real-valued X always exist.

Stochastic processes

A crucial role is played by the information which is available at any particular point in time. Decisions concerning e.g. buying or selling assets can only be based on the present knowledge about prices or about the market as a whole. The flow of information is expressed mathematically in terms of a so-called *filtration*.

Definition 3.1. A *filtration* $\mathscr{F}_0, \mathscr{F}_1, \mathscr{F}_2, \dots$ on a probability space $(\Omega, \mathfrak{P}(\Omega), P)$ is an increasing sequence of σ-fields on Ω, i.e. $\mathscr{F}_m \subset \mathscr{F}_n$ for $m \leq n$. $(\Omega, \mathfrak{P}(\Omega), (\mathscr{F}_n)_{n \in \mathbb{N}}, P)$ is called a *filtered probability space*.

The σ-field $\mathscr{F}_n$ represents the information that is available at time n. An event A belongs to $\mathscr{F}_n$ if it is known at time n whether it has happened or not. For an illustration cf. Figures 3.1-3.3 where σ-fields are represented by their generating partition.

From now on we fix a filtered probability space $(\Omega, \mathfrak{P}(\Omega), (\mathscr{F}_n)_{n \in \mathbb{N}}, P)$.

Definition 3.2. A *stochastic process* $X = (X_0, X_1, X_2, \dots)$ is a sequence of random variables. It is called *adapted* if X_n is $\mathscr{F}_n$-measurable for $n = 0, 1, \dots$

A stochastic process represents the random state of a process through time. n stands for the corresponding time as above. The range of X_n are usually numbers or, more generally, vectors. X_n could denote e.g. the price(s) of one or several stocks at time n.

Unless otherwise noted, we assume the values of all processes to be numbers rather than vectors.

Adaptedness means that we observe or know the present state of the process, at least as far as it is random, cf. Figures 3.5, 3.6, 3.8. In particular, all deterministic processes are adapted. From now on we consider only such adapted processes.

Notation. Occasionally we write $n- := n-1$ for the previous time and set $0- := 0$. Moreover, we denote by $\Delta X_n := X_n - X_{n-}$ the *increment* or *jump* of X in period n. Finally, we write X_- for the time-shifted process X, more specifically $(X_-)_n := X_{n-}$.

The following notion is stronger than adaptedness.

Definition 3.3. A process X is called *predictable* if X_n is $\mathscr{F}_{n-1}$-measurable for any n, i.e X_n is known already at time $n-1$.

We will encounter predictable processes in the context of stochastic integration below. Note that X_- is predictable if X is an adapted process.

Where does the filtration in our general mathematical setup come from? Often there is no need to specify it in detail. But if it is, then it is typically of the following form.

Definition 3.4. A filtration $\mathscr{F}_0, \mathscr{F}_1, \mathscr{F}_2, \dots$ is said to be the *filtration generated by a process* X if $\mathscr{F}_n = \sigma(X_0, \dots, X_n)$ for any n, i.e. $\mathscr{F}_n$ is the σ-field generated by $X_0, \dots, X_n$.

The filtration generated by X is the smallest filtration relative to which X is adapted, cf. Figure 3.7. It means that all our information on random events comes from observing X.

We will later discuss American-style options in detail. For these we need to consider random times which are only random in the sense that they depend on random events in the past. They are called *stopping times*.

Definition 3.5. A *stopping time* is a random variable τ with values in $\{0, 1, 2, \dots, \infty\}$ which satisfies $\{\tau = n\} \in \mathscr{F}_n$ (or equivalently $\{\tau \leq n\} \in \mathscr{F}_n$) for any $n \in \mathbb{N}$.

Intuitively this means that the decision whether we say "Stop!" at time n can be based on the information that is available a time n. The time of a vulcanic eruption is a stopping time if observing the vulcano belongs to the information encoded in the filtration $\mathscr{F}_0, \mathscr{F}_1, \dots$ Any deterministic (i.e. non-random) time is a stopping time as well. The time *exactly three hours before the vulcanic eruption* is not a stopping time because we cannot look into the future. We cannot say for sure whether this point in time has come or not.

A typical instance of a stopping time is the first time when a process enters a given set, e.g. the first time a stock index exceeds a given threshold. In Figure 3.9, τ corresponds to the first time that process X exceeds the level 10.5.

Lemma 3.6. *If X denotes an adapted process and B a (measurable) set of real numbers, then $\tau := \inf\{n \in \mathbb{N} : X_n \in B\}$ is a stopping time.*

τ in the previous definition is the first time n such that X_n is in B. If this does not happen at all, τ attains the value ∞. Note that the argument ω representing randomness is suppressed in most formulas.

We also need a mathematical notion for "freezing" a stochastic process at a particular time.

Definition 3.7. If X denotes an adapted process, we call the process X^τ with $X_n^\tau := X_{\tau\wedge n}$ *the process stopped at time* τ, where $\tau \wedge n := \min\{\tau, n\}$ denotes the minimum of τ and n.

The stopped process X^τ coincides with X up to time τ and stays constant afterwards.

Martingales

The *martingale* is a key concept in stochastic analysis. It turns out be pivotal for mathematical finance as well.

Definition 3.8. A *martingale* (resp. *submartingale, supermartingale*) is an adapted stochastic process X such that

$$E(X_n|\mathscr{F}_m) = X_m \quad \text{(resp.} \geq X_m, \leq X_m) \tag{3.1}$$

for any $m \leq n$.

For a martingale we expect on average the present value for the future, cf. Figure 3.10. In the submartingale (resp. supermartingale) case we expect at least (resp. at most) the present value.

The wealth in a fair game follows a martingale. As an example consider a roulette game, where the stake doubles in case that *red* turns up. We play several rounds betting €1 on *red* and we denote the wealth process as X. If *red* turns up with probability $\frac{1}{2}$, our wealth stays on average the same after each round, i.e. X is a martingale. For the (bounded) future we expect neither a profit nor a loss on average.

Real roulette games are biased in favour of the casino because *red* appears only with probability $\frac{18}{37}$. Hence we are rather facing a supermartingale. By contrast, stocks and other securities are commonly believed to behave as submartingales, at least if we adjust for dividends. Economic theory claims that investors would not take the risk involved in such assets unless they are compensated by a positive return on average.

Lemma 3.9. *In the previous definition it suffices to show*

$$E(X_n|\mathscr{F}_{n-1}) = X_{n-1} \quad \textit{(resp.} \geq, \leq)$$

or, equivalently,

$$E(\Delta X_n|\mathscr{F}_{n-1}) = 0 \quad \textit{(resp.} \geq, \leq)$$

for any $n \geq 1$ if one wants to prove the (sub-, super-)martingale property of X.

Proof. This follows by induction from $E(X_n|\mathscr{F}_{n-2}) = E(E(X_n|\mathscr{F}_{n-1})|\mathscr{F}_{n-2})$ etc., where we used Lemma 2.A.4(4) or Theorem 2.B.10(4). For sub- and supermartingales we additionally need monotonicity according to Lemma 2.A.4(6) or Theorem 2.B.10(5). □

The simplest martingales are obtained by successively adding independent, centered random variables.

Example 3.10. Denote by $X_1, X_2, \dots$ a sequence of independent random variables with expected value $E(X_n) = 0$ for any n. Then $S_n := \sum_{m=1}^{n} X_m$ defines a martingale relative to the filtration that is generated by X (or equivalently S). Here one uses Lemma 2.A.4 or Theorem 2.B.10 again.

A random variable generates a martingale in a canonical way.

Lemma 3.11. *If Y is a random variable and $(\mathscr{F}_n)_{n\in\mathbb{N}}$ a filtration, then*

$$X_n := E(Y|\mathscr{F}_n)$$

defines a martingale X, the martingale generated by Y *or* Doob martingale.

Proof. X is adapted by Lemma 2.A.3 or Theorem 2.B.6. The martingale property follows from $E(X_n|\mathscr{F}_m) = E(E(Y|\mathscr{F}_n)|\mathscr{F}_m) = E(Y|\mathscr{F}_m) = X_m$ for $m \leq n$, where we used Lemma 2.A.4(4) or Theorem 2.B.10(4). □

One may wonder whether any martingale is generated by a random variable, which means that one can basically identify martingales with their generating random variables. This is indeed the case in our present setup of finite probability spaces. In the more general case of infinite spaces it is not the case.

Example 3.12. Let $Q \sim P$ denote another probability measure. The martingale Z generated by the density $\frac{dQ}{dP}$, with conditional expectations computed with respect to P, is called *density process of Q relative to P*. The random variable Z_n is the density of Q relative to P, both restricted to $\mathscr{F}_n$. This means that

$$Q(A) = E(Z_n 1_A)$$

holds for any event $A \in \mathscr{F}_n$. More generally, we have

$$E_Q(X) = E(Z_n X)$$

for any $\mathscr{F}_n$-measurable random variable X.

An example is provided in Figure 3.11.

For a martingale, the expected future value is the present value. A more general process could exhibit a positive, negative, or varying trend. This is formalized in terms of *Doob's decomposition*. It decomposes the increment ΔX_n into a predictable trend ΔA_n and a random deviation ΔM_n from that trend.

Theorem 3.13 (Doob decomposition). *An adapted process X can be decomposed uniquely in the form*

$$X = X_0 + M + A$$

where M is a martingale with $M_0 = 0$ and A a predictable process with $A_0 = 0$. A is called compensator *of X.*

For an illustration cf. Figure 3.12. The one-period prediction

$$\Delta A_n := E(\Delta X_n | \mathscr{F}_{n-1}) \tag{3.2}$$

of the increment of X can be interpreted as the present trend. The cumulative sum $A_n := \sum_{m=1}^n \Delta A_m$ of these trends is the compensator A from the previous theorem. Its meaning is less obvious. The difference of X and its compensator A is a martingale because it increments $\Delta M_n := \Delta X_n - \Delta A_n$ have conditional expectation 0.

If X is a submartingale (resp. supermartingale), then its compensator A is increasing (resp. decreasing).

Stochastic integral

As before we consider a filtered probability space $(\Omega, \mathfrak{P}(\Omega), (\mathscr{F}_n)_{n\in\mathbb{N}}, P)$. We assume that all outcomes happen with non-zero probability. The stochastic integral is a key concept in stochastic analysis. It is nothing but a sum in our discrete-time setup.

Definition 3.14. Let X denote an $\mathbb{R}^d$-valued adapted process and H an $\mathbb{R}^d$-valued predictable (or at least adapted) process. We call the adapted process $H \bullet X$ of the form

$$H \bullet X_n := \sum_{m=1}^{n} H_m^\top \Delta X_m, \tag{3.3}$$

the *stochastic integral* of H relative to X, with $H \bullet X_0 = 0$ as the sum is empty then. Here, $H_m^\top \Delta X_m := \sum_{i=1}^d H_m^i \Delta X_m^i$ denotes the scalar product of the vectors $H_m = (H_m^1, \dots, H_m^d)^\top$ and $\Delta X_m = (\Delta X_m^1, \dots, \Delta X_m^d)^\top$. In the univariate case $d = 1$, we simply have $H \bullet X_n = \sum_{m=1}^n H_m \Delta X_m$.

An example is provided in Figure 3.13. Note that the dot is the symbol for the operation on the right-hand side of (3.3) and does not denote a simple product. The name *integral* is motivated from its continuous-time extension, in which case the sum indeed generalizes to an integral. Its relevance for mathematical finance stems from the fact that it stands for financial gains. Suppose that X stands for the stock price evolution and H for the investor's position through time. More specifically, let X_n denote the stock price at time n and H_n the number of shares held in the period from $n-1$ to n. Due to stock price changes, the investor makes a profit of $H_n(X_n - X_{n-1}) = H_n \Delta X_n$ in this period. Consequently, the integral $H \bullet X_n$ stands for the cumulative gais resp. losses between times 0 und n, as they are due to price changes (and not from buying resp. selling assets).

If we consider a portfolio of d stocks, both X and H turn vector-valued. In this case, X_n^i denotes the price at time n of stock No. i. Accordingly, H_n^i denotes the number of shares of stock No. i in the portfolio in the period from $n-1$ to n. In order to compute the cumulative gains $H \bullet X_n$ from 0 to n we now have to sum up over all stocks $i = 1, \dots, d$ and all periods $m = 1, \dots, n$.

In order to get the bookkeeping right, we need to be careful about the order in which things happen at time n. If we interpret $H_n(X_n - X_{n-1})$ as profit at time n, this implies that we have acquired H_n shares *before* the stock price changed from X_{n-1} to X_n. Put differently, the shares are bought at the end of period $n-1$, after the stock price settled at X_{n-1}. This implies that only the information up to time $n-1$ can be used when H_n is chosen because time n and in particular the stock price X_n still belongs to the unknown future. This motivates why we typically assume the process H in the above definition to be predictable.

Definition 3.15. If X, Y are adapted processes, we denote by

$$[X, Y]_n := \sum_{m=1}^{n} \Delta X_m \Delta Y_m,$$

the *covariation* process $[X, Y]$ of X and Y, again with $[X, Y]_0 = 0$. It is called *quadratic variation* of X if $X = Y$.

The covariation process is primarily needed for calculations as in the integration by parts rule below (cf. Lemma 3.16). It does not have an obvious interpretation in terms of financial mathematics. A large covariation $[X, Y]$ means that X and Y tend to move in the same direction. A large quadratic variation $[X, X]$ occurs if the process changes a lot. Both terms therefore can loosely be regarded as some sort of random (co)variance.

The following rules are helpful in the context of stochastic integration. As noted before, we assume $d = 1$ for simplicity.

Lemma 3.16. *Let X, Y be adapted and H, K predictable processes.*

1. $H \bullet (K \bullet X) = (HK) \bullet X$
2. $[H \bullet X, Y] = H \bullet [X, Y]$
3. Integration by parts*:*

$$\begin{aligned} XY &= X_0Y_0 + X_- \bullet Y + Y \bullet X & (3.4)\\ &= X_0Y_0 + X_- \bullet Y + Y_- \bullet X + [X, Y] & (3.5) \end{aligned}$$

4. *If X is a martingale so is $H \bullet X$.*
5. *If X is a supermartingale and $H \geq 0$ then $H \bullet X$ is a supermartingale as well.*
6. *If X is a martingale and τ a stopping time then X^τ is a martingale as well.*
7. *If X is a supermartingale and τ a stopping time then X^τ is a supermartingale as well.*

The last four rules mean that fair resp. unfavourable games cannot be turned into favourable ones by clever stopping or trading. The restriction $H \geq 0$ in Statement 6 effectively stands for short sale restrictions. If short sales are allowed, one could profit from decreasing prices.

The integration by parts rule (3.5) can be illustrated by considering a US stock whose price X is quoted in US dollars. If Y denotes the dollar exchange rate, more precisely the price of a US dollar in Euro, then XY represents the stock price quoted in Euro. Its changes may be due to changes in the dollar stock price or due to changes in the exchange rate. The two effects are expressed by $Y_- \bullet X$ resp. $X_- \bullet Y$ in (3.5). The less intuitive term $[X, Y]$ indicates that another contribution occurs if X and Y change at the same time. We will come back to this issue after Lemma 3.21.

Proof. 1. $H \bullet (K \bullet X)_n = \sum_{m=1}^n H_m \Delta(K \bullet X)_m = \sum_{m=1}^n H_m K_m \Delta X_m = (HK) \bullet X_n$

2.

$$\begin{aligned} [H \bullet X, Y]_n &= \textstyle\sum_{m=1}^n \Delta(H \bullet X)_m \Delta Y_m \\ &= \textstyle\sum_{m=1}^n H_m \Delta X_m \Delta Y_m \\ &= \textstyle\sum_{m=1}^n H_m \Delta[X, Y]_m \\ &= H \bullet [X, Y]_n \end{aligned}$$

3.

$$\begin{aligned} X_n Y_n &= X_0 Y_0 + \textstyle\sum_{m=1}^n (X_m Y_m - X_{m-1} Y_{m-1}) \\ &= X_0 Y_0 + \textstyle\sum_{m=1}^n (X_{m-1}(Y_m - Y_{m-1}) + Y_m(X_m - X_{m-1})) \\ &= X_0 Y_0 + X_- \bullet Y_n + Y \bullet X_n \end{aligned}$$

and

$$\begin{aligned} Y \bullet X_n &= \textstyle\sum_{m=1}^n Y_m \Delta X_m \\ &= \textstyle\sum_{m=1}^n (Y_{m-1} \Delta X_m + (Y_m - Y_{m-1}) \Delta X_m) \\ &= Y_- \bullet X_n + [X, Y]_n \end{aligned}$$

4. $H \bullet X$ is again an adapted process. The martingale property follows from

$$\begin{aligned} E(H \bullet X_n | \mathscr{F}_{n-1}) &= E(H \bullet X_{n-1} + H_n \Delta X_n | \mathscr{F}_{n-1}) \\ &= H \bullet X_{n-1} + H_n E(\Delta X_n | \mathscr{F}_{n-1}) \\ &= H \bullet X_{n-1} + H_n (E(X_n | \mathscr{F}_{n-1}) - X_{n-1}) \\ &= H \bullet X_{n-1} \end{aligned}$$

where we have used Lemma 2.A.4 or Theorem 2.B.10 several times.

5. This follows along the same lines as 5.

6.,7. The stopped process can be written as a stochastic integral of the form $X^\tau = X_0 + H \bullet X$ with $H_n := 1_{\{\tau \geq n\}}$. H is predictable because $\{H_n = 1\} = \{\tau \geq n\} = \{\tau \leq n - 1\}^C \in \mathscr{F}_{n-1}$. The identity for X^τ is clear if it is interpreted in terms of financial gains: $X^\tau - X_0$ stands for the profit if the stock is held until time τ. Since H equals 1 until τ and 0 afterwards, the expression $H \bullet X$ represents the same thing. The assertion follows from Statements 5 and 6. $\square$

Remark.

1. The stochastic integral $H \bullet X$ is linear in H and X.
2. The covariation $[X, Y]$ is linear in both X and Y.
3. The above rules hold for vector-valued processes as well if they make sense, e.g. $H \bullet (K \bullet X) = (HK) \bullet X$ if K, X are $\mathbb{R}^d$-valued and H is $\mathbb{R}$-valued.

From the previous results, we immediately get the following corollary which is important in many applications and illustrates again that it is not possible to beat a martingale:

Theorem 3.17 (Optional Sampling Theorem)**.** *Let M be a martingale and τ a bounded stopping time*[1]*. Then,*

$$E(M_\tau) = E(M_0).$$

Analogously, it holds for all supermartingales X that

$$E(X_\tau) \leq E(X_0)$$

and for submartingales

$$E(X_\tau) \geq E(X_0).$$

Proof. We just prove the martingale case. Let N be an integer such that $\tau \leq N$. By 7. in the previous lemma, it holds that

$$E(M_\tau) = E(M_{\tau\wedge N}) = E(M^\tau_N) = E(E(M^\tau_N|\mathscr{F}_0)) = E(M^\tau_0) = E(M_0).$$

$\square$

The most important rule in continuous-time stochastic calculus is Itô's formula. In discrete time it reduces to a simple telescoping sum and it does not play an important role. It is stated here only for the sake of completeness.

Theorem 3.18 (Itô's formula)**.** *If X denotes an adapted process and $f : \mathbb{R} \to \mathbb{R}$ a differentiable function, then the adapted process $f(X)$ satisfies*

$$\begin{aligned} f(X_n) &= f(X_0) + \sum_{m=1}^{n} (f(X_m) - f(X_{m-})) \\ &= f(X_0) + f'(X_-) \bullet X_n + \sum_{m=1}^{n} \Big(f(X_m) - f(X_{m-}) - f'(X_{m-})\Delta X_m \Big) \end{aligned}$$

Proof. The first equation follows because almost all terms in the telescopic sum cancel. The second equation holds because $f'(X_-) \bullet X_n = \sum_{m=1}^n f'(X_{m-})\Delta X_m$. $\square$

[1] i.e. there exists an integer N such that $\tau(\omega) \leq N$ for all ω

Remark. If the jumps ΔX are small and f is twice continuously differentiable, we have the approximation

$$f(X_n) \approx f(X_0) + f'(X_-) \bullet X_n + \frac{1}{2} f''(X_-) \bullet [X,X]_n \tag{3.6}$$

An explanation is as follows: A sufficiently smooth function can be approximated by its second order Taylor polynomial $f(x+h) \approx f(x) + f'(x)h + \frac{1}{2}f''(x)h^2$ if h is small enough. In our setup this yields

$$\begin{aligned} f(X_m) &= f(X_{m-1} + \Delta X_m) \\ &\approx f(X_{m-1}) + f'(X_{m-1})\Delta X_m + \frac{1}{2} f''(X_{m-1})(\Delta X_m)^2 \end{aligned}$$

and hence

$$\begin{aligned} \sum_{m=1}^{n} \Big(f(X_m) - f(X_{m-}) - f'(X_{m-})\Delta X_m \Big) &\approx \frac{1}{2} \sum_{m=1}^{n} f''(X_{m-})\Delta[X,X]_m \\ &= \frac{1}{2} f''(X_-) \bullet [X,X]_n. \end{aligned}$$

Stochastic exponentials are processes of multiplicative structure, which play an important role in stochastic calculus. They have an obvious financial interpretation. If ΔX_n denotes a (random) interest rate in period n (i.e. each Euro at time $n-1$ turns into €$1 + \Delta X_n$ at time n), then the process $\mathscr{E}(X)$ describes how an initial capital of €1 grows by interest and compound interest. Conversely, if the price of an asset can be written as $S = \mathscr{E}(X)$, then

$$\Delta X_n = \frac{\Delta S_n}{S_{n-1}}$$

is the *return* of this asset in period n. In that sense, X can be called *return process* of S.

Definition 3.19. If X is an adapted process, then the *stochastic exponential* is the unique adapted process Z satisfying

$$Z = 1 + Z_- \bullet X$$

(which is equivalent to $Z_0 = 1$ and $\Delta Z_n = Z_{n-1}\Delta X_n$). We denote it as $\mathscr{E}(X) = Z$.

An example can be found in Figure 3.15. The stochastic exponential has a simple explicit representation.

Lemma 3.20. *We have* $\mathscr{E}(X)_n = \prod_{m=1}^{n}(1 + \Delta X_m)$.

Proof. The formula holds for $n = 0$ trivially, as an empty product is defined to be one. For $n \geq 1$, if we set $Z_n := \prod_{m=1}^{n}(1 + \Delta X_m)$, we conclude

$$Z_n = Z_{n-1} + Z_{n-1}\Delta X_n$$

as desired. □

Remark.

1. If the jumps ΔX are small, we may use the Taylor expansion $\log(1+x) \approx x - \frac{1}{2}x^2$ to obtain the approximation

$$\begin{aligned} \mathscr{E}(X)_n &= \exp\left(\sum_{m=1}^{n} \log(1+\Delta X_m)\right) \\ &\approx \exp\left(\sum_{m=1}^{n} (\Delta X_m - \frac{1}{2}(\Delta X_m)^2)\right) = \exp\left(X_n - X_0 - \frac{1}{2}[X,X]_n\right). \end{aligned} \tag{3.7}$$

2. If X is a martingale, so is $\mathscr{E}(X)$. This follows from the fact that the latter is — up to an additional constant — an integral relative to a martingale.

3. If c denotes a constant, then $c\mathscr{E}(X)$ is the unique process Z satisfying $Z = c + Z_- \bullet X$.

The following rule for stochastic exponentials resembles a similar statement for the ordinary exponential function. But by contrast to $e^X e^Y = e^{X+Y}$, we have an additonal term involving the covariation.

Lemma 3.21 (Yor's formula)**.** *For adapted processes X, Y we have*

$$\mathscr{E}(X)\mathscr{E}(Y) = \mathscr{E}(X + Y + [X,Y]).$$

Proof. By Lemma 3.16 we have

$$\begin{aligned} \mathscr{E}(X)\mathscr{E}(Y) &= \mathscr{E}(X)_0\mathscr{E}(Y)_0 + \mathscr{E}(X)_- \bullet \mathscr{E}(Y) + \mathscr{E}(Y)_- \bullet \mathscr{E}(X) + [\mathscr{E}(X), \mathscr{E}(Y)] \\ &= 1 + (\mathscr{E}(X)_-\mathscr{E}(Y)_-) \bullet Y + (\mathscr{E}(X)_-\mathscr{E}(Y)_-) \bullet X + (\mathscr{E}(X)_-\mathscr{E}(Y)_-) \bullet [X,Y] \\ &= 1 + (\mathscr{E}(X)\mathscr{E}(Y))_- \bullet (Y + X + [X,Y]), \end{aligned}$$

which yields the claim. □

Let us illustrate Yor's formula by coming back to the example illustrating integration by parts. Suppose that $\mathscr{E}(X)$ denotes the price of a stock, expressed in US dollar. If $\mathscr{E}(Y)$ stands for the price in Euro of a US dollar, then $\mathscr{E}(X)\mathscr{E}(Y) = \mathscr{E}(X+Y+[X,Y])$ represents the price of the stock in Euro. Yor's formula means that the return $\Delta X + \Delta Y + \Delta[X,Y]$ differs from the sum of the returns of $\mathscr{E}(X)$ and $\mathscr{E}(Y)$ by the covariation term $\Delta[X,Y]$. E.g. if the stock price in dollar grows by 3% and the dollar exchange rate increases by 3% as well, then the stock price in Euro grows by 6.09% rather than 6%.

Appendix QF

3.A Intuitive summary of some notions from this chapter

A filtration on a probability space can be represented as a tree (or several trees if $\mathscr{F}_0$ is not trivial which means that there is no unique root), cf. Figures 3.1-3.3. A filtered probability space is such a tree together with probabilities. These can be specified as unconditonal probabilities of all outcomes or as transition probabilities on all edges of the tree, cf. Figure 3.4. Suppose first that the unconditonal probabilities of all outcomes are given. The probabilities of the events corresponding to the vertices are obtained by adding the probabilities of all outcomes which are descendents of any particular vertex under consideration. The transition probabilities on the edges are then obtained as the ratio of the probabilities of the adjacent vertices, more specifically child probability over parent probability. If, on the other hand, only the transition probabilities on all edges are given, one can compute the probabilities of the vertices by multiplying all transition probabilities on the edges leading from the root to the vertex under consideration.

Specifying an adapted process means to assign numbers resp. vectors to all vertices in the tree, cf. Figure 3.8. A predictable process has the same value on all children of any vertex. Sometimes, it is useful to write this value X_n at the corresponding parent vertex at time $n-1$. A stopping time means to cut the tree at certain vertices, cf. Figure 3.9. Stopping a process means to change the value at all following vertices to the value where the tree was cut.

A martingale is a process where the value at any parent vertex is the conditional expectation of the values at the child vertices, cf. Figure 3.10. For super- resp. submartingales we have inequalities instead of equality in each vertex.

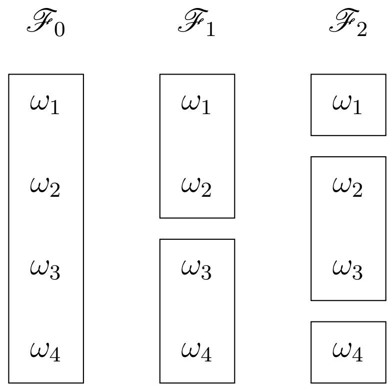

Figure 3.1: σ-fields on $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, no filtration

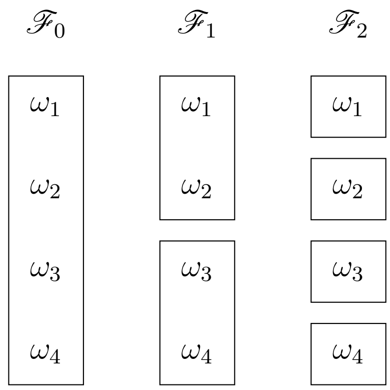

Figure 3.2: Filtration on $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$

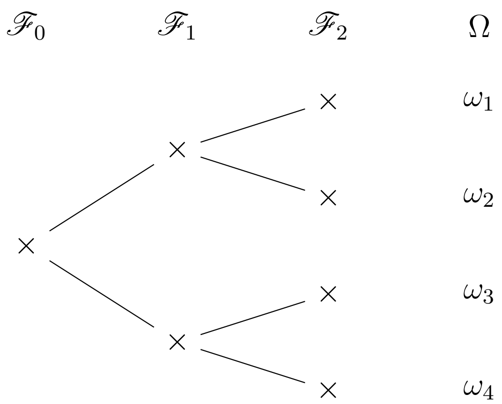

Figure 3.3: Alternative representation of Figure 3.2

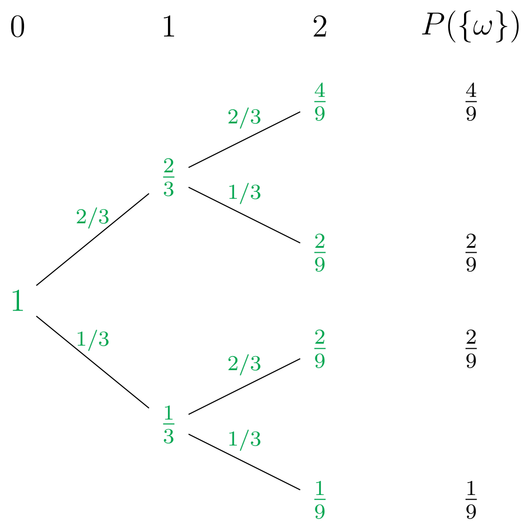

Figure 3.4: Filtered probability space, unconditional and transition probabilities

	$\mathscr{F}_0$	$\mathscr{F}_1$	$\mathscr{F}_2$
$X_n(\omega_1)$	1	2	3
$X_n(\omega_2)$	2	3	3
$X_n(\omega_3)$	2	2	4
$X_n(\omega_4)$	2	2	4
	$n{=}0$	$n{=}1$	$n{=}2$

Figure 3.5: Non-adapted process X

	$\mathscr{F}_0$	$\mathscr{F}_1$	$\mathscr{F}_2$
$X_n(\omega_1)$	1	2	4
$X_n(\omega_2)$	1	2	4
$X_n(\omega_3)$	1	3	4
$X_n(\omega_4)$	1	3	5
	$n{=}0$	$n{=}1$	$n{=}2$

Figure 3.6: Adapted process X

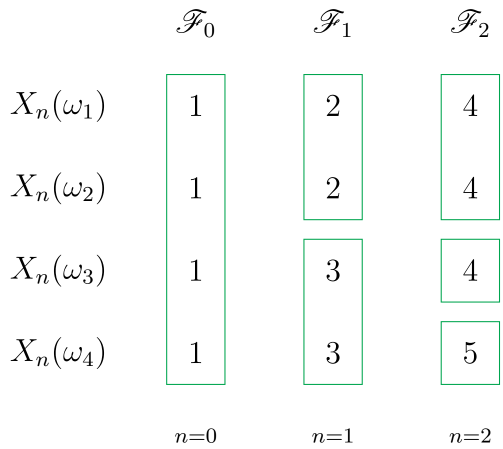

Figure 3.7: Filtration generated by X

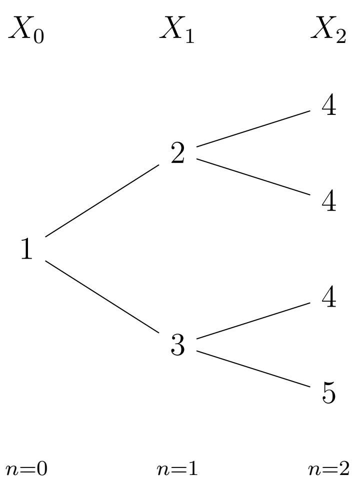

Figure 3.8: Alternative representation of Figure 3.6

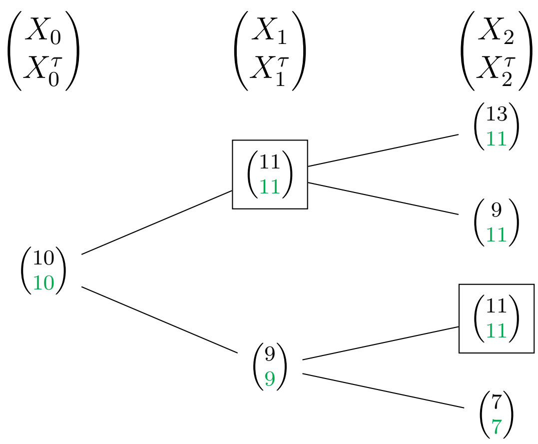

Figure 3.9: Process X, stopping time τ, and stopped process X^τ

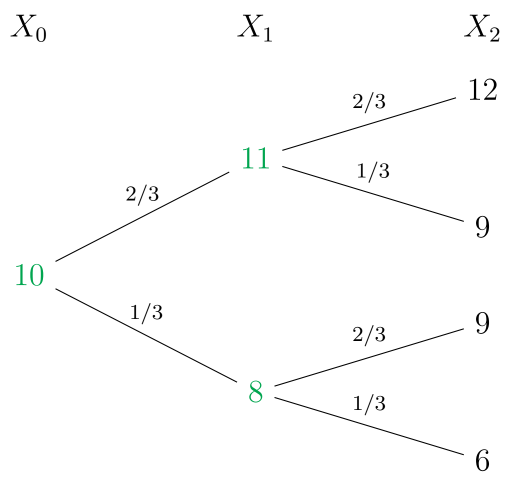

Figure 3.10: Martingale X

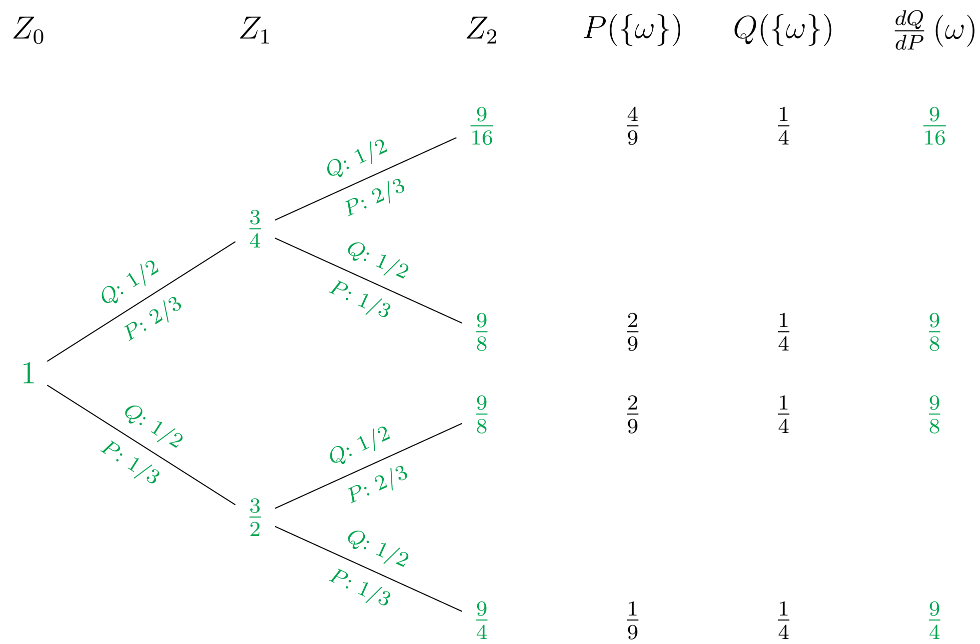

Figure 3.11: Probability measures P, Q and density process Z

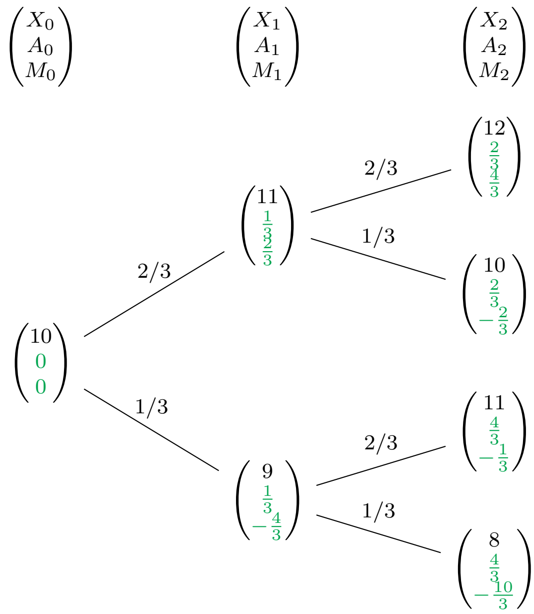

Figure 3.12: Doob decomposition of X

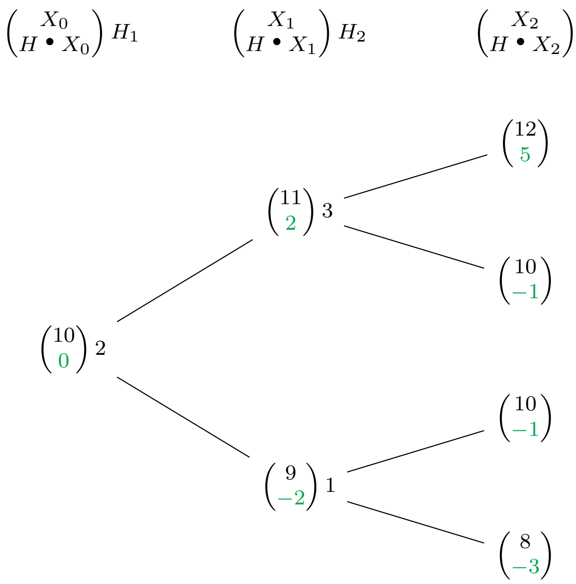

Figure 3.13: Stochastic integral $H \bullet X$

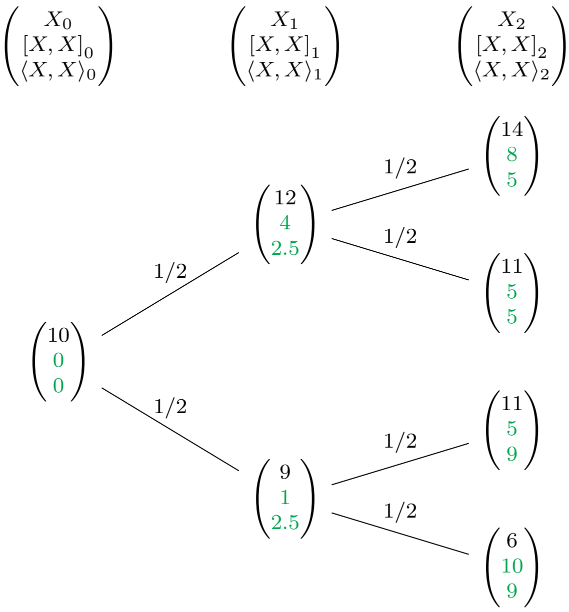

Figure 3.14: Quadratic variation (and predictable quadratic variation, which will be introduced later and can be ignored now) of X

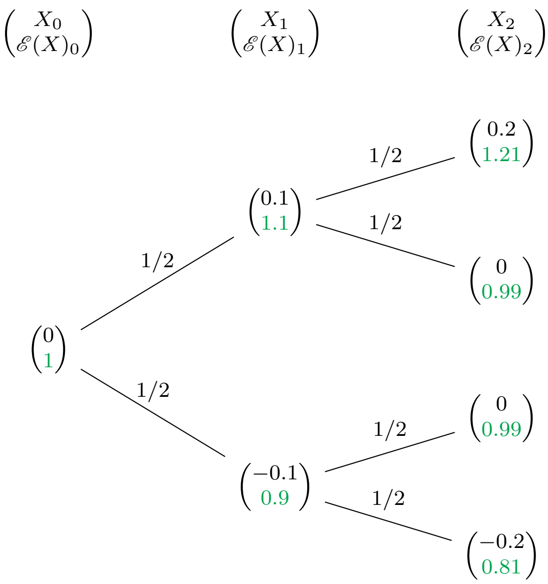

Figure 3.15: Stochastic exponential of X

Appendix MF

3.B Some more results

The density process is useful in order to compute conditional expectations relative to a new probability measure:

Lemma 3.B.1 (Generalized Bayes formula)**.** *Let $Q \sim P$ denote a probability measure with density process Z. Moreover, let X denote an $\mathscr{F}_n$-measurable random variable for some n. Then*

$$E_Q(X|\mathscr{F}_m) = \frac{E(XZ_n|\mathscr{F}_m)}{Z_m}$$

for any $m \leq n$.

The proof is easy and is left as an (optional) exercise.

Martingales do not remain martingales if the probability measure is changed. The following result shows how the martingale property relative to some new probability measure Q can be expressed in terms of the density process of Q relative to P. Such measure changes play an important role in financial mathematics.

Lemma 3.B.2. *Let $Q \sim P$ be a probability measure with density process Z. An adapted process X is a Q-martingale if and only if XZ is a P-martingale.*

Proof. According to the generalized Bayes rule (Lemma 3.B.1) we have

$$Z_{n-1}E_Q(X_n|\mathscr{F}_{n-1}) = E_P(Z_nX_n|\mathscr{F}_{n-1}).$$

X is a Q-martingale if and only if the left-hand side coincides with $Z_{n-1}X_{n-1}$. Similarly, ZX is a P-martingale if and only if the right-hand side coincides with $Z_{n-1}X_{n-1}$. $\square$

Definition 3.B.3. Let X, Y be adapted processes. The compensator of $[X, Y]$ is called *predictable covariation* of X and Y and it is denoted as $\langle X, Y\rangle$. For $X = Y$ it is called *predictable quadratic variation* of X.

Note that

$$\Delta\langle X, Y\rangle_n = E(\Delta[X, Y]_n|\mathscr{F}_{n-1}) = E(\Delta X_n \Delta Y_n|\mathscr{F}_{n-1}) \tag{3.8}$$

and hence

$$\langle X, Y\rangle_n = \sum_{m=1}^{n} E(\Delta X_m \Delta Y_m|\mathscr{F}_{m-1}) \tag{3.9}$$

by (3.2). The predictable covariation can be interpreted as a generalization of the covariance for stochastic processes. It is used e.g. in Girsanov's theorem below. It does not have an obvious interpretation in terms of financial mathematics either. An example for both quadratic variation and predictable quadratic variation can be found in Figure 3.14.

We sometimes need the following.

Lemma 3.B.4. *Let X, Y be adapted and H a predictable process. Then*

$$\langle H \bullet X, Y\rangle = H \bullet \langle X, Y\rangle.$$

Proof. It suffices to show that the increments of both sides coincide.

$$\begin{aligned}
\Delta(H \bullet \langle X, Y\rangle)_n &= H_n \Delta\langle X, Y\rangle_n \\
&\overset{(3.8)}{=} H_n E(\Delta[X, Y]_n|\mathscr{F}_{n-1}) \\
&\overset{\text{Theorem } 2.B.10\ 9.}{=} E(H_n \Delta[X, Y]_n|\mathscr{F}_{n-1}) \\
&\overset{\text{Lemma } 3.16\ 2.}{=} E(\Delta[H \bullet X, Y]_n|\mathscr{F}_{n-1}) \\
&\overset{(3.9)}{=} \Delta\langle H \bullet X, Y\rangle_n
\end{aligned}$$

□

Since a P-martingale does not generally remain a martingale under Q, it exhibits a trend relative to Q. The Doob decomposition of X relative to Q can be expressed in terms of the density process of Q.

Lemma 3.B.5 (Girsanov's theorem)**.** *Let $Q \sim P$ be a probability measure with density process Z. If X is a P-martingale,*

$$X - \frac{1}{Z_-} \bullet \langle Z, X\rangle$$

is a Q-martingale, where the predictable covariation is to be interpreted relative to P.

Proof. $A := \frac{1}{Z_-} \bullet \langle Z, X\rangle$ is a predictable process with $A_0 = 0$. Applying Lemma 3.16 several times proves that

$$\begin{aligned}
(X - A)Z &= XZ - AZ \\
&= X_0 Z_0 + X_- \bullet Z + Z_- \bullet X + [Z, X] - Z_- \bullet A - A \bullet Z \\
&= X_0 Z_0 + X_- \bullet Z + Z_- \bullet X + ([Z, X] - \langle Z, X\rangle) - A \bullet Z
\end{aligned}$$

is a P-martingale. The assertion follows now from Lemma 3.B.2. □

3.B.1 A fun application

We consider the following game: We draw cards (one after the other) from a deck of N cards which contains M red and $N - M$ black cards. We can stop whenever we think that the next card is red and win 1 € if our prediction is correct. Clearly, if we stop at the very beginning, then our probability of winning is M/N. But can we do better?

To model the situation, we can take Ω the set of all permutations of $1, ..., N$ with the Laplace measure. Let $C_1, C_2, ..., C_N$ be random variables with values in $\{0, 1\}$ where $\{C_n = 1\}$ denotes the event that the n-th card is red. The filtration is given by $\mathscr{F}_n = \sigma(C_1, .., C_n)$. When stopping at time n, we win if $C_{n+1} = 1$. The strategies we use are stopping times $\tau : \Omega \to \{0, ..., N-1\}$ and our probability of winning is $P(C_{\tau+1} = 1) = E(C_{\tau+1})$.

As C_{n+1} is not known at time n, it is natural to consider $R_n := E(C_{n+1}|\mathscr{F}_n)$. Then,

$$\begin{aligned} E(C_{\tau+1}) &= \sum_{k=0}^{N-1} E(C_{k+1} 1_{\{\tau=k\}}) \\ &= \sum_{k=0}^{N-1} E(E(C_{k+1} 1_{\{\tau=k\}}|\mathscr{F}_k)) \\ &= \sum_{k=0}^{N-1} E(1_{\{\tau=k\}} E(C_{k+1}|\mathscr{F}_k)) \\ &= \sum_{k=0}^{N-1} E(1_{\{\tau=k\}} R_k) \\ &= E(R_\tau). \end{aligned}$$

Hence, the question is whether we can find a stopping time τ such that $E(R_\tau) > M/N$.

We now show that the answer is *No* by showing that R is a martingale, so that by the Optional Sampling Theorem 3.17 we have $E(R_\tau) = E(R_0) = M/N$ for all τ.

First note that

$$R_n = E(C_{n+1}|\mathscr{F}_n) = P(C_{n+1} = 1|C_1, ..., C_n) = \frac{M - \sum_{i=1}^n C_i}{N-n}.$$

Using this, we indeed obtain

$$\begin{aligned} E(R_{n+1}|\mathscr{F}_n) &= E\left(\frac{M - \sum_{i=1}^{n+1} C_i}{N-n-1} \middle| C_1, ..., C_n \right) \\ &= \frac{1}{N-n-1} \left[M - E\left(\sum_{i=1}^{n+1} C_i \middle| C_1, ..., C_n \right) \right] \\ &= \frac{1}{N-n-1} \left[M - \sum_{i=1}^{n} C_i - E\left(C_{n+1} \middle| C_1, ..., C_n \right) \right] \\ &= \frac{1}{N-n-1} \left[M - \sum_{i=1}^{n} C_i - \frac{M - \sum_{i=1}^n C_i}{N-n} \right] \\ &= \frac{M - \sum_{i=1}^n C_i}{N-n-1} \frac{N-n-1}{N-n} \\ &= R_n. \end{aligned}$$

Part II

Financial Modeling: Pricing, Hedging, Portfolio Optimization

Chapter 4

Modelling financial markets

In this chapter we introduce the general mathematical framework for modeling financial markets. Our starting point is as before a fixed filtered probability space $(\Omega, \mathfrak{P}(\Omega), (\mathscr{F}_n)_{n\in\mathbb{N}}, P)$.

Assets and trading strategies

Mathematical Finance considers mostly questions that concern trading securities. Therefore, the mathematical model involves primarily two kinds of stochastic processes: *securities price processes* representing the up and down of quotations, and *trading strategies*, which stand for the investor's portfolio.

We suppose that a fixed adapted process $S = (S^0, \dots, S^d)^\top$ represents the *price process* of the $d+1$ traded assets that are traded in the market. The random variable S_n^i stands for the price of security i at time n. We consider concrete models at the end of this chapter. The assets can be traded. This is expressed in terms of stochastic processes as well.

Definition 4.1. A *trading strategy* (or *portfolio*) is a predictable process $\varphi = (\varphi^0, \dots, \varphi^d)^\top$. The *value process* or *wealth process* of this portfolio is

$$V(\varphi) := \varphi^\top S := \sum_{i=0}^{d} \varphi^i S^i.$$

φ_n^i denotes the number of shares of security i in our portfolio at time n. It is random e.g. in the sense that the investor may choose it depending on the random price changes up to time n. This explains why trading strategies should be adapted. Why do we assume predictability? This has to do with the chronological order of events. In the period from $n-1$ to n prices as well as the portfolio changes. We follow the convention that the portfolio transition from φ_{n-1} to φ_n precedes the securities' price changes from S_{n-1} to S_n. Portfolio φ_n is hence bought at the old prices S_{n-1}. In particular, the information on S_n is not yet available; the investor's decision can only depend on events that have happened up to time $n-1$. This is precisely the essence of predictability.

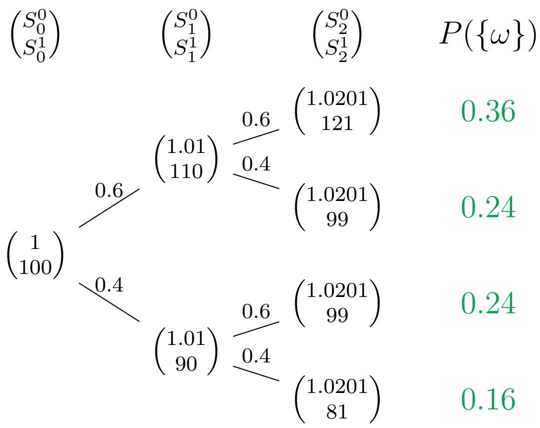

Figure 4.1: A market with two assets (bond and stocks)

The portfolio value $V_n(\varphi)$ is determined after prices have changed, i.e. relative to prices S_n. Note that its definition involves a scalar product, which means that the total value is the sum of the positions in the various securities. Cash itself does not appear in the model. Instead, we consider a bank account as traded security, cf. the subsection on concrete models.

In this introductory course we confine ourselves to an idealized market ("dry water"). We assume that for all securities an arbitrary — even fractional or negative — number of shares can be held at any time. Even though negative positions do not seem to make sense, such *short sales* are in fact possible in real markets under some limitations. Moreover, our mathematical model does not involve any transaction costs and dividend payments; interest rate for debit and credit coincide; prices are not affected by the investor's trade. In fact, this means that we consider neither very large investors, who move prices by their transactions, nor very small ones, who are truly affected by fees and transaction costs. Finally, our framework does not fit to illiquid markets, where the difference between bid and ask prices cannot be neglected.

By contrast to securities prices the portfolio can be chosen according to the investor's preferences. We restrict ourselves to *self-financing* strategies in the following sense.

Definition 4.2. A trading strategy φ is called *self financing*, if

$$(\Delta\varphi_n)^\top S_{n-1} := \sum_{i=0}^{d} \Delta\varphi_n^i S_{n-1}^i = 0$$

for $n = 1, 2, \ldots$

The self-financing condition means that wealth may be redistributed among assets, but no funds are added or withdrawn after initiation at time 0, cf. Figure 4.2 for a numerical example. By expressing cumulative profits and losses in terms of the stochastic integral, the following lemma provides an alternative statement of self-financability: The portfolio value is the initial value plus profits and losses due to price changes.

Lemma 4.3. *A trading strategy φ is self financing if and only if*

$$V(\varphi) = V_0(\varphi) + \varphi \bullet S.$$

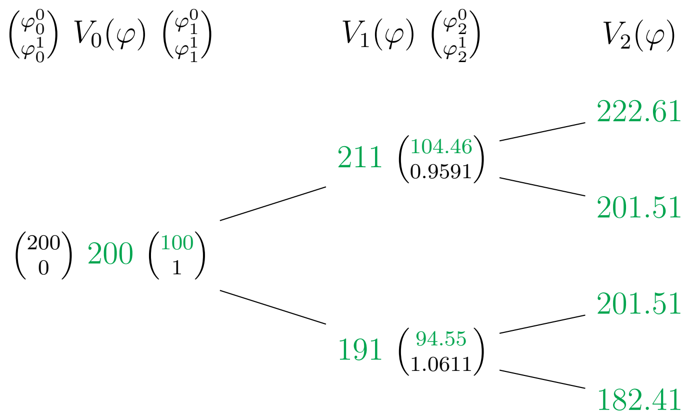

Figure 4.2: A self-financing strategy φ (namely, half of the wealth invested in stock) and its value process in the market of Figure 4.1

Proof. The integration by parts formula (3.4) yields

$$V(\varphi) = \varphi^\top S = V_0(\varphi) + \varphi \bullet S + S_- \bullet \varphi. \tag{4.1}$$

Since the increments of the second integral are $S_{n-1}^\top \Delta\varphi_n$, it vanishes if and only if φ is self financing.

Observe that Equation (4.1) means that the portfolio value is affected by two effects. The first integral stands for profits and lossed due to price changes. The second integral, on the other hand, keeps track of supply and withdrawal of funds after time 0. Self financability means that this second integral vanishes. □

Both bookkeeping and mathematical theory simplify considerably if prices are not expressed in currency but in multiples of a given reference asset, called *numeriare*. Often a particularly simple asset is chosen for this purpose, e.g. a riskless bond with fixed interest rate (called *bond*, *bank account*, or *savings account*). We denote the price process of the numeriare by S^0, which is considered as strictly positive from now on.

Definition 4.4.

$$\hat{S} := \frac{1}{S^0} S = \left(1, \frac{S^1}{S^0}, \dots, \frac{S^d}{S^0}\right)$$

is called *discounted price process*. Moreover,

$$\hat{V}(\varphi) := \frac{1}{S^0} V(\varphi) = \varphi^\top \hat{S}$$

is called *discounted value process* or *discounted wealth process* of strategy φ.

Example 4.5. If the numeraire is of the form $S_n^0 = (1+r)^n$ with fixed deterministic interest rate r, the discounted asset price $\hat{S}_n^i = (1+r)^{-n} S_n^i$ coincides with the *present value* of S_n^i, i.e. the price of S_n^i in terms of currency units as of time 0. In other words, the computation of $\hat{S}$ corresponds to discounting in the usual sense.

The self-financing condition can be expressed in terms of discounted quantities.

Lemma 4.6. *A strategy φ is self financing if and only if*

$$\hat{V}(\varphi) = \hat{V}_0(\varphi) + \varphi \bullet \hat{S}.$$

Proof. φ is self-financing if and only if $(\Delta\varphi_n)^\top \hat{S}_{n-1} = 0$ for any n. The assertion follows from Lemma 4.3 for $\hat{S}$ instead of S. □

Note that the stochastic integral $\varphi \bullet \hat{S}$ for discounted price processes does not depend on the numeraire part φ^0.

The self-financing condition limits the choice of the $d + 1$ assets by a constraint. The following theorem shows that the investor can chose her investment in d securities arbitrarily and that the position in the remaining (e.g. numeraire) security is determined uniquely by this choice.

Lemma 4.7. *For any predictable process $(\varphi^1, \dots, \varphi^d)$ and any $V_0 \in \mathbb{R}$ there exists a unique predictable process φ^0 such that $\varphi = (\varphi^0, \dots, \varphi^d)$ is self financing with $V_0(\varphi) = V_0$.*

Proof. By the previous lemma φ is self-financing if and only if

$$\varphi^0_n \hat{S}^0_n + (\varphi^1, \dots, \varphi^d)^\top_n (\hat{S}^1, \dots, \hat{S}^d)_n = \hat{V}_0 + (\varphi^1, \dots, \varphi^d) \bullet (\hat{S}^1, \dots, \hat{S}^d)_n,$$

i.e. if and only if

$$\begin{aligned} \varphi^0_n &= \hat{V}_0 + (\varphi^1, \dots, \varphi^d) \bullet (\hat{S}^1, \dots, \hat{S}^d)_n - (\varphi^1, \dots, \varphi^d)^\top_n (\hat{S}^1, \dots, \hat{S}^d)_n \\ &= \hat{V}_0 + (\varphi^1, \dots, \varphi^d) \bullet (\hat{S}^1, \dots, \hat{S}^d)_{n-1} - (\varphi^1, \dots, \varphi^d)^\top_n (\hat{S}^1, \dots, \hat{S}^d)_{n-1}. \end{aligned}$$

This is a predictable process. □

Discounting simplifies the bookkeeping in two ways. Firstly, self-financing strategies can be identified in a canonical way with the last d components $(\varphi^1, \dots, \varphi^d)$, which can be arbitrarily chosen. In addition, the last component φ^0 is not needed for the computation of the wealth process $V(\varphi) = S^0 \hat{V}(\varphi) = S^0(V_0 + \varphi \bullet \hat{S})$. Should it be of interest, according to the previous proof it can be calculated e.g. by $\varphi^0_n = \hat{V}(\varphi)_n - (\varphi^1, \dots, \varphi^d)^\top_n (\hat{S}^1, \dots, \hat{S}^d)_n$. From now on, we work primarily with discounted processes.

Notation. Occasionally, we identify $(\hat{S}^1, \dots, \hat{S}^d)$ with $(\hat{S}^0(= 1), \hat{S}^1, \dots, \hat{S}^d)$ and predictable processes $(\varphi^1, \dots, \varphi^d)$ with the self-financing strategies $(\varphi^0, \varphi^1, \dots, \varphi^d)$ from Lemma 4.7 with $V_0 = 0$.

The discounted version of Figure 4.2 is to be found in Figure 4.3.

Arbitrage

In this section we fix a terminal date N, i.e. the time index set is now $\{0, \dots, N\}$ instead of $\mathbb{N}$. In mathematical finance one generally assumes that riskless profits (*Arbitrage*) cannot

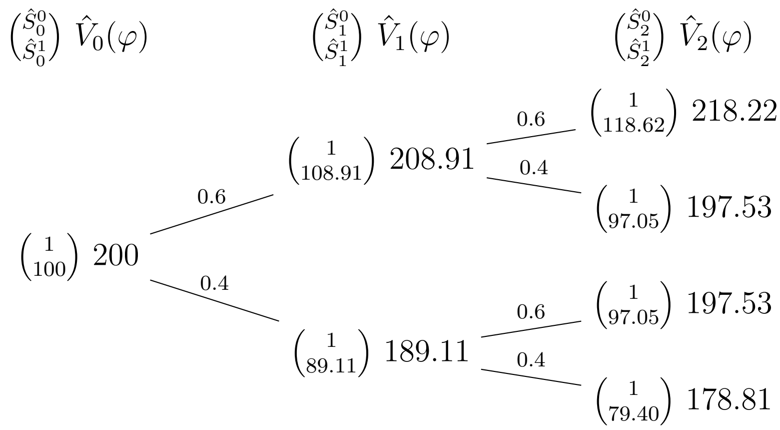

Figure 4.3: Discounted price process $\hat{S}$ and discounted wealth process corresponding to Figures 4.1, 4.2

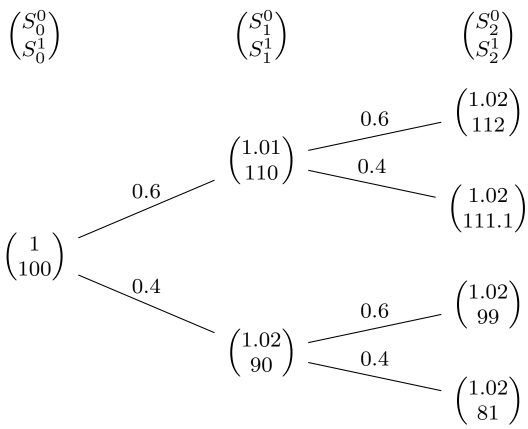

Figure 4.4: A market allowing for arbitrage

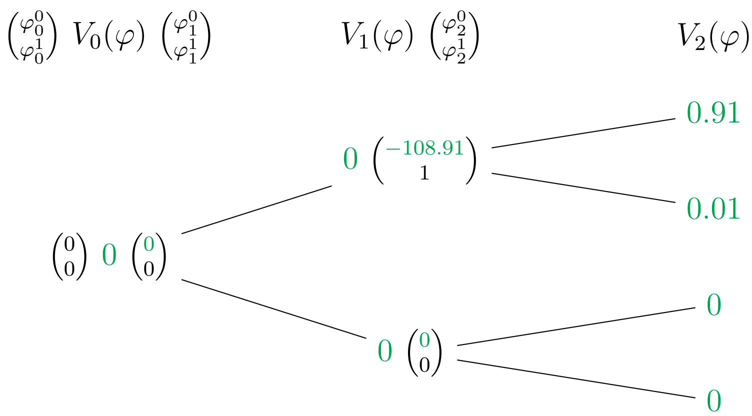

Figure 4.5: An arbitrage strategy φ in the market of Figure 4.4

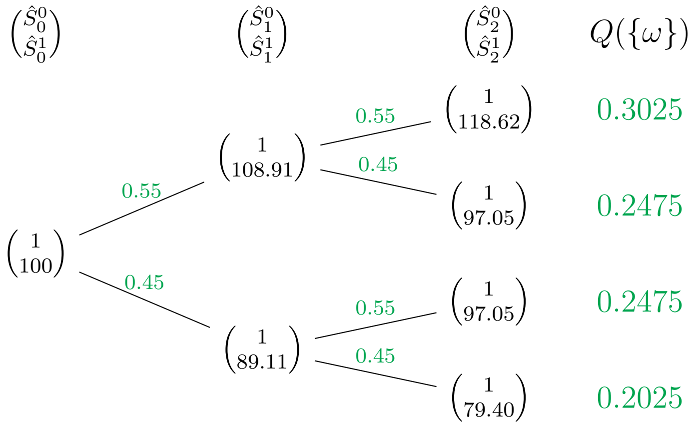

Figure 4.6: Equivalent martingale measure probabilities for the market in Figures 4.1 and 4.3

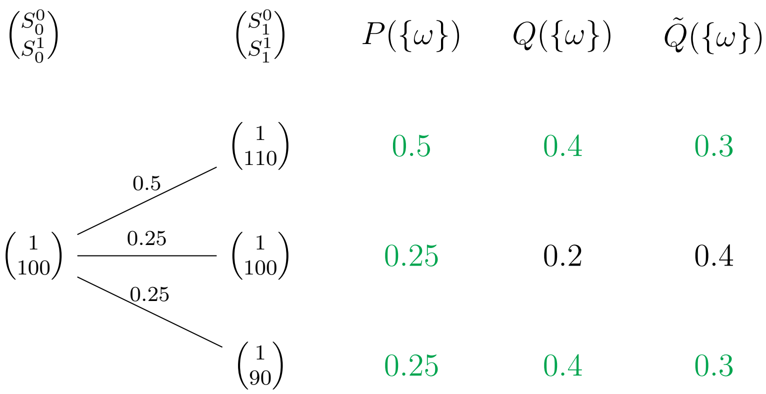

Figure 4.7: A market with two different equivalent martingale measures $Q, \tilde{Q}$

be made without initial capital. This is justified by the argument that such arbitrage opportunities are exploited immediately by so-called arbitrage traders whose presence makes them disappear immediately. Consequently, arbitrage opportunities exist only on small scale and for short times.

And in fact prices and exchange rates differ only marginally if they are quoted at different exchanges. Otherwise, buying and selling an asset at the same time in different places would consitute an arbitrage. In our setup, arbitrage does not arise from transactions in different places but from trading dynamically in time.

Definition 4.8. A self-financing strategy φ is called *arbitrage*, if

$$V_0(\varphi) = 0, \quad V_N(\varphi) \geq 0, \quad P(V_N(\varphi) > 0) > 0,$$

i.e. without any initial endowment one can create a possible profit without venturing any losses. We say that there are no *arbitrage opportunities*, if such strategies do not exist.

The seemingly weak assumption of absence of arbitrage is the foundation of many statements in mathematical finance. Above we argued that it makes prices of the same asset in different places coincide. Here, we need it to conclude that portfolios with the same future value must have the same value today.

Lemma 4.9 (Law of one price)**.** *Denote by φ, ψ self-financing strategies with $V_N(\psi) = V_N(\varphi)$ (resp. $V_N(\psi) \leq V_N(\varphi)$). If there are no arbitrage opportunities, we have $V_n(\psi) = V_n(\varphi)$ (resp. $V_n(\psi) \leq V_n(\varphi)$) for $n = 0, \dots, N$.*

In particular, for the underlying asset price processes S^i we have $S^i = V(\varphi)$ if φ denotes a self-financing strategy with $S^i_N = V_N(\varphi)$ for some i.

Proof. We proceed by an indirect reasoning: If the prices of the two portfolios do not coincide for all $n \leq N$, then there exist arbitrage opportunities.

To this end, suppose that $V_n(\varphi) < V_n(\psi)$ with positive probability for some $n < N$. The idea now is to go long in (i.e. buy) the relatively cheap portfolio and go short in (i.e. sell) the relatively expensive one at the same time. More specifically, we buy φ and we sell ψ at time n if $V_n(\varphi) < V_n(\psi)$ occurs. Otherwise we do nothing. The difference $V_n(\psi) - V_n(\varphi) > 0$ is invested in any security, e.g. S^0. At time N, we liquidate our portfolio. The revenues from the long position in φ and the obligations from the short position in ψ cancel each other. The positive wealth from the investment in S^0 constitute our riskless arbitrage gain.

The second statement follows from considering the portfolio ψ that contains one share of S^i and nothing else. $\square$

The following deep thereom links absence of arbitrage to the existence of *equivalent martingale measures*, which play a key role in financial mathematics. Intuitively, it states that there exist fictitious probabilities Q such that the market can be interpreted as a fair game, where discounted profits and losses cancel on average. Note that this may not hold under the physical or “real” probabilities, which are expressed by P. Under real probabilities one would rather expect risky assets to have a higher return on average as a compensation for the higher risk, cf. e.g. the capital asset pricing model (*CAPM*) in economic theory. Since taking risks is not rewarded relative to Q, the corresponding probabilities are often called *risk neutral*.

Theorem 4.10 (First fundamental theorem of asset pricing)**.** *There are no arbitrage opportunities if and only if there exists an* equivalent martingale measure *(EMM), i.e. a probability measure $Q \sim P$ such that the discounted price process $\hat{S}$ is a martingale relative to Q.*

Proof. We only show the simpler implication that the existence of an EMM implies that there are no arbitrage opportunities. Let φ denote a self-financing strategy with $V_0(\varphi) = 0$ and $V_N(\varphi) \geq 0$. Strategy φ serves as a potential arbitrage because it starts with zero initial capital and does not take any risk of losses. The Q-expectation of discounted terminal wealth equals

$$E_Q(\hat{V}_N(\varphi)) = E_Q(\hat{V}_0(\varphi) + \varphi \bullet \hat{S}_N) = E_Q(0 + \varphi \bullet \hat{S}_0) = 0.$$

We have used the fact that $\varphi \bullet \hat{S}$ is a martingale relative to Q, which implies that its expectation stays constant over time. A nonnegative random variable with expected value 0 equals 0 with probability 1. This means $\hat{V}_N(\varphi) = 0$ and hence also $V_N(\varphi) = 0$. Consequently, φ is not an arbitrage opportunity. □

Theorem 4.10 is particularly useful for showing that a given market model does not allow for arbitrage. It suffices to find an EMM, which is typically much simpler than to absence of arbitrage directly. Note that the EMM Q may not be unique, cf. e.g. Figure 4.7. Moreover, it depends on the choice of the numeraire security.

Concrete models

The statements so far do not depend on the distribution of price processes. In order to calculate prices and strategies explicitly in the following chapters, we need to consider concrete models. Its construction is a delicate task as it has to compromise between different goals. Models should be mathematically tractable without contradicting economic intuition. But ultimately they have to be compatible with real financial data, which must be examined by statistical means. Here we discuss a simple model for two traded securities, which gives the general framework for many easy models:

General random walk model

Cash does not appear in our general setup. Its place is taken by a *bond* (or *money market account*, *savings account*, *bank account*)

$$S_n^0 = S_0^0 \exp(rn) = S_0^0(1+\tilde{r})^n = S_0^0 \mathscr{E}(\tilde{r}I)_n, \tag{4.2}$$

where $r \in \mathbb{R}$, $\tilde{r} := e^r - 1$ and $I_n = n$ for $n \in \mathbb{N}$. It corresponds to an entirely riskless investment with fixed interest rate, which is not paid in cash. Instead it is reinvested in the bond and hence compounded. For simplicity we assume that the interest rate $\tilde{r}$ is a fixed constant.

By its simple structure the bond recommends itself as a natural numeriare. But in principle any tradable assets could be used for this purpose. Note that the existence of the bond does not contradict absence of arbitrage in the sense of Definition 4.8, even though it leads to some sort of riskless gains.

The bond can be held in negative quantities, which corresponds to a loan involving the same interest rate as a deposit. The term structure of interest rates in real markets is of course much more complicated. There exist long and short term investments with more or less fixed interest rates, which are paid at different times.

The only nontrivial security in our market is a *stock* or *foreign currency* whose price is assumed to be of the form

$$S_n^1 = S_0^1 \exp(X_n) = S_0^1 \prod_{m=1}^{n} (1 + \Delta\widetilde{X}_m) = S_0^1 \mathscr{E}(\widetilde{X})_n. \tag{4.3}$$

Here, $X_0 = 0$ and the increments $\Delta X_1, \Delta X_2, \dots$ of X are supposed to be independent and identically distributed. Moreover, we set

$$\widetilde{X}_n := \sum_{m=1}^{n} (e^{\Delta X_m} - 1). \tag{4.4}$$

In principle, the stock price process has the same structure as the bond. However, the return $\Delta\widetilde{X}_n$ for period n varies randomly. This may be due to unexpected news which affect prices favourably or unfavourably.

Note that $\log(S^1)$ is a sum of i.i.d. random variables, i.e., a *random walk*. S^1 is therefore often called a *geometric random walk*.

We assume that the filtration $(\mathscr{F}_n)_{n=0,\dots,N}$ is generated by S^1 and that $\mathscr{F} = \mathscr{F}_N$. Since any of the random vectors $(S_1^1, \dots, S_n^1)$, $(\hat{S}_1^1, \dots, \hat{S}_n^1)$, $(\Delta\hat{X}_1, \dots, \Delta\hat{X}_n)$, and $(\hat{X}_1, \dots, \hat{X}_n)$ can be expressed in terms of any of the other ones, they all carry the same information. Put differently, the filtration is also generated by $\hat{S}^1$, by $\Delta\hat{X}$, or $\hat{X}$, respectively.

Cox-Ross-Rubinstein binomial tree model

The easiest model of the type above often used in mathematical finance is the *Cox-Ross-Rubinstein* model, also called *binomial tree model*, which is presented below.

The market consists of a bond and a stock in the general classical market model. More specifically, let

$$S_n^0 = S_0^0(1 + \tilde{r})^n$$

with constants $S_0^0 > 0$, $\tilde{r} \geq 0$ and

$$S_n^1 = S_0^1 \prod_{m=1}^{n} (1 + \Delta\widetilde{X}_m)$$

with constant $S_0^1 > 0$ and independent, identically distributed random variables $\Delta\widetilde{X}_1, \dots, \Delta\widetilde{X}_N$. The major assumption is now that

$$P(\Delta\widetilde{X}_n = u-1) = p = 1 - P(\Delta\widetilde{X}_n = d-1)$$

with $0 < d < 1+\tilde{r} < u$ and $0 < p < 1$. In particular, the stock price may rise or fall in a single period only by a fixed factor u or d, respectively. The discounted price process is of the form

$$\hat{S}_n^1 = \frac{S_0^1}{S_0^0} \prod_{m=1}^n \frac{1+\Delta\widetilde{X}_m}{1+\tilde{r}} = \hat{S}_0^1 \prod_{m=1}^n (1+\Delta\hat{X}_m),$$

where $(\hat{X}_n)_{n=0,\dots,N}$ is defined by $\hat{X}_n := \sum_{m=1}^n (\frac{1+\Delta\widetilde{X}_m}{1+\tilde{r}} - 1)$ and satisfies

$$P(\Delta\hat{X}_n = \tfrac{u}{1+\tilde{r}} - 1) = p = 1 - P(\Delta\hat{X}_n = \tfrac{d}{1+\tilde{r}} - 1).$$

Classical Gaussian model

In the classical standard model the daily *logarithmic returns* ΔX_n are assumed be *Gaussian* or normally distributed, i.e. the model is determined entirely by two parameters μ, σ^2. These parameters can be estimated based on stock price data from the past. Since the logarithmic returns

$$\Delta X_n = \log\left(\frac{S_n^1}{S_{n-1}^1}\right) \sim \mathtt{N}(\mu, \sigma^2)$$

in this model are independent and identically distributed with mean μ and variance σ^2, one may use the standard estimates

$$\begin{aligned} \hat{\mu} &:= \frac{1}{N}\sum_{n=1}^N \Delta X_n, \\ \hat{\sigma}^2 &:= \frac{1}{N-1}\sum_{n=1}^N (\Delta X_n - \hat{\mu})^2 \end{aligned}$$

if the observations S_n^1, $n = 0, \dots, N$ are available.

In view its very simple structure, the standard model represents real data quite well. However, it does not stand up to a more careful examination. Repeatedly observed *stylized facts* have lead to a variety of alternative models. Let us illustrate the issue by having a look at real data. Figure 4.8 depicts the evolution of the German stock index (DAX) from April 1991 to April 2001. This should be compared to the simulation of stock index data in Figure 4.9, which is based on Equation (4.3) with Gaussian logarithmic returns. At first glance these diagrams seem to resemble each other quite closely. However, the representation of the daily logarithmic returns $\Delta X_n = \log(S_n^1/S_{n-1}^1)$ in Figure 4.10 is more revealing. According to the standard model, they should form a sample of independent, identically distributed (i.i.d.) random variables. But they differ in two ways from the corresponding simulation of i.i.d. data in Figure 4.11. Firstly, large daily price changes happen much more often than in standard model (4.3) with comparable variance (*heavy tails*, *leptokurtosis*). Repeatedly, price changes of more than 6% happened in the observation period. According to the Gaussian model, this should happen only once in 1300 years. Of course,

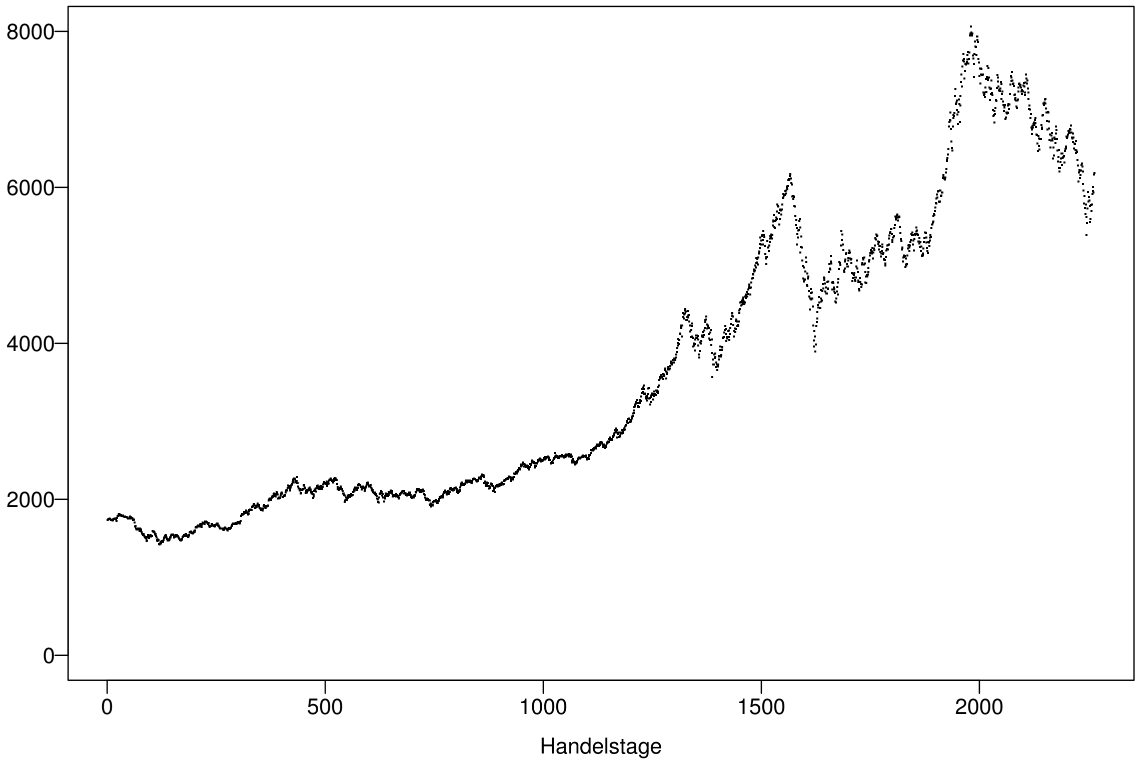

Figure 4.8: DAX from April 1992 to April 2001

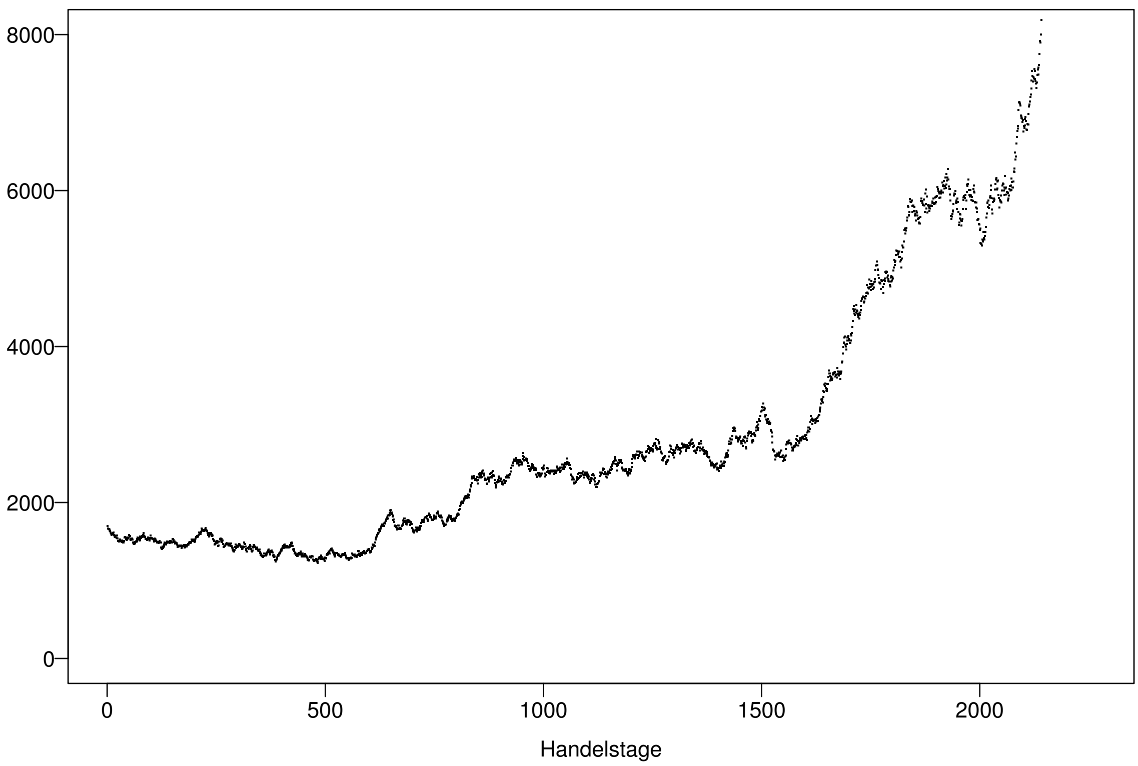

Figure 4.9: Simulation according to Equation (4.3)

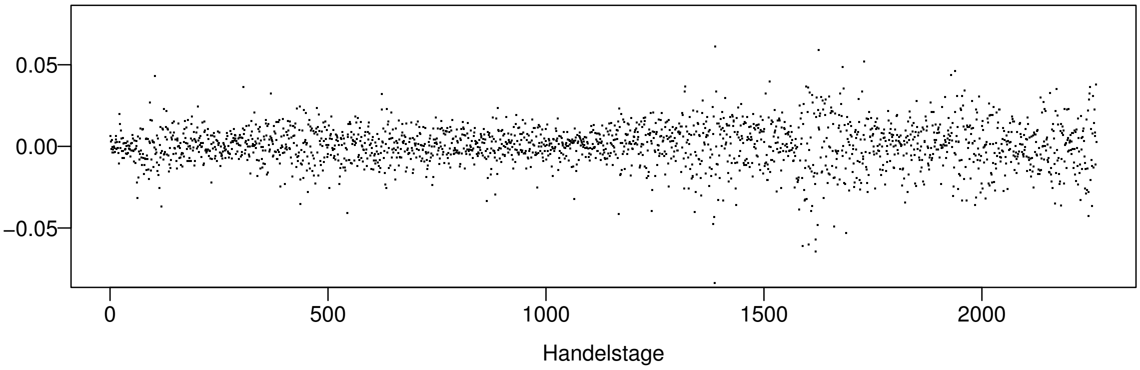

Figure 4.10: Daily logarithmic DAX returns

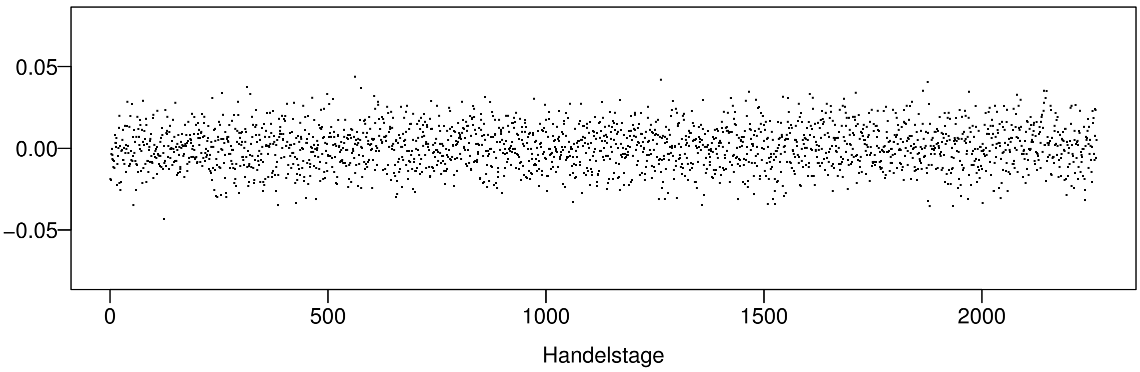

Figure 4.11: Simulation of daily normally distributed returns

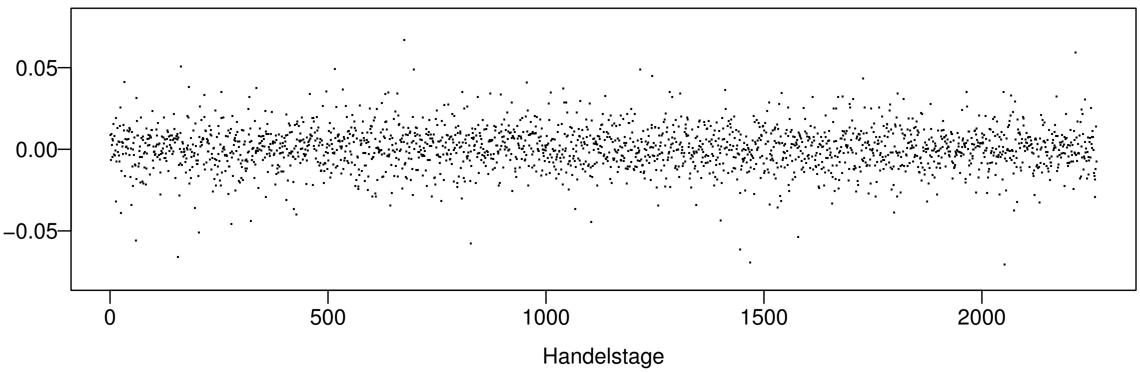

Figure 4.12: Simulation of daily normal-inverse Gaussian returns

such sudden major price changes involve a considerable risk for investors. Therefore, one should carefully check to what extent theoretical results based on an unsatisfactory model are really applicable in practice.

The above deficiency is not so much caused by model (4.3) itself, but rather by choosing the Gaussian law. The latter is typically motivated by the central limit theorem, according to which sums of small and approximately independent factors are approximately normally distributed. However, daily returns could very well be dominated by a few large observations, which means that the Gaussian law lacks foundation. As an example, Figure 4.12 depicts a simulation of normal-inverse Gaussian data. Compared to the normal law with the same variance, this law puts more weight on the tails and the centre.

But a comparison of Figures 4.10, 4.11, 4.12 reveals a further difference, which is indeed linked to model (4.3). Relatively large resp. small returns are clustered in the real data set (*volatility clustering*). Processes with independent and identically distributed increments do no show this behaviour. Therefore, they are not compatible with the observed data, even if we drop the assumption of Gaussian returns.

A way out is to model the increments ΔX_n in (4.3) in the form

$$\Delta X_n = \sigma_n \Delta Z_n,$$

where independent, identically distributed random variables $\Delta Z_1, \Delta Z_2, \dots$ as above are responsible for daily price changes. These are multiplied by a slowly varying process σ, who makes calm periods with small σ_n take turns with busy periods caused by relatively large σ_n. The concrete specification and estimation of a parametric model for $\Delta Z, \sigma$ belongs to the field of statistics and econometrics. A sensible model should be compatible with real data, easy to estimate, and tractable for purposes of mathematical finance. According to a general rule, simple models should be preferred to complex ones unless they do not match the data well enough. It depends on the concrete purpose of the model whether this is the case.

Note that Gaussian returns contradict our assumption of a finite probability space. In order to avoid mathematical inconsistencies one could approximate continuous laws by appropriate discrete ones. Alternatively, one could extend the mathematical theory to infinite sample spaces. Indeed, most results of our setup hold with only minor modifications in general discrete-time models.

Appendix QF

4.A Dividend payments

So far we assumed that no dividends or interest is paid on traded assets. In practice this assumption is obviously violated in many cases. There are two ways to account for dividend payments. In the *indirect approach* one identifies them with price gains. More precisely, one defines a fictitious security without dividend payments which yields the same profits and losses as the original one if dividends are reinvested by buying additional shares of the same asset. As an example consider a coupon-paying bond whose value of 1 Euro stays constant over time. If it pays 3% interest in any period (i.e. 0.03 Euro per share), reinvesting these payments yields a wealth of 1.03^n at time n. Generally, the dividend-free price process $S_n^0 = (1+r)^n$ can be used to model a constant bond paying a fixed interest rate r per period. In case of more complex dividend payments the transition from real to fictitious securities becomes somewhat messy. But from a theoretical point of view it provides a way to generalize results as e.g. the fundamental theorems of asset pricing to dividend-paying securities.

In this section we follow the *direct approach*, which means that we explicitly keep track of dividend payments in the bookkeeping. As before we denote the $d+1$-dimensional securities' price process by $S = (S^0, \dots, S^d)$. Moreover, we consider a $d+1$-dimensional adapted process $D = (D^0, \dots, D^d)$ with $D_0 = (0, \dots, 0)$, called *cumulative dividend process.* D_n^i stands for the dividends that are paid for security S^i up to time n, i.e. at time n the dividend ΔD_n^i is paid for secuity i.

Even more than in the dividend-free case we must pay attention to the chronological order of events. Three things happen in the period from $n-1$ to n. The portfolio changes from φ_{n-1} to φ_n, the prices move from S_{n-1} to S_n and dividends ΔD_n^i are paid. We suppose that these events happen precisely in this order.

The dividend payments affect the portfolio because the bank account to which the dividend is transferred to appears in our model as a traded security as well, typically as the numeraire. We interpret φ_n as the portfolio before dividends are paid, $V_n(\varphi)$, on the other hand, as the value of the portfolio after dividends at time n have been paid. This motivates the following definitions.

Definition 4.A.1. A *trading strategy* (or a *portfolio*) is a $d+1$-dimensional predictable process $\varphi = (\varphi^0, \dots, \varphi^d)$. Its *value process* or *wealth process* is

$$V(\varphi) := \varphi^\top (S + \Delta D).$$

Strategy φ is called *self financing*, if

$$\varphi_n^\top S_{n-1} = \varphi_{n-1}^\top (S_{n-1} + \Delta D_{n-1})$$

or, equivalently,

$$(\Delta\varphi_n)^\top S_{n-1} = \varphi_{n-1}^\top \Delta D_{n-1}$$

for $n = 1, 2, \ldots$

The analogue of Lemma 4.3 reads as follows.

Lemma 4.A.2. *A trading strategy φ is self financing if and only if*

$$V(\varphi) = V_0(\varphi) + \varphi \bullet (S + D).$$

Proof. This is left as an exercise. □

As before the bookkeeping simplifies by working in discounted terms. We suppose that the numeraire asset is positive and we define the discounted price process $\hat{S} := S/S^0$ as before.

Definition 4.A.3.

$$\hat{D} := \frac{1}{S^0} \bullet D \tag{4.5}$$

is called *discounted dividend process*. The *discounted value process* or *discounted wealth process* of strategy φ is defined as

$$\hat{V}(\varphi) := \frac{1}{S^0} V(\varphi) = \varphi^\top (\hat{S} + \Delta\hat{D}).$$

Note that Equation (4.5) involves a stochastic integral instead of a product. Dividends $\Delta\hat{D}_n$ are paid at time n, which is why they have to be discounted by S_n^0. Formula

$$\Delta\hat{D}_n = \frac{\Delta D_n}{S_n^0}$$

for the *present* payments leads to

$$\hat{D}_n = \sum_{m=1}^{n} \Delta\hat{D}_m = \sum_{m=1}^{n} \frac{1}{S_m^0} \Delta D_m = \frac{1}{S^0} \bullet D_n$$

as in the previous definition. Lemma 4.A.2 can be expressed in discounted terms as well:

Lemma 4.A.4. *A trading strategy φ is self financing if and only if*

$$\hat{V}(\varphi) = \hat{V}_0(\varphi) + \varphi \bullet (\hat{S} + \hat{D}).$$

Proof. φ is self financing if and only if $(\Delta\varphi_n)^\top \hat{S}_{n-1} = \varphi_{n-1}^\top \Delta\hat{D}_{n-1}$ holds for any n. Hence, the assertion follows from Lemma 4.A.2, applied to $\hat{S}$ instead of S. □

Lemma 4.7 remains literally true in the case of dividend payments.

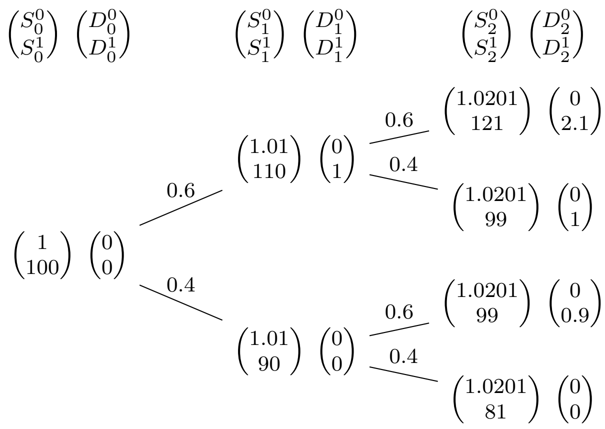

Figure 4.13: A market with two assets and dividend payments on S^1

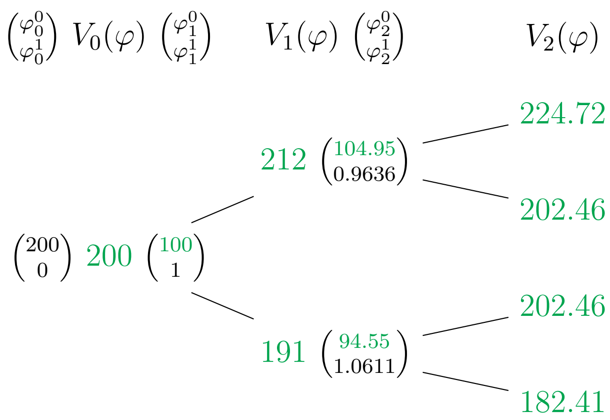

Figure 4.14: A self-financing strategy φ and its value process in the market with dividends of Figure 4.13

Lemma 4.A.5. *For any predictable process $(\varphi^1, \dots, \varphi^d)$ and any $V_0 \in \mathbb{R}$ there exists a unique predictable process φ^0 such that $\varphi = (\varphi^0, \dots, \varphi^d)$ is self financing with $V_0(\varphi) = V_0$.*

Proof. The proof is left as an exercise. □

In order to link arbitrage and equivalent martingale measures we consider as before a finite time horizon $N \in \mathbb{N}$. Moreover, we assume that no dividends are paid for Security S^0 (i.e. $D^0 = 0$). *Arbitrage* is defined as in Section 4. Note that the value process of any self-financing strategy is of the same form as in the case without dividends, but with $\hat{S}+\hat{D}$ instead of $\hat{S}$. This implies that the first fundamental theorem holds with this modification.

Corollary 4.A.6 (First fundamental theorem of asset pricing)**.** *The market does not allow for arbitrage if and only if there exists an equivalent martingale measure for $\hat{S}+\hat{D}$, i.e. a probability measure $Q \sim P$ such that $\hat{S}+\hat{D}$ is a Q-martingale.*

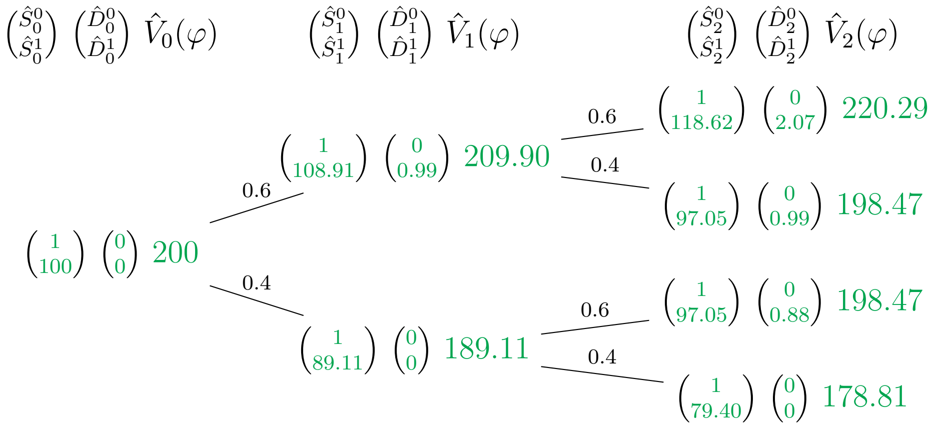

Figure 4.15: Discounted price process $\hat{S}$, dividend process $\hat{D}$, and value process $\hat{V}(\varphi)$ corresponding to Figures 4.13 and 4.14

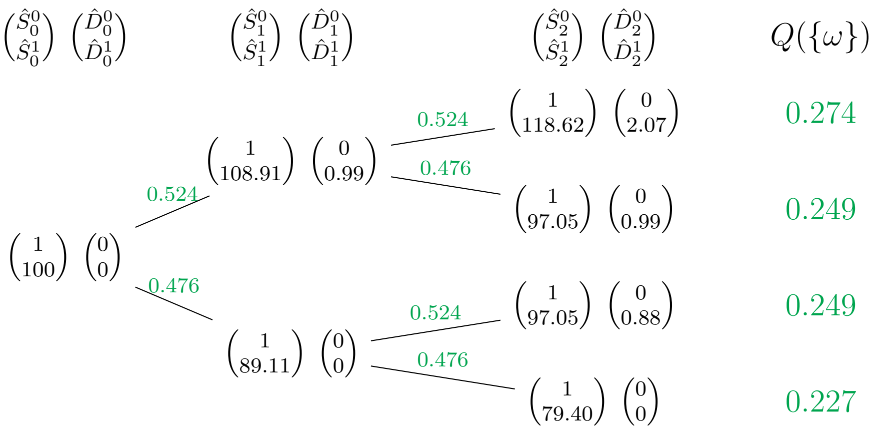

Figure 4.16: Equivalent martingale measure probabilities for the market in Figures 4.13 and 4.15

Appendix MF

4.B On the proof of the first FTAP

In order to prove the fundamental theorem, we need the following separation theorem.

Theorem 4.B.1 (Separation theorem for cones)**.** *Let $U \subset \mathbb{R}^n$ be a nonempty, closed convex cone, $K \subset \mathbb{R}^n$ nonempty, compact, and convex, $U \cap K = \varnothing$. Then there exist a $\lambda \in \mathbb{R}^n$ with $\lambda^\top x \leq 0$ for all $x \in U$ and $\lambda^\top x > 0$ for all $x \in K$. If U is a subspace, we have $\lambda^\top x = 0$ for all $x \in U$.*

We base the proof on the following lemma:

Lemma 4.B.2 (Basic Separation Theorem)**.** *Let $C \subseteq \mathbb{R}^n$ be nonempty, closed, convex and $0 \in C^c$. Then there exists $\lambda \in \mathbb{R}^n$, $\gamma > 0$ such that*

$$\lambda^T x \geq \gamma \text{ for all } x \in C.$$

Proof. We obviously find $r > 0$ such that $C \cap B_r \neq \varnothing$, where B_r denotes the closed ball around 0 with radius r. $C \cap B_r$ is compact, hence the function

$$C \cap B_r \to \mathbb{R},\ x \mapsto \|x\|$$

attains its minimum in some point $x^* \in C$. As $\|x^*\| \leq r \leq \|x\|$ for all $x \notin B_r$, x^* is actually a minimizer of

$$C \to \mathbb{R},\ x \mapsto \|x\|.$$

As $0 \notin C$, $\|x^*\| > 0$. Let $x \in C$. As C is convex,

$$\alpha x + (1-\alpha)x^* \in C \text{ for all } \alpha \in (0,1).$$

By definition of x^*

$$\|\alpha x + (1-\alpha)x^*\| \geq \|x^*\|,$$

that is

$$(\alpha x + (1-\alpha)x^*)^T(\alpha x + (1-\alpha)x^*) \geq (x^*)^T x^*,$$

yielding

$$\alpha(x - x^*)^T(x - x^*) + 2(x - x^*)^T x^* \geq 0.$$

As this inequality holds for all $\alpha \in (0,1)$, we obtain ($\alpha \searrow 0$)

$$(x - x^*)^T x^* \geq 0,$$

i.e.

$$(x^*)^T x^* \leq x^T x^*.$$

Furthermore, as $x^* \neq 0$, it holds that $0 < (x^*)^T x^*$. Now, $\gamma = (x^*)^T x^*$ and $\lambda := x^*$ yields the claim. □

Proof of 4.B.1. Write

$$C := K - U = \{b - x : b \in K, x \in U\}.$$

Then, C fulfills the assumptions in the Basic Separation Theorem (note that the sum of a compact and convex set and a closed and convex set is closed and convex) and $0 \notin C$. Hence, there exists $\lambda \in \mathbb{R}^n$ such that for every $c = b - x \in C$ it holds that $\lambda^T c > 0$, i.e. for any $x \in U, b \in K$

$$\lambda^T (b - x) > 0, \text{ i.e. } \lambda^T b > \lambda^T x.$$

As U is a cone, $0 \in U$, hence for any $b \in K$

$$\lambda^T b > 0.$$

On the other hand, as U is a cone and if $x \in U$, then for any $\gamma > 0$ and arbitrary $b \in K$

$$\lambda^T (\gamma x) < \lambda^T b,$$

i.e.

$$\lambda^T x < \frac{1}{\gamma} \lambda^T b$$

and as $\gamma \to \infty$

$$\lambda^T x \leq 0,$$

yielding the first claim. If U is even a subspace, we can choose $\gamma < 0$ as well in the previous argument and additionally obtain

$$\lambda^T x \geq 0.$$

□

The previous separation theorems are formulated in $\mathbb{R}^n$. For a finite Ω, we can identify random variables $X : \Omega \to \mathbb{R}$ with vectors in $\mathbb{R}^\Omega$. W.l.o.g we may assume that the underlying σ-algebra contains all atoms $\{\omega\}$, $\omega \in \Omega$ and $P(\{\omega\}) > 0$, which we do for notational convenience in the following proof.

Proof of 4.10. $\Leftarrow$: Let φ be a self financing strategy with $V_0(\varphi) = 0$ and $V_N(\varphi) \geq 0$. Then we have $\varphi \bullet \hat{S}_N = \hat{V}_N(\varphi) \geq 0$. Since $\varphi \bullet \hat{S}$ is a Q-martingale according to Lemma 3.16(5), we have $E_Q(\varphi \bullet \hat{S}_N) = E_Q(\varphi \bullet \hat{S}_0) = 0$, meaning $\varphi \bullet \hat{S}_N = 0$ is Q-almost surely and therefore $V_N(\varphi) = 0$ P-almost surely.

$\Rightarrow$: Let $U := \{\varphi \bullet \hat{S}_N \in \mathbb{R}^\Omega : \varphi \text{ self financing strategy with } \varphi_0 = 0\}$ and

$$K := \Big\{X \in \mathbb{R}^\Omega : X \geq 0; \textstyle\sum_{\omega \in \Omega} X(\omega) = 1\Big\}.$$

Then U is a subspace of $\mathbb{R}^\Omega$ and K is a compact, convex set. Because of absence of arbitrage it follows that $U \cap K = \varnothing$. According to Theorem 4.B.1 there exists a $\lambda \in \mathbb{R}^\Omega$

with $\sum_{\omega\in\Omega}\lambda(\omega)x(\omega) = 0$ for all $x \in U$ and $\sum_{\omega\in\Omega}\lambda(\omega)x(\omega) > 0$ for all $x \in K$. We define a probability measure Q by

$$Q(\{\omega\}) = \frac{\lambda(\omega)}{\sum_{\widetilde{\omega}\in\Omega}\lambda(\widetilde{\omega})}.$$

Because $1_{\{\omega\}} \in K$ we have $\lambda(\omega) > 0$ for all $\omega \in \Omega$. Therefore Q is a probability measure equivalent to P.

Let $i \in \{1,\ldots,d\}$, $m \le n$, $C \in \mathscr{F}_m$. Let φ be a self financing strategy with $\varphi_0 = 0$ and

$$\varphi^j_\ell(\omega) := \begin{cases} 1_C(\omega)1_{\{m+1,\ldots,n\}}(\ell) & \text{for } i = j, \\ 0 & \text{else.} \end{cases}$$

Then we have

$$E_Q(1_C(\hat{S}^i_n - \hat{S}^i_m)) = \left(\sum_{\widetilde{\omega}\in\Omega}\lambda(\widetilde{\omega})\right)^{-1}\sum_{\omega\in\Omega}\lambda(\omega)(\varphi \bullet \hat{S}_N)(\omega) = 0,$$

thus $E_Q(\hat{S}^i_n - \hat{S}^i_m|\mathscr{F}_m) = 0$. Consequently, $\hat{S}^i$ is a Q-martingale. □

For later use we need the following version of the fundamental theorem.

Lemma 4.B.3. *In a no-arbitrage market let X_0 be a random variable with*

$$X_0 \notin \{\varphi \bullet \hat{S}_N - X : \varphi \text{ self financing strategy and } X \text{ non-negative random variable}\}.$$

Then we can choose the equivalent martingale measure from Theorem 4.10 so that $E_Q(X_0) > 0$.

Proof. Replace K in the previous proof by the convex closure $\tilde{K}$ of the sets K and X_0, i.e.

$$\tilde{K} = \{cX_0 + (1-c)X : c \in [0,1], X \in K\}.$$

$\tilde{K}$ is compact as a convex closure of $\{X_0\}$ and finitely many $1_{\{\omega\}}$, $\omega \in \Omega$. Suppose $\tilde{K} \cap U \neq \varnothing$. Then it follows that because of $K \cap U = \varnothing$ there exists a self financing strategy φ, a $X \in K$ and $c \in (0,1]$ with $cX_0 + (1-c)X = \varphi \bullet \hat{S}_N$, and therefore $X_0 = \frac{\varphi}{c} \bullet \hat{S}_N - \frac{1-c}{c}X$ which contradicts our assumption.

For λ as in the previous proof we then have $\sum_{\omega\in\Omega}\lambda(\omega)x(\omega) > 0$ for all $x \in \tilde{K}$, especially for $x = X_0$, which gives us $E_Q(X_0) > 0$. □

Chapter 5

Derivative pricing and hedging

In this chapter we study derivative pricing and hedging based on the Principle of No Arbitrage.

Derivatives are securities whose price depends in a more or less complex way on the price of other, so-called *underlying* securities or assets. Often these are forward deals. This means that it is negotiated now that a particular transaction will be made in the future at a certain price. These contracts may be signed in order to reduce the own exposure to price fluctuations. However, for the counterparty they involve a potentially larger exposure to risks. Consequently, a key issue in mathematical finance is how the risk caused e.g. by selling derivatives can be reduced or *hedged* by trading the underlying securities skillfully. This problem is related to the likewise important question how to assign an appropriate price to a derivative contract.

More specifically, we distinguish two situations. First, the derivative is itself liquidly traded at the exchange, which means that the results from the previous chapter can be applied. A pivotal role will be played by the no arbitrage assumption. This seemingly weak and innocent assumption has in some cases far-reaching consequences. In others, however, it does not help much.

Alternatively, a forward deal may be contracted *over the counter* (OTC) by two parties. In this case, which is discussed starting on page 79, we must take an individual rather than a market perspective. In the end, the two situations lead to similar formulas and results. This may explain why the literature does not often distinguish very clearly between OTC and liquidly traded derivatives.

Our general market model rests as before on a finite filtered probability space $(\Omega, \mathfrak{P}(\Omega), (\mathscr{F}_n)_{n=0,\dots,N}, P)$ with finite time horizon $N \in \mathbb{N}$. As before we assume that all outcomes happen with positive probability and that the initial σ-field is trivial ($\mathscr{F}_0 = \{\Omega, \varnothing\}$), which means that we consider the initial state of the model as deterministic or fixed. We also set $\mathscr{F}_N = \mathfrak{P}(\Omega)$. As in the previous chapter we suppose that the $d+1$-dimensional securities price process $S = (S^0, \dots, S^d)$ with positive numeraire S^0 is given. This market of so-called *underlying* securities is supposed not to allow for arbitrage in the sense of Definition 4.8.

5.1 Contingent claims

Many — but not all — forward deals can be represented as a random variable X which stands for the random payoff at time N. We consider some examples.

1. A *European call option* gives the owner the right to buy a fixed quantity of a fixed asset at a fixed time *(maturity)* and for a fixed price *(strike)*. There is no obligation to buy the asset.

   Such an option can be represented as a random variable X. If the market price of the asset at maturity is below the strike, it does not make sense to exercise the option, which is hence worthless. If, on the other hand, the market price settles above the strike, the value of the option is the difference of the two. Indeed, even if the option holder is not interested in the asset, he can realize the profit by selling it immediately at the market price. Hence, the value of a call option written on one share of security S^1 with maturity N and strike K amounts to

   $$X = (S_N^1 - K)^+ = \max(S_N^1 - K, 0) \tag{5.1}$$

   at time N.

   Often, the asset is not actually physically delivered. Instead a pure cash settlement takes place. In practice this may in fact make a difference because of transaction and storage cost involved in physical delivery. This difference will be neglected in our analysis.

   In case of a pure cash settlement, the price of the asset in (5.1) could in fact be replaced by any well-defined quantity, even if it is not a proper price. There are e.g. options on indices.

2. The *European put option* corresponds to the call but the owner has the right to sell rather than buy the asset. Otherwise, the above comments hold for the put as well. The value of a put option written on one share of security S^1 with maturity N and strike K amounts to

   $$X = (K - S_N^1)^+ = \max(K - S_N^1, 0) \tag{5.2}$$

   at time N.

3. The owner of an *American call* resp. *put* can exercise her option any time before maturity. She does not necessarily have to wait until maturity. In constrast to the European option, this involves a true choice. This complicates the mathematical treatment. Indeed, the seller faces the additional uncertainty at which time the option to buy resp. sell the underlying will be exercised. American options cannot generally be expressed in terms of a single random variable.

4. In a *forward contract* one agrees to buy a fixed quantity of a fixed asset at a fixed date (*maturity*) and a fixed price (*forward price*). In contrast to a call option this involves an obligation. The forward price is chosen such that no money has to be paid when the contract is settled, i.e. the contract itself is worthless. Observe the different meaning of the word *price* compared to the price of call and put options.

Whereas the forward price resembles the strike and hence the price of the underlying at maturity, the *option price* typically refers to the price of the option contract as an asset in its own right.

In order to express its payoff as a random variable, we consider a forward contract on asset S^1 with maturity N. We assume that it is entered at time $n \in \{0, \ldots, N\}$ at a forward price O_n. The value of this contract at maturity equals

$$X = S_N^1 - O_n$$

because the holder needs to pay "only" O_n instead of the market price S_N^1 in order to buy one share of S^1. Note that the forward price may well lie above the market price, in which case X is negative.

5. Forwards are usually contracted directly by two parties (*over the counter*). The exchange-traded version is called *futures contract*. It differs from a forward by the bookkeeping, which makes it less transparent. Whereas money is paid only at maturity for a forward contract, daily payments are made for a futures contract. Depending on the market price of the underlying, its price change is credited resp. debited to the counterparties' *margin accounts (marking to market)*.

   A future can be viewed as a contract which can be entered and terminated free of charge at any time. This contract is based on a time varying quotation, the so-called *futures price* U_n, which resembles the above forward price. At maturity N, the futures price is laid down as the market price of a corresponding underlying asset (e.g. $U_N = S_N^1$). During the holding period of the future, payments of the daily futures price change $U_n - U_{n-1}$ are made to the holder's margin account. Neglecting interest payments, the credit and debit notes from time n to N sum up to $U_N - U_n = S_N^1 - U_n$. This corresponds to the payoff of a forward at maturity if the futures price U_n is replaced by the forward price O_n. In general, however, futures cannot be expressed in terms of a random variable expressing a payoff at maturity because interest is earned or charged for the intermediate payments.

5.2 Liquidly traded derivatives

Even if not all traded derivatives allow for such a representation, we call $\mathscr{F}_N$-measurable random variables X interchangeably (contingent) *claim*, *derivative*, or *option*. The random variable X represents the value of the contract at time N and

$$\hat{X} := \frac{X}{S_N^0}$$

the *discounted payoff*, respectively. We consider any contingent claim in this section as a liquidly traded security S^{d+1}, whose market price at any time $n = 0, \ldots, N$ equals S_n^{d+1}. Based on modest assumptions, we would like to draw conclusions on possible or reasonable market prices S^{d+1} for $n \leq N$ besides $S_N^{d+1} = X$, in particular for $n = 0$. Common sense probably suggests two answers. Firstly, the claim's market price is subject to supply and demand and hence unpredictable. Secondly, the conditional expectation $E(X|\mathscr{F}_n)$ may appear as an at least not unreasonable suggestion.

The following corollary implies that absence of arbitrage restricts the set of possible option price processes by a martingale condition. The only feasible derivative prices can be represented as conditional expectations under equivalent martingale measures.

Corollary 5.1. *Let X denote a contingent claim. An adapted derivative price process S^{d+1} satisfying $S^{d+1}_N = X$ leads to an arbitrage-free market $(S^0, \dots, S^{d+1})$ if and only if there exists an equivalent martingale measure Q for the market $(S^0, \dots, S^d)$ such that*

$$\hat{S}^{d+1}_n = E_Q(\hat{X}|\mathscr{F}_n)$$

for $n = 0, \dots, N$.

Proof. $\Leftarrow$: This implication follows from the fundamental theorem 4.10 because Q is an equivalent martingale measure for $(S^0, \dots, S^{d+1})$.

$\Rightarrow$: By the first fundamental theorem 4.10 there exists an EMM Q for $(S^0, \dots, S^{d+1})$. □

In some cases absence of arbitrage determines the derivative price process uniquely. This is the case for replicable claims in the sense of the following definition.

Definition 5.2. A contingent claim X is called *replicable* or *attainable* if there exists a self-financing strategy φ (trading only in the primary assets $S^0, \dots, S^d$) which satisfies $X = V_N(\varphi)$. In this case φ is called (perfect) *hedging strategy* for X.

Remark. Observe that a contingent claim is attainable if and only if there is some $x \in \mathbb{R}$ and some $\mathbb{R}^d$-valued predictable process $\varphi = (\varphi^1, \dots, \varphi^d)$ with

$$\hat{X} = x + \varphi \bullet \hat{S}_N.$$

Here, $\hat{S}$ is to be interpreted as the d-dimensional process $(\hat{S}^1, \dots, \hat{S}^d)$, cf. the notation following Lemma 4.7. The question whether such an integral representation exists motivates Theorem 5.B.2.

For attainable claims it does not make a difference whether one owns the claim or its replicating portfolio φ — the value at maturity is the same. Hence, such a derivative is redundant in the sense that it is already available in the market, namely in the shape of the dynamic strategy φ. This suggests that the market price of a replicable claim should coincide with the value of its replicating portfolio.

Theorem 5.3. *For any contingent claim X the following statements are equivalent.*

1. *X is attainable by a self-financing strategy φ.*
2. *There is one and only one derivative price process S^{d+1} with $X = S^{d+1}_N$ and such that the market $(S^0, \dots, S^{d+1})$ does not allow for arbitrage.*
3. *For any EMM Q the definition $\hat{S}^{d+1}_n := E_Q(\hat{X}|\mathscr{F}_n)$ leads to the same process S^{d+1}.*

In this case we have $S^{d+1} = V(\varphi)$.

In the situation of the previous theorem, S^{d+1} is the only price process that is fair in the sense that it does not lead to riskless gains in the market.

Definition 5.4. If a contingent claim X is attainable, we call the process S^{d+1} in Theorem 5.3 its (unique) *fair price process.*

Recall that we asked for reasonable market prices of derivatives. For attainable claims we have found a satisfactory answer. In lucky but somewhat rare cases, any contingent claim is in fact replicable.

Definition 5.5. The market is called *complete* if any contingent claim X is attainable.

Completeness can be characterized in terms of equivalent martingale measures, similarly as absence of arbitrage in Theorem 4.10.

Theorem 5.6 (Second fundamental theorem of asset pricing)**.** *If the market does not allow for arbitrage, we have equivalence between:*

1. *The market is complete.*
2. *There is a unique equivalent martingale measure.*

Proof. 1⇒2: For any fixed $A \in \mathscr{F}_N$ define the discounted contingent claim $\hat{X} := 1_A$. By Theorem 5.3 (1⇒3) $E_Q(\hat{X}|\mathscr{F}_0) = Q(A)$ coincides for all EMM's Q. Hence there is only one EMM.

2⇒1: This follows from Theorem 5.3 (3⇒1). □

Let us come back to the intuitive discussion in the beginning of Section 5.2. We observe that at least in complete markets both intuitive answers are more or less partly wrong. Firstly, absence of arbitrage leaves only one possible option price. Supply and demand cannot have an effect on the option price — at least not without affecting the underlying as well. Secondly, $S_n^{d+1} = E(X|\mathscr{F}_n)$ or its discounted variant $\hat{S}_n^{d+1} = E(\hat{X}|\mathscr{F}_n)$ may not yield acceptable prices unless P happens to be an EMM. As reasonable as they may seem, they both typically lead to arbitrage.

In complete markets it suffices to specify the law of the underlying securities; it already determines the dynamics of any derivative uniquely. We study such a complete market model (the Cox-Ross-Rubinstein market) below. In general incomplete markets, absence of arbitrage still imposes constraints on possible derivative price processes (cf. Corollary 5.1). Sometimes, however, they do not tell us much about option prices as we will see in Example 5.5.

The following results show that the number of possible outcomes is rather limited in complete market models.

Theorem 5.7. *Suppose that the market is complete and does not allow for arbitrage. Then the number of children of each node in the tree representing the market does not exceed the number of traded assets, i.e. $d+1$.*

Corollary 5.8. *Suppose that the market is complete and does not allow for arbitrage. Then the sample space* Ω *has at most* $(d+1)^N$ *elements.*

Proof. From the previous theorem it follows that the partition generating $\mathscr{F}_n$ has at most $(1+d)^n$ elements. Since we assumed $\mathscr{P}(\Omega) = \mathscr{F} = \mathscr{F}_N$, $\mathscr{F}_N$ is generated by the partition $\{\{\omega\} : \omega \in \Omega\}$, which has as many elements as Ω. $\square$

5.3 Individual perspective

In this section we consider derivatives that are contracted between two counterparties and may not be exchange traded. In particular, it is not obvious whether and at what price the contract can be sold between inception and maturity. Hence we cannot base our theory on a derivatice price *process* because this usually refers to a market price at which the claim can be bought and sold at any time.

We work with the same mathematical setup as in the previous section, i.e. we assume price processes $S^0, \dots, S^d$ to be given, along with some random variable X which represents the random payoff of a contingent claim at time N. As before, $\hat{X}$ denotes the corresponding discounted payoff.

We consider an asymmetric situation where the potential buyer is interested in the claim for unknown reasons. The seller — e.g. a bank — acts as a pure service provider who is not interested in the derivative on its own account. The bank faces two basic questions which are considered in the sequel.

1. What price does the bank need to charge from the potential buyer?
2. How can the bank hedge against the risk of losses that are involved by the unknown random payment which is due at maturity?

Generally, the answer to these questions depends on many factors, in particular on the bank's attitude towards risk. The minimum acceptable price for the bank cannot be determined without additional information of this kind. Therefore, we confine ourselves to reasonable rough bounds which are based on general considerations.

Definition 5.1. Let X denote a contingent claim. We call

$$\begin{aligned}\pi_U(X) \quad := \quad & \inf\Big\{x \in \mathbb{R} : \text{There is a self-financing strategy } \varphi \\ & \text{satisfying } V_0(\varphi) = x \text{ and } V_N(\varphi) \geq X\Big\}\end{aligned}$$

upper price and

$$\begin{aligned}\pi_L(X) \quad := \quad & \sup\Big\{x \in \mathbb{R} : \text{There is a self-financing strategy } \varphi \\ & \text{satisfying } V_0(\varphi) = x \text{ and } V_N(\varphi) \leq X\Big\}\end{aligned}$$

lower price of the option.[1]

[1] Here, inf and sup denote the infimum and supremum of the corresponding sets of real numbers. If you are not familiar with these concepts, you can – in this case – use min and max, resp., instead (see the next theorem).

In what sense do these represent rational bounds for the price that a bank can or will actually charge for the contingent claim? If it obtains a premium $x > \pi_U(X)$ for the option, this allows the bank to buy a self-financing portfolio whose terminal value $V_N(\varphi)$ suffices to meet its obligations at maturity. Consequently, the bank does not face any risk of losses. Except for administrative costs, there is no reason not to accept any potential buyer's offer to pay $x > \pi_U(X)$ for the option.

On the other hand, the bank should not accept any premium below the lower price. Indeed, for any initial wealth $x < \pi_L(X)$ it can sell a self-financing portfolio with terminal value $V_N(\varphi) \leq X$. The involved obligations at time N do not exceed those of a shorted contingent claim. Consequently, the bank is better off selling such a portfolio at the exchange than to sell an option to a potential buyer for less than $\pi_L(X)$. Altogether it follows that any reasonably charged or offered price for the option should lie between $\pi_L(X)$ and $\pi_U(X)$.

In the following theorem, these bounds are characterized in terms of equivalent martingale measures.

Theorem 5.2. *Let X denote a contingent claim.*

1. *For the upper price we have*

$$\begin{aligned}\pi_U(X) &= \min\Big\{x \in \mathbb{R} : \textit{There exists some self-financing strategy } \varphi \\ &\qquad \textit{with } V_0(\varphi) = x \textit{ and } V_N(\varphi) \geq X\Big\} \\ &= \sup\left\{S_0^0 E_Q(\hat{X}) : Q \textit{ EMM}\right\}.\end{aligned}$$

2. *Accordingly, the lower price satisfies*

$$\begin{aligned}\pi_L(X) &:= \max\Big\{x \in \mathbb{R} : \textit{There exists some self-financing strategy } \varphi \\ &\qquad \textit{with } V_0(\varphi) = x \textit{ and } V_N(\varphi) \leq X\Big\} \\ &= \inf\left\{S_0^0 E_Q(\hat{X}) : Q \textit{ EMM}\right\}.\end{aligned}$$

3. *If X is not replicable, then*

$$\left\{S_0^0 E_Q(\hat{X}) : Q \textit{ EMM}\right\} \tag{5.3}$$

is an open interval (and a singleton otherwise).

Above we argued that the premium charged by the bank should belong to the interval

$$[\pi_L(X), \pi_U(X)] = \left[\inf\left\{S_0^0 E_Q(\hat{X}) : Q \text{ EMM}\right\}, \sup\left\{S_0^0 E_Q(\hat{X}) : Q \text{ EMM}\right\}\right].$$

The preceding theorem shows that this interval coincides except for the boundary with the set (5.3). The latter represents the set of initial values of arbitrage-free derivative price processes in the sense of the previous section. Consequently, the two seemingly different situations and approaches lead ultimately to similar pricing formulas.

In Theorem 5.2 it was stated that the infimum in the definition of the upper price is actually attained. Put differently, one needs exactly the upper price to be able to afford such a *superhedge*, which provides perfect protection against the risk of losses from selling the claim.

Definition 5.3. Let X be a contingent claim. A self-financing strategy φ with $V_0(\varphi) = \pi_U(X)$ and $V_N(\varphi) \geq X$ is called *cheapest superhedge* for X.

If the interval (5.4) is a singleton, the option can be perfectly hedged by Theorem 5.2. In such cases, the questions from the beginning of this section have a clear answer. The bank will charge the premium $\pi_U = \pi_L$ (plus administrative costs), and it is perfectly protected against losses by investing in the replicating portfolio. In the general case, the bank could be tempted to charge the upper price and invest this premium in the cheapest superhedge. As we will see in Example 5.5, however, the upper price may be so large that its cannot be obtained in real markets. We will come back to this issue in Chapter 7.

Examples

We consider now some concrete contingent claims which can be replicated and hence have a unique fair price.

Example 5.4 (European call and put options in the binomial model)**.** In general, European call and put options are not replicable. The situation, however, simplifies in the setting of the Cox-Ross-Rubinstein model, see page 62.

In order to study absence of arbitrage and completeness in this case, we determine the set of equivalent martingale measures Q. Since $\hat{S}^1_n = \hat{S}^1_0 \prod_{m=1}^n (1 + \Delta\hat{X}_m)$ we have that $\hat{S}^1_n = \hat{S}^1_{n-1}(1 + \Delta\hat{X}_n)$. For $\hat{S}^1$ to be a Q-martingale, we need to satisfy

$$E_Q(1 + \Delta\hat{X}_n | \mathscr{F}_{n-1}) = 1 \tag{5.4}$$

for all n. In order to analyse this condition, we focus on any fixed one-period subtree, say in period n. Two children descend from the parent node of this subtree, corresponding to the subevents where $S^1_n/S^1_{n-1} = 1+\Delta\widetilde{X}_n = u$ and $S^1_n/S^1_{n-1} = 1+\Delta\widetilde{X}_n = d$, respectively. In discounted terms, this means $1+\Delta\hat{X}_n = u/(1+\tilde{r})$ and $1+\Delta\hat{X}_n = d/(1+\tilde{r})$, respectively. If we denote the conditional Q-probabilities on the edges by q and $1-q$, respectively, the martingale condition (5.4) boils down to $q\frac{u}{1+\tilde{r}} + (1-q)\frac{d}{1+\tilde{r}} = 1$, i.e.

$$q := \frac{1+\tilde{r}-d}{u-d}.$$

Note that this conditional probability depends neither on the period n nor on the particular one-period subtree under consideration. Considering the tree as a whole, we have found unique transition probabilities that warrant the martingale property of $\hat{S}^1$. Since there is a one-to-one correspondence between probability measures and their transition probabilities on the edges, we conclude that there is a unique equivalent martingale measure Q. In financial terms the Cox-Ross-Rubinstein model under consideration is arbitrage free and complete by Theorems 4.10 and 5.6. Since the transition probabilities q resp. $1-q$ are the same all over the tree, the model has basically the same probabilistic structure under Q as under P, except for p being replaced by q.

Market completeness entails that European call and put options can be hedged perfectly. We want to determine their fair price and their replicating strategy. To this end denote by $Y := (S^1_N - K)^+$ the payoff of a European call with strike $K > 0$. The discounted stock price at maturity equals

$$\hat{S}^1_N = \hat{S}^1_0 (\tfrac{u}{1+\tilde{r}})^U (\tfrac{d}{1+\tilde{r}})^{N-U}$$

where $U = |\{n \in \{1, \dots, N\} : \Delta\hat{X}_n = \frac{u}{1+\tilde{r}} - 1\}|$ denotes the number of upward movements of the stock. Observe that $\Delta\hat{X}_1, \dots \Delta\hat{X}_N$ is a sequence of independent random variables which attain the value $\frac{u}{1+\tilde{r}} - 1$ with probability q under Q. Therefore the number of upward movements U is a binomial random variable with parameters N and q relative to the pricing measure Q. Moreover $\hat{S}^1_N > K/S^0_N$ (i.e. the call finishes in the money) if and only if

$$U > a := \frac{\log(K/S^1_0) - N\log(d)}{\log(u/d)}$$

(i.e. if there are more than a upward movements of the stock price). By Theorem 5.3 the discounted fair price of the European call is obtained as

$$\begin{aligned}
\hat{S}^2_0 &= E_Q(\tfrac{Y}{S^0_N}) \\
&= E_Q((\hat{S}^1_N - \tfrac{K}{S^0_N})^+) \\
&= E_Q((\hat{S}^1_0(\tfrac{u}{1+\tilde{r}})^U(\tfrac{d}{1+\tilde{r}})^{N-U} - \tfrac{K}{S^0_N})1_{(a,\infty)}(U)) \\
&= \textstyle\sum_{\ell=0}^N \binom{N}{\ell} q^\ell (1-q)^{N-\ell} \Big(\hat{S}^1_0(\tfrac{u}{1+\tilde{r}})^\ell(\tfrac{d}{1+\tilde{r}})^{N-\ell} - \tfrac{K}{S^0_N}\Big) 1_{(a,\infty)}(\ell) \\
&= \hat{S}^1_0 \textstyle\sum_{\ell=0}^N \binom{N}{\ell} \tilde{q}^\ell (1-\tilde{q})^{N-\ell} 1_{(a,\infty)}(\ell) - \tfrac{K}{S^0_N} \textstyle\sum_{\ell=0}^N \binom{N}{\ell} q^\ell (1-q)^{N-\ell} 1_{(a,\infty)}(\ell) \\
&= \hat{S}^1_0 (1 - b_{N,\tilde{q}}(a)) - \tfrac{K}{S^0_N}(1 - b_{N,q}(a)) \\
&= \hat{S}^1_0 b_{N,1-\tilde{q}}(N-a) - \tfrac{K}{S^0_N} b_{N,1-q}(N-a),
\end{aligned}$$

where $b_{N,p}$ denotes the cumulative distribution function (cdf) of the binomial distribution with N trials and success probability p and where we set

$$\tilde{q} := \frac{u}{1+\tilde{r}} q = \frac{u(1+\tilde{r}-d)}{(1+\tilde{r})(u-d)}.$$

Along the same lines we obtain

$$\hat{S}^2_n = E_Q(\tfrac{Y}{S^0_N}|\mathscr{F}_n) = \hat{S}^1_n b_{N-n,1-\tilde{q}}(a_n) - \tfrac{K}{S^0_N} b_{N-n,1-q}(a_n)$$

for any intermediate time $n = 0, \dots, N$, where

$$a_n := \frac{\log(S^1_n/K) + (N-n)\log(u)}{\log(u/d)}.$$

(Note that $a_0 = N - a$.) Hence we obtain

$$S^2_n = S^1_n b_{N-n,1-\tilde{q}}(a_n) - K\tfrac{S^0_n}{S^0_N} b_{N-n,1-q}(a_n). \tag{5.5}$$

for the undiscounted price of the option.

In order to compute the hedging strategy we write $S_n^2 = C(S_n^1, n)$ with some function $C : \mathbb{R}_+ \times \{0, \dots, N\} \to \mathbb{R}_+$, which is determined by Equation (5.5). Denote by $\varphi = (\varphi^0, \varphi^1)$ a self-financing strategy which replicates the call by trading (S^0, S^1), i.e. $V_N(\varphi) = (S_N^1 - K)^+ = S_N^2$. For the discounted value process we have $\hat{V}(\varphi) = \hat{S}^2$ and hence $\Delta \hat{S}_n^2 = \varphi_n^1 \Delta \hat{S}_n^1$. If the stock price rises we obtain

$$\tfrac{1}{S_n^0} C(S_{n-1}^1 u, n) - \tfrac{1}{S_{n-1}^0} C(S_{n-1}^1, n-1) = \varphi_n^1 \hat{S}_{n-1}^1 (\tfrac{u}{1+\tilde{r}} - 1),$$

if it falls we get

$$\tfrac{1}{S_n^0} C(S_{n-1}^1 d, n) - \tfrac{1}{S_{n-1}^0} C(S_{n-1}^1, n-1) = \varphi_n^1 \hat{S}_{n-1}^1 (\tfrac{d}{1+\tilde{r}} - 1).$$

Subtracting the two and solving for the number of stocks we have

$$\varphi_n^1 = \frac{C(S_{n-1}^1 u, n) - C(S_{n-1}^1 d, n)}{S_{n-1}^1 (u-d)}.$$

The hedge for the European put is obtained along the same lines or using the *call-put parity* from Example 1.4. At maturity N we have $(S_N^1 - K)^+ - (K - S_N^1)^+ = S_N^1 - K$. The fair price at time n of the contingent claim on the right is $S_n^1 - K S_n^0 / S_N^0$, as will be shown in Example 5.7. The law of one price implies that $S_n^2 - S_n^3 = S_n^1 - K S_n^0 / S_N^0$, $n \in \{0, \dots, N\}$ holds for the call price S_n^2 and the put price S_n^3. After a straightforward calculation one obtains

$$S_n^3 = K \tfrac{S_n^0}{S_N^0} b_{N-n, 1-q}(b_n) - S_n^1 b_{N-n, 1-\widetilde{q}}(b_n)$$

for the put price at time n, where

$$b_n := \frac{\log(K/S_n^1) - (N-n)\log(d)}{\log(u/d)}.$$

The hedging strategy can be determined in the same way as above, up to replacing C by some $\tilde{C}$ such that $S_n^3 = \tilde{C}(S_n^1, n)$ holds.

Example 5.5 (European call and put options in the standard model)**.** Now we turn to the standard market model from Section 4 with Gaussian random variables ΔX_n having mean μ and variance σ^2. According to Theorem 5.8 this cannot be a complete market model. Hence we cannot expect to obtain unique option prices. Instead we want to determine the interval of possible initial prices of a European call, which is limited by the lower and the upper price as in the preceding section. Strictly speaking, we cannot apply the earlier results, which were stated for a finite probability space Ω. However, the results can be extended to the general case with only minor modifications.

Note that $\varphi := (0,1)$ is a superhedge for the call because $(S_N^1 - K)^+ \le S_N^1$. This yields $\pi_U \le S_0^1$. Considering $\varphi := (0,0)$ and $(S_N^1 - K)^+ \ge 0$ we conclude $\pi_L \ge 0$ for the lower price. Moreover, looking at portfolio $\varphi := (-Ke^{-rN}/S_0^0, 1)$ and $(S_N^1 - K)^+ \ge -Ke^{-rN} S_N^0 / S_0^0 + S_N^1$ we obtain $\pi_L \ge S_0^1 - Ke^{-rN}$. Together, this yields

$$\pi_L \ge \max\{0, S_0^1 - Ke^{-rN}\} = (S_0^1 - Ke^{-rN})^+.$$

These almost trivial price bounds π_L, π_L hold irrespective of the underlying process S^1.

We will now construct a one-parametric family of equivalent martingale measures, which are parametrized by $\widetilde{\sigma} > 0$. Since both $N(\mu, \sigma^2)$ and $N(r - \frac{\widetilde{\sigma}^2}{2}, \widetilde{\sigma}^2)$ are equivalent to Lebesgue measure, there exists a density

$$f := \frac{dN(r - \frac{\widetilde{\sigma}^2}{2}, \widetilde{\sigma}^2)}{dN(\mu, \sigma^2)}.$$

We now define a probability measure $Q \sim P$ by its density

$$\frac{dQ}{dP} := \prod_{n=1}^{N} f(\Delta X_n).$$

Relative to this measure, the price process has the same structure as under P, but with different parameters μ and σ^2:

Lemma 5.6. *1. Q is a probability measure which is equivalent to P. Relative to Q, the random variables $\Delta X_1, \ldots, \Delta X_N$ are independent with law $N(r - \frac{\widetilde{\sigma}^2}{2}, \widetilde{\sigma}^2)$.*

2. Q is an equivalent martingale measure.

Proof. We skip the proof of the first statement. For the second note that

$$\begin{aligned}
E_Q(\hat{S}_n^1 | \mathscr{F}_{n-1}) &= E_Q(\hat{S}_{n-1}^1 \exp(\Delta X_n - r) | \mathscr{F}_{n-1}) \\
&= \hat{S}_{n-1}^1 E_Q(e^{\Delta X_n - r}) \\
&= \hat{S}_{n-1}^1 \frac{1}{\sqrt{2\pi\widetilde{\sigma}^2}} \int e^x \exp\left(-\frac{1}{2} \frac{(x + \frac{\widetilde{\sigma}^2}{2})^2}{\widetilde{\sigma}^2} \right) dx \\
&= \hat{S}_{n-1}^1 \frac{1}{\sqrt{2\pi\widetilde{\sigma}^2}} \int \exp\left(-\frac{1}{2} \frac{(x - \frac{\widetilde{\sigma}^2}{2})^2}{\widetilde{\sigma}^2} \right) dx \\
&= \hat{S}_{n-1}^1,
\end{aligned}$$

which yields the Q-martingale property of $\hat{S}$. $\square$

Let us compute the option price that is obtained as an expectation under measure Q, i.e.

$$\pi_Q := S_0^0 E_Q((S_N^1 - K)^+ / S_N^0) = E_Q\Big((S_0^1 e^{X_N - rN} - K e^{-rN})^+ \Big).$$

By the previous lemma $X_N - rN = \sum_{n=1}^N \Delta X_n - rN$ is Gaussian with mean $-\frac{\widetilde{\sigma}^2}{2} N$ and variance $\widetilde{\sigma}^2 N$. Therefore, we have

$$\begin{aligned}
\pi_Q &= \frac{1}{\sqrt{2\pi\widetilde{\sigma}^2 N}} \int \left(S_0^1 e^x - K e^{-rN} \right)^+ \exp(-\tfrac{(x + N\widetilde{\sigma}^2/2)^2}{2\widetilde{\sigma}^2 N}) dx \\
&= \frac{1}{\sqrt{2\pi}} \int \left(S_0^1 \exp(\sqrt{\widetilde{\sigma}^2 N} y - \tfrac{\widetilde{\sigma}^2}{2} N) - K e^{-rN} \right)^+ \exp(-\tfrac{y^2}{2}) dy \\
&= \frac{S_0^1}{\sqrt{2\pi}} \int_{y_0}^{\infty} \exp(\sqrt{\widetilde{\sigma}^2 N} y - \tfrac{\widetilde{\sigma}^2}{2} N) e^{-\frac{y^2}{2}} dy - \frac{K e^{-rN}}{\sqrt{2\pi}} \int_{y_0}^{\infty} e^{-\frac{y^2}{2}} dy
\end{aligned}$$

for

$$y_0 := \frac{\log \frac{K}{S_0^1} - rN + \frac{\widetilde{\sigma}^2}{2}N}{\widetilde{\sigma}\sqrt{N}}.$$

If Φ_{μ,σ^2} denotes the cumulative distribution function of the Gaussian law with mean μ and variance σ^2, we have

$$\begin{aligned}
\pi_Q &= \frac{S_0^1}{\sqrt{2\pi}}\int_{y_0}^{\infty} e^{-\frac{1}{2}(y-\widetilde{\sigma}\sqrt{N})^2}dy - \frac{Ke^{-rN}}{\sqrt{2\pi}}\int_{y_0}^{\infty} e^{-\frac{y^2}{2}}dy \\
&= S_0^1(1-\Phi_{\widetilde{\sigma}\sqrt{N},1}(y_0)) - Ke^{-rN}(1-\Phi_{0,1}(y_0)) \\
&= S_0^1\Phi_{0,1}(-y_0+\widetilde{\sigma}\sqrt{N}) - Ke^{-rN}\Phi_{0,1}(-y_0) \\
&= S_0^1\Phi_{0,1}(d_1) - Ke^{-rN}\Phi_{0,1}(d_2) \qquad (5.6)
\end{aligned}$$

with

$$d_1 := \frac{\log \frac{S_0^1}{K} + rN + \frac{\widetilde{\sigma}^2}{2}N}{\widetilde{\sigma}\sqrt{N}} \quad \text{and} \quad d_2 := \frac{\log \frac{S_0^1}{K} + rN - \frac{\widetilde{\sigma}^2}{2}N}{\widetilde{\sigma}\sqrt{N}}.$$

For $\widetilde{\sigma} \to 0$ we obtain

$$\Phi_{0,1}(d_1), \Phi_{0,1}(d_2) \to \begin{cases} 0 & \text{if } S_0^1 < Ke^{-rN}, \\ \frac{1}{2} & \text{if } S_0^1 = Ke^{-rN}, \\ 1 & \text{if } S_0^1 > Ke^{-rN} \end{cases}$$

and hence $\pi_Q \to (S_0^1 - Ke^{-rN})^+$. For $\widetilde{\sigma} \to \infty$ we get instead $\Phi_{0,1}(d_1) \to 1$, $\Phi_{0,1}(d_2) \to 0$ and hence $\pi_Q \to S_0^1$. Since $\pi_L \leq \pi_Q \leq \pi_U$ and in view of the estimates above we get

$$\pi_L = (S_0^1 - Ke^{-rN})^+ \quad \text{and} \quad \pi_U = S_0^1.$$

Therefore the price limits in the standard model coincide with the trivial ones. Put differently, absence of arbitrage does not provide much information on European call prices.

Example 5.7 (Forward contract)**.** The forward contract is related to a contingent claim with payoff $X := S_N^1 - K$ for some specific $K \in \mathbb{R}$. We assume that the numeraire has deterministic terminal value S_N^0, e.g. because the numeriare asset is altogether deterministic or because it represents a zero coupon bond with maturity N. Consider now the constant and hence self-financing strategy $\varphi := (-K/S_N^0, 1)$ in the market $S = (S^0, S^1)$. Its value process is $V_n(\varphi) = S_n^1 - K\frac{S_n^0}{S_N^0}$. Since $V_N(\varphi) = S_N^1 - K = X$, the claim X is replicable and $V(\varphi)$ is the only reasonable price process of the claim, both in the sense of Sections 5.2 und 5.3. Suppose that the forward contract is settled at time $n \in \{0, 1, \dots, N\}$. The forward price O_n is the value K that makes the contract worthless at time n, i.e.

$$O_n = S_n^1 \frac{S_N^0}{S_n^0} = \hat{S}_n^1 S_N^0.$$

Observe that a forward contract can be hedged perfectly whenever there exists a bond with deterministic payoff at time N. No assumptions must be made about the dynamics of the underlying security S^1. Moreover, the hedge is static in the sense that it does not involve frequent rebalancing. In practice, however, forward contracts may involve a risk that does not turn up in our mathematical model, namely the *counterparty risk* that the other party does not meet its obligations.

Appendix QF

5.A Futures and Illustration

On first glance, futures contracts do not seem to be compatible with the theory because they involve a complex cashflow. But on closer look, they can be interpreted as securities with — sometimes negative — dividend payments. The value of the asset itself is always 0 because entering and terminating the contract is always free of charge. Since a futures contract involves a payment of $\Delta U_n = U_n - U_{n-1}$ at time n, the futures price U can be interpreted as a dividend process satisfying the constraint $U_N = S^1_N$.

Let us formally consider the market $S^0, S^1, (S^2, D^2)$, where S^0 denotes the numeraire, S^1 the underlying of the futures contract, and $(S^2, D^2) = (0, U - U_0)$ the futures contract itself as a dividend-paying asset. We assume that the numeraire asset S^0 is predictable.

If this market does not allow for arbitrage, there exists an equivalent martingale measure Q by Corollary 4.A.6, i.e. there is a probability measure $Q \sim P$ such that $\hat{S}^1$ and $\hat{S}^2 + \hat{D}^2 = \frac{1}{S^0} \bullet U$ are Q-martingales. Hence $U = U_0 + S^0 \bullet (\frac{1}{S^0} \bullet U)$ is a Q-martingale as well. Consequently, we have

$$U_n = E_Q(U_N|\mathscr{F}_n) = E_Q(S^1_N|\mathscr{F}_n)$$

for $n = 0, \dots, N$. Note that the futures price process itself is a Q-martingale, not the *discounted* price process as usual. This does not contradict absence of arbitrage because the futures price does not represent the price of a liquidly traded asset.

If the money market account S^0 is deterministic, it follows that

$$U_n = E_Q(\hat{S}^1_N S^0_N|\mathscr{F}_n) = S^0_N E_Q(\hat{S}^1_N|\mathscr{F}_n) = \hat{S}^1_n S^0_N = S^1_n \frac{S^0_N}{S^0_n},$$

i.e. the futures price coincides with the forward price above (Example 5.7).

In this case the cashflow of the futures can be replicated by trading the numeraire and the underlying. Indeed, by Section 4.A the self-financing strategy $\varphi = (\varphi^0, 0, 1)$ corresponding to one futures contract has the discounted value process

$$\hat{V}(\varphi) = 0 + \frac{1}{S^0} \bullet D^2 = \frac{1}{S^0} \bullet (S^0_N \hat{S}^1) = \frac{S^0_N}{S^0} \bullet \hat{S}^1.$$

This coincides with the discounted value of the self-financing strategy $\psi = (\psi^0, \psi^1, 0)$ which has initial value 0 and holds $\psi^1_n = \frac{S^0_N}{S^0_n}$ shares of the underlying S^1 at time n.

We have derived the equality of futures and forward price processes in the case of deterministic interest rates, without any assumptions on the law of S^1. Interestingly, however,

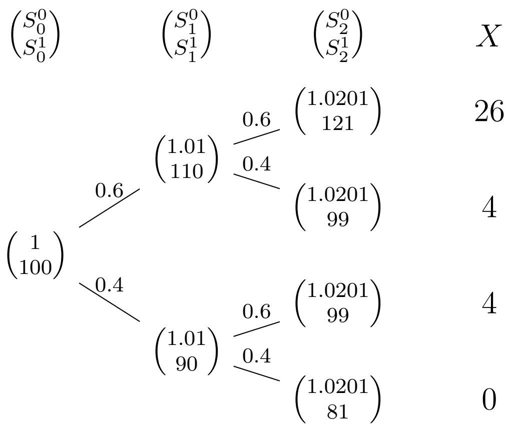

Figure 5.A.1: A contingent claim X in the market of Figure 4.1

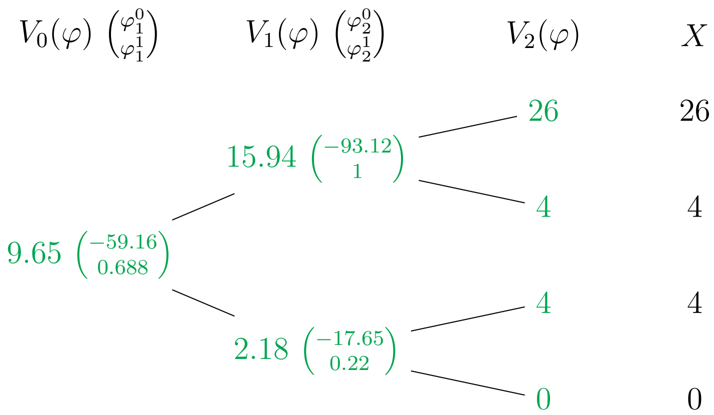

Figure 5.A.2: The replicating strategy φ for X and its value process

the corresponding replicating strategies differ. While a sold forward is hedged by buying exactly one share of the underlying asset, the replicating portfolio of the futures contract contains S_N^0/S_n^0 shares of stock, which is a time-varying number.

Illustration

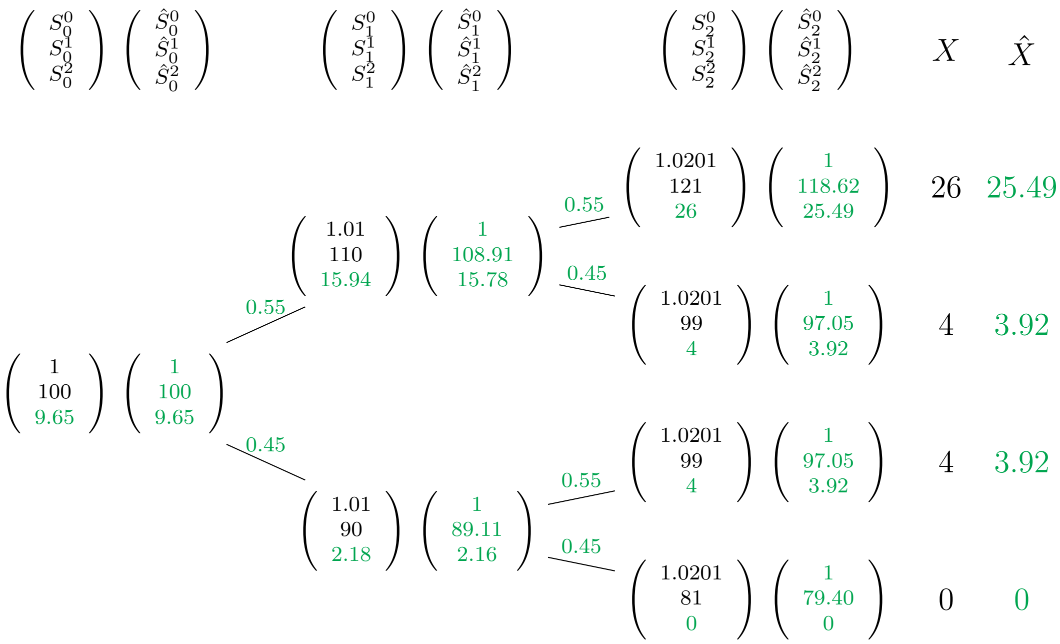

Figure 5.A.3: Computation of the fair price process S^2 of contingent claim X in Figures 5.A.1, 5.A.2 using the martingale measure of Figure 4.6

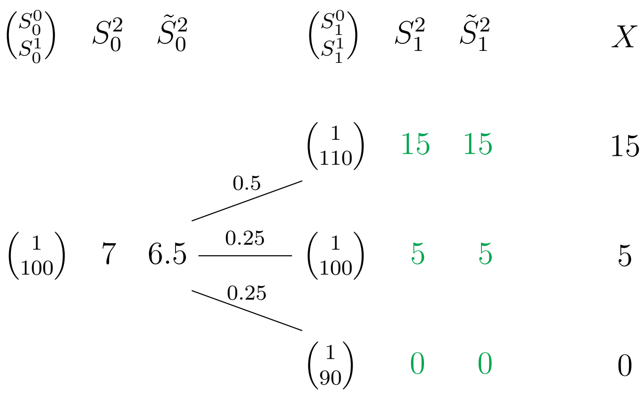

Figure 5.A.4: A contingent claim X in the market of Figure 4.7 and two alternative price processes $S^2, \tilde{S}^2$ for X (corresponding to $Q, \tilde{Q}$ in Figure 4.7) that do not lead to arbitrage

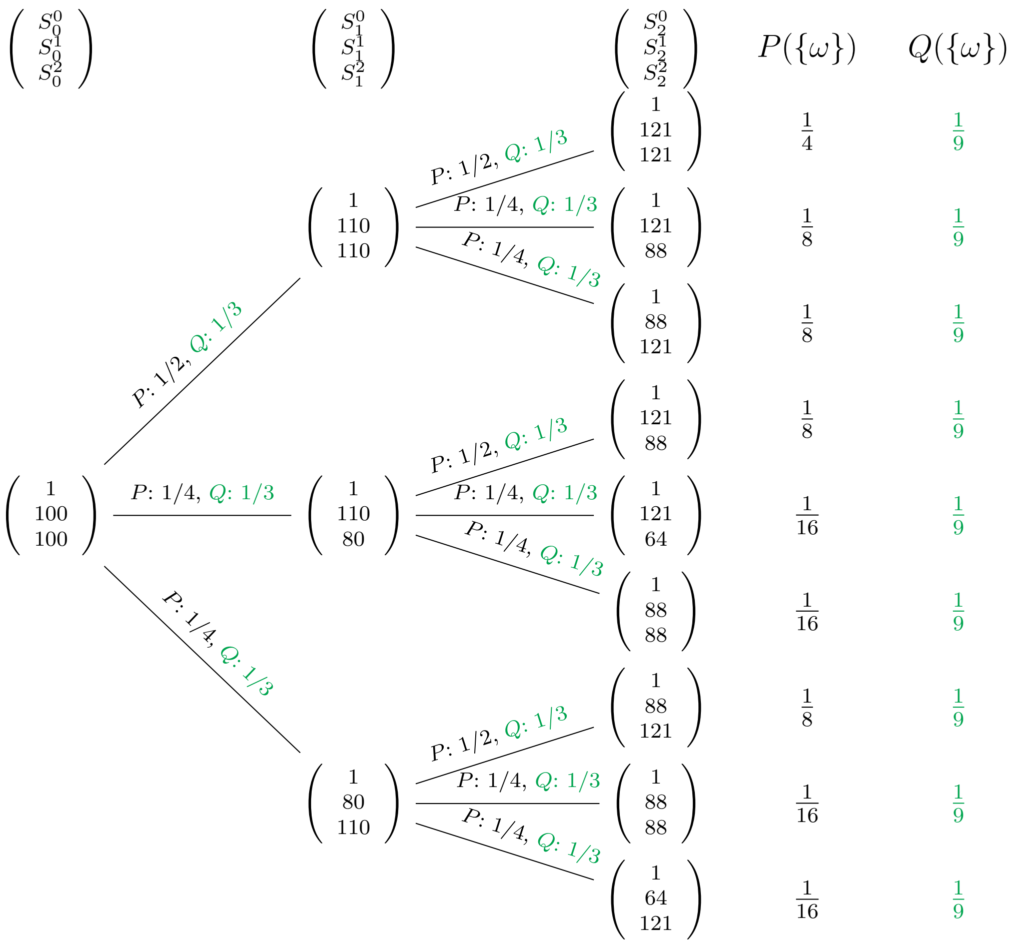

Figure 5.A.5: A complete market with three assets and the corresponding equivalent martingale measure Q

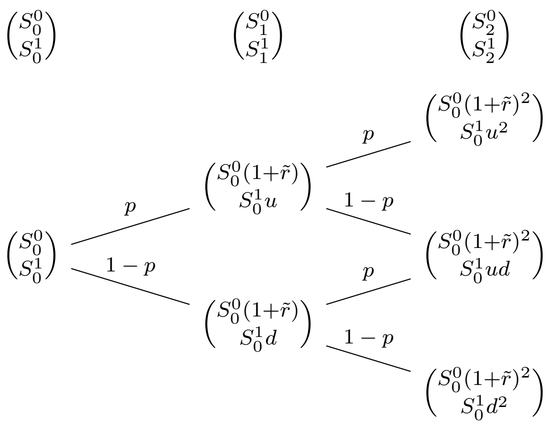

Figure 5.A.6: Cox-Ross-Rubinstein model

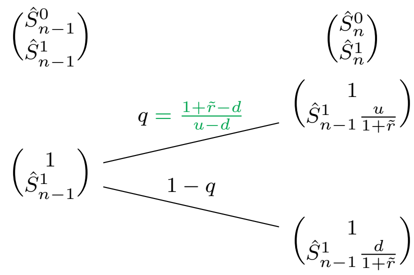

Figure 5.A.7: Subtree of the Cox-Ross-Rubinstein model and transition probabilities of the EMM Q

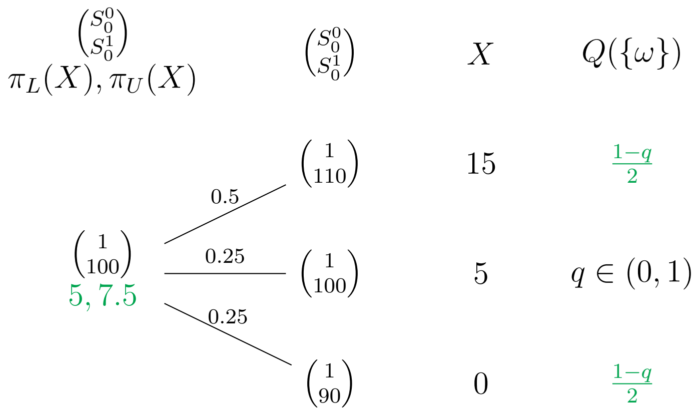

Figure 5.A.8: Lower and upper price of the claim X in Figure 5.A.4 as well as possible EMM's Q

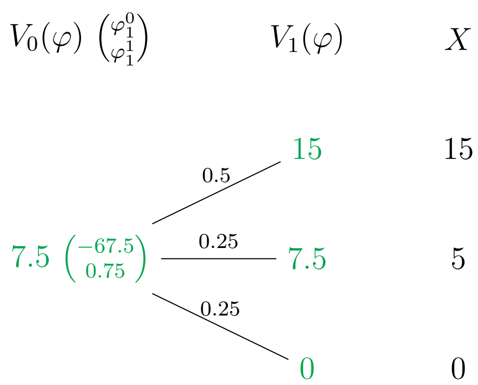

Figure 5.A.9: The cheapest superhedge of X in Figure 5.A.8

Appendix MF

5.B Some Proofs and the Martingale Representation Theorem

Some proofs

Proof of Theorem 5.3. 1⇒2: In order to prove existence set $S^{d+1} = V(\varphi)$ which clearly satisfies $S^{d+1}_N = X$.

We need to show that this choice does not lead to arbitrage in the extended market $(S^0, \dots, S^{d+1})$. So let ψ denote some $\mathbb{R}^{d+1}$-valued predictable process with $\psi \bullet (\hat{S}^1, \dots, \hat{S}^{d+1})_N \geq 0$. We have

$$\begin{aligned} \psi \bullet (\hat{S}^1, \dots, \hat{S}^{d+1})_N &= \textstyle\sum_{i=1}^d \psi^i \bullet \hat{S}^i_N + \psi^{d+1} \bullet (\varphi \bullet \hat{S})_N \\ &= \textstyle\sum_{i=1}^d \psi^i \bullet \hat{S}^i_N + (\psi^{d+1}\varphi) \bullet \hat{S}_N \\ &= \Big(\psi^1 + \psi^{d+1}\varphi^1, \dots, \psi^d + \psi^{d+1}\varphi^d\Big) \bullet \hat{S}_N. \end{aligned}$$

Investing in S according to this strategy and with zero initial money would lead to an arbitrage unless $\Big(\psi^1 + \psi^{d+1}\varphi^1, \dots, \psi^d + \psi^{d+1}\varphi^d\Big) \bullet \hat{S}_N = 0$. Since the original market S is free of arbitrage we thus conclude $\psi \bullet (\hat{S}^1, \dots, \hat{S}^{d+1})_N = 0$.

In order to show uniqueness let S^{d+1} denote an arbitrary semimartingale with terminal value $X = V_N(\varphi)$ and such that $(S^0, \dots, S^{d+1})$ does not allow for arbitrage. By Lemma 4.9 we conclude that $S^{d+1} = V(\varphi)$.

2⇒3: Corollary 5.1

3⇒1: Let Q be a EMM. If X is not replicable, then $\hat{X}$ is not an element of the subspace

$$U := \{\hat{V}_N(\varphi) : \varphi \text{ self-financing strategy } \}$$

of the space of all random variables $\mathbb{R}^\Omega$. By Theorem 4.B.1 there exists $\lambda \in \mathbb{R}^\Omega$ such that $\sum_{\omega\in\Omega} \lambda(\omega)\hat{X}(\omega) > 0$ and $\sum_{\omega\in\Omega} \lambda(\omega)Y(\omega) = 0$ for all $Y \in U$. For the random variable $\xi(\omega) := \lambda(\omega)/Q(\{\omega\})$ it holds that $E_Q(\xi\hat{X}) > 0$ and $E_Q(\xi Y) = 0$ for all $Y \in U$. Define $d\widetilde{Q}/dQ := 1 + \frac{\xi}{2\|\xi\|_\infty}$. Denote the vector which only has entries 1 by $\mathbb{1} \in \mathbb{R}^\Omega$, then $\mathbb{1} \in U$ (invest the money into S^0). Therefore $E_Q(\xi) = E_Q(\xi\mathbb{1}) = \sum_{\omega\in\Omega} \lambda(\omega) = 0$, hence it holds that $\widetilde{Q}$ is a probability measure. Furthermore $E_{\widetilde{Q}}(Y) = E_Q(Y) + \frac{1}{2\|\xi\|_\infty} E_Q(\xi Y) = E_Q(Y)$ for all $Y \in U$. As in the proof of 4.10, we can see that $\hat{S}^1, \dots, \hat{S}^d$ are $\widetilde{Q}$-martingales.

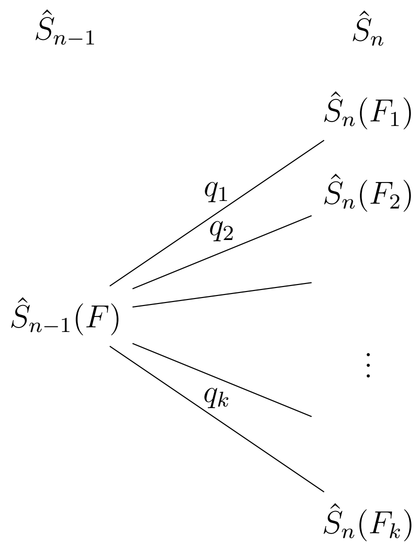

Figure 5.B.1: One-period subtree in the proof of Theorem 5.7

On the other hand $E_{\widetilde{Q}}(\hat{X}) = E_Q(\hat{X}) + \frac{1}{2\|\xi\|_\infty} E_Q(\xi\hat{X}) > E_Q(\hat{X})$ and hence $E_{\widetilde{Q}}(\hat{X}|\mathscr{F}_0) \neq E_Q(\hat{X}|\mathscr{F}_0)$, i.e. claim 3 does not hold true for X. $\square$

Proof of Theorem 5.7. If we represent the filtered probability measure by a tree as in Figure 3.4, any probability measure Q on Ω can be uniquely characterized by conditional probabilities on the edges of this tree. Let us consider a one-period subtree, where one parent event F in $\mathscr{F}_{n-1}$ splits into k disjoint child events $F_1, \dots, F_k$ in $\mathscr{F}_n$. Denote the conditional probabilities on the edges of the subtree by $q_j := Q(F_j|F) = Q(F_j)/Q(F)$, $j = 1, \dots, k$. Moreover, we write $\hat{S}^i_{n-1}(F)$, $i = 0, \dots, d$ for the discounted prices of the $d+1$ assets at time $n-1$ if event F happens. Similarly, $\hat{S}^i_n(F_j)$ for $i = 0, \dots, d$ and $j = 1, \dots, k$ denotes the price of asset i if event F_k happens at time n. Since $F = F_1 \cup \dots \cup F_k$, we have

$$\sum_{j=1}^{k} q_j = 1. \tag{5.7}$$

In order for Q to be a martingale measure, the coefficients q_j must satisfy the $d+1$ equations

$$\sum_{j=1}^{k} \hat{S}^i_n(F_j) q_j = \hat{S}^i_{n-1}(F), \quad i = 0, \dots, d. \tag{5.8}$$

Since $\hat{S}^0 = 1$, the equation for $i = 0$ coincides with (5.7). Hence altogether $d+1$ linear equations have to be met in order for $q_1, \dots, q_k$ to satisfy the requirements of an EMM. If $k > d+1$, then these $d+1$ equations have infinitely many solutions if they have a solution at all. Note that the conditional probabilities can be chosen independently on each one-period subtree because the martingale property holds if and only (5.8) holds separately on any of these subtrees. Market completeness implies that there is only one EMM. Therefore, $k > d+1$ cannot hold, which proves the claim. $\square$

Proof of Theorem 5.2. 1. Let φ be a self-financing strategy with $V_N(\varphi) \geq X$ and let Q denote an EMM. Since $\hat{V}(\varphi)$ is a Q-martingale, it follows that $\hat{V}_0(\varphi) = E_Q(\hat{V}_N(\varphi)) \geq E_Q(\hat{X})$ and hence $\pi_U \geq \sup\left\{S_0^0 E_Q(\hat{X}) : Q \text{ EMM}\right\}$.

To prove the other inequality, let $x = \sup\{S_0^0 E_Q(\hat{X}) : Q \text{ EMM}\}$. It remains to show that there exists a self-financing strategy φ with $V_0(\varphi) = x$ and $V_N(\varphi) \geq X$. If not, then

$$\hat{X} \notin \left\{\frac{x}{S_0^0} + \varphi \bullet \hat{S}_N - Y : \varphi \text{ predictable process, } Y \geq 0 \text{ random variable}\right\}.$$

By Lemma 4.B.3 there exists an EMM Q such that $E_Q(\hat{X} - \frac{x}{S_0^0}) > 0$, hence $E_Q(\hat{X})S_0^0 > x$ which is the desired contradiction.

2. This statement follows from the first one by considering $-X$ instead of X.

3. $\{S_0^0 E_Q(\hat{X}) : Q \text{ EMM}\}$ is an interval because the set of EMM's is convex. To prove that the boundary points do not belong to the interval for non-replicable X, suppose that $\pi_U(X) = S_0^0 E_Q(\hat{X})$ for some EMM Q. Furthermore, let φ be a self-financing strategy with $V_0(\varphi) = \pi_U(X)$ and $V_N(\varphi) \geq X$. As

$$E_Q(\hat{V}_N(\varphi) - \hat{X}) = \hat{V}_0(\varphi) - E_Q(\hat{X}) = \frac{\pi_U(X)}{S_0^0} - \frac{\pi_U(X)}{S_0^0} = 0,$$

we obtain $V_N(\varphi) = X$, hence X is replicable. A similar reasoning leads to the lower bound. If X is replicable, then the set contains one element by Theorem 5.3. □

Martingale Representation Theorem

The following class of processes turns out to be useful for concrete models.

Definition 5.B.1. A *random walk* is a process X with $X_0 = 0$ whose increments $\Delta X_1, \Delta X_2, \ldots$ are independent and identically distributed random variables. We call it *binomial random walk* if the random variables ΔX_n have only two values.

The following property of binomial random walks turns out to be important in the context of option pricing. It is shared only by very few other processes.

Theorem 5.B.2 (Martingale representation theorem)**.** *Let X be a binomial random walk that is a martingale. If $(\mathscr{F}_n)_{n\in\mathbb{N}}$ is the filtration generated by X, then for every martingale Y there exists a predictable process H such that $Y = Y_0 + H \bullet X$.*

Proof. If $a, -b$ are the values of ΔX_n, we have $P(\Delta X_n = a) = 1 - P(\Delta X_n = -b) = p$, whereas we have $ap = b(1-p)$ for p because of the martingale property. Since ΔY_n is $\sigma(\Delta X_1, \ldots, \Delta X_n)$-measurable, there exists a function $f_n : \{-b, a\}^n \to \mathbb{R}$ with $\Delta Y_n = f_n(\Delta X_1, \ldots, \Delta X_n)$. Since Y is a martingale, we have

$$0 = E(\Delta Y_n | \mathscr{F}_{n-1}) = p f_n(\Delta X_1, \ldots, \Delta X_{n-1}, a) + (1-p) f_n(\Delta X_1, \ldots, \Delta X_{n-1}, -b)$$

due to Lemma 2.B.12, therefore

$$\tfrac{1}{a} f_n(\Delta X_1, \ldots, \Delta X_{n-1}, a) = -\tfrac{1}{b} f_n(\Delta X_1, \ldots, \Delta X_{n-1}, -b) =: H_n.$$

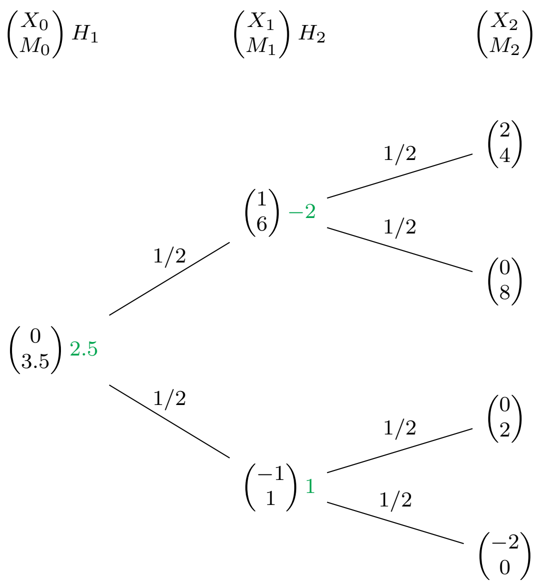

Figure 5.B.2: Martingale representation $M = M_0 + H \bullet X$

Then we have

$$H_n \Delta X_n = f_n(\Delta X_1, \ldots, \Delta X_{n-1}, a) = \Delta Y_n$$

if $\Delta X_n = a$ and analogous for $\Delta X_n = -b$. $\square$

Theorem 5.B.2 is illustrated in Figure 5.B.2. Since any random variable generates a martingale as in Lemma 3.11, the preceding theorem implies that it can be written as the sum of its expected value and the endpoint of some stochastic integral.

The martingale representation theorem can be used to prove completeness in the binomial tree model directly. To this end, consider an arbitrary $\mathcal{F}_N$-measurable random variable $\hat{Y}$ and the associated Q-martingale $Z_n = E_Q(\hat{Y}|\mathcal{F}_n)$. By the martingale representation theorem 5.B.2 applied to the Q-random walk $\hat{X}$ (so to the entire process), there is a real number y and a predictable process H satisfying

$$\hat{Y} = Z_N = y + H \bullet \hat{X}_N = y + \tfrac{H}{\hat{S}^1_-} \bullet \hat{S}^1_N.$$

The remark following Definition 5.2 yields that the market is complete.

Some comments on calls and puts

See Exercises.

Chapter 6

American Options

In this chapter we study how to price American options.

For American options, the payoff depends on the (random) time point where the holder exercises it. Therefore, the theory developed in the previous chapter does not apply: we have no fixed claim X, but rather a process $(X_n)_{n=0,\dots,N}$ of possible payoffs. The exercise time of an American option can therefore be viewed as a stopping time that should be chosen optimally. This indicates why American options are related to *optimal stopping problems*. As before we work in a filtered probability space $(\Omega, \mathfrak{P}(\Omega), (\mathscr{F}_n)_{n=0,\dots,N}, P)$ with finite time horizon $N \in \mathbb{N}$ and trivial initial σ-field.

Theory of optimal stopping

We first study the problem where an expected payoff $E(X_\tau)$ is to be maximized over all stopping times τ. Here X denotes a given stochastic process and $(X_\tau)(\omega) := X_{\tau(\omega)}(\omega)$ the value of this random process at the random time τ. This stopping problem is closely related to the notion of a *Snell envelope* in the following sense.

Definition 6.1. The *Snell envelope* U of an adapted process X is defined recursively via $U_N := X_N$ and

$$U_n := \max\{X_n, E(U_{n+1}|\mathscr{F}_n)\}.$$

Its relation to optimal stopping is stated in the following theorem. The proofs of all results of this section can be found in the appendix.

Theorem 6.2. *The Snell envelope U of an adapted process X has the following properties:*

1. *U is the smallest supermartingale satisfying $U \geq X$.*

2. $$U_n = \max_{\tau \in \mathscr{T}_n} E(X_\tau|\mathscr{F}_n) = E(X_{\tau_f}|\mathscr{F}_n) = E(X_{\tau_s}|\mathscr{F}_n)$$

for $n = 0, \dots, N$, where $\mathscr{T}_n$ denotes the set of stopping times with values in $\{n, n+1, \dots, N\}$ and

$$\begin{aligned}
\tau_f &:= \min\{m \geq n : U_m = X_m\}, \\
\tau_s &:= \min\{m \geq n : X_m > E(U_{m+1}|\mathscr{F}_m) \text{ or } m = N\}.
\end{aligned}$$

The following characterization of Snell envelopes will turn out to be useful later.

Lemma 6.3. *The following statements are equivalent for adapted processes X.*

1. *U is the Snell envelope of X.*
2. *U is a supermartingale with $U \geq X$, $U_N = X_N$ and such that $1_{\{U_- \neq X_-\}} \bullet U$ is a martingale.*

Valuation of American options

The valuation of European options is based on absence of arbitrage and the first fundamental theorem of asset pricing. These arguments in turn involve the possibility to trade any asset at any time in any positive or negative amount. This is only partially true for American options. No problems are involved with buying American options and selling them later. However, holding negative quantities is less obvious. A short position in an American option can always be terminated by the holder, whose right to execute the option may make the asset and hence the seller's short position disappear. On the other hand, we can safely assume that this will not happen as long as the option's market value exceeds the exercise price. Indeed, in this case selling the option on the market results in a higher profit than exercising it. Altogether, any trader of an American option faces certain short selling constraints that depend on the option's market price. We consider now a version of the fundamental theorem covering this situation.

As in Chapter 4 we work in a market with $d+1$ securities, terminal date N and positive numeraire S^0. Short selling constraints are expressed in terms of predictable processes γ^i, $i = 1, \dots, d$, whose only values are 0 and 1. We call the $\{0,1\}^d$-valued process $\gamma = (\gamma^1, \dots, \gamma^d)$ the *short-sale constraint indicator process.* The idea is that asset i cannot be held in negative amounts as long as γ^i equals 1. As long as γ^i equals 0, on the other hand, short-selling of asset i is not restricted. Put differently, we consider the set

$$\Theta := \left\{\varphi \text{ self-financing} : \varphi^i \gamma^i \geq 0 \text{ for } i = 1, \dots, d\right\}$$

of trading strategies.

Definition 6.4. Let γ be a short-sale constraint indicator process. We say that the market does not allow for *γ-arbitrage* if there exists no arbitrage strategy $\varphi \in \Theta$, where Θ is defined as above in terms of γ.

The following result generalizes Theorem 4.10 to such strategies.

Theorem 6.5 (First fundamental theorem under short sale constraints)**.** *If $\gamma = (\gamma^1, \dots, \gamma^d)$ denotes a short-sale constraint indicator process as above, we have equivalence between:*

1. *The market does not allow for γ-arbitrage.*
2. *There exists a probability measure $Q \sim P$ such that $\hat{S}^i$ is a Q-supermartingale and $(1-\gamma^i) \bullet \hat{S}^i$ is a Q-martingale for $i = 1, \dots, d$.*

We skip the proof, which is similar to the one of Theorem 4.10.

As in Chapter 5.2 we consider American options as liquidly traded assets. The goal is to determine the derivative price processes that are consistent with absence of arbitrage. To this end, we suppose that the numeraire asset S^0 is positive and arbitrage opportunities do not exist in our market of $d+1$ assets $(S^0, \dots, S^d)$ excluding the American option.

From the mathematical perspective, an *American option* is defined by a nonnegative adapted process X. The random variable X_n represents the payoff that the holder receives if she exercises the option at time n. The corresponding *discounted exercise process* is defined as $\hat{X} := \frac{X}{S^0}$. We denote the option's market price process as S^{d+1}. If the option has not been exercised prematurely, its terminal value equals $S^{d+1}_N = X_N$. Evidently, $S^{d+1} \geq X$ should always hold. Indeed, if the option's market price fell below the exercise price, one could buy it for S^{d+1}_n and at the same time receive the higher value X_n by exercising it immediately. This can be viewed as a simple form of arbitrage.

As argued above, trading American options involves certain short sale constraints. A negative number φ^{d+1}_n of options in the portfolio may not be upheld while $S^{d+1}_{n-} = X_{n-}$ because the option may be exercised and hence vanish from the market. As motivated above, this is not going to happen as long as the market value exceeds the exercise price. Consequently, we are facing trading constraints of the form

$$\gamma^{d+1}_n := \begin{cases} 1 & \text{if } S^{d+1}_{n-} = X_{n-}, \\ 0 & \text{otherwise,} \end{cases}$$

i.e.

$$\gamma^{d+1} = 1_{\{S^{d+1}_- = X_-\}}.$$

We can now derive an analogue of Corollary 5.1 for American options.

Corollary 6.6. *The following statements are equivalent for an American option with exercise process X.*

1. *The market $(S^0, \dots, S^{d+1})$ obtained by adding derivative price process S^{d+1} does not allow for arbitrage. More precisely, S^{d+1} is an adapted process with $S^{d+1} \geq X$, $S^{d+1}_N = X_N$ and such that the market does not allow for $(0, \dots, 0, \gamma^{d+1})$-arbitrage.*
2. *There is an equivalent martingale measure Q for the market $(\hat{S}^0, \dots, \hat{S}^d)$ such that the discounted derivative price process $\hat{S}^{d+1}$ is the Snell envelope of $\hat{X}$ relative to probability measure Q.*

Proof. 2⇒1: Lemma 6.3 yields that $\hat{S}^{d+1}$ is a Q-supermartingale and $(1-\gamma^{d+1}) \bullet \hat{S}^{d+1}$ is a Q-martingale. The assertion follows now from Theorem 6.5.

1⇒2: According to Theorem 6.5 there is a probability measure $Q \sim P$ such that $\hat{S}^1, \dots, \hat{S}^d$ are Q-martingales, $\hat{S}^{d+1}$ is a Q-supermartingale, and $1_{\{S_-^{d+1} \neq X_-\}} \bullet \hat{S}^{d+1}$ is a Q-martingale. By Lemma 6.3, we conclude that $\hat{S}^{d+1}$ is the Q-Snell envelope of $\hat{X}$. □

In complete markets, there exists a unique price process for the American option which does not lead to arbitrage because there is only one EMM Q for the underlying market. We call it the *fair price process* of the claim.

Let us now consider a market where both an American option with exercise process X and a European option with terminal payoff X_N are traded. If this market does not allow for arbitrage, the American option must be at least as expensive as the European one at any instance. Indeed, this follows from Corollary 6.6 but it can also be seen directly. If the American option happens to be cheaper than the European one, one buys the American option, shorts the European one at the same time, and invests the positive difference in the numeraire. At maturity N the payoff X_N and the liability $-X_N$ of the two options cancel. The investment in the numeraire remains as arbitrage gain.

By contrast, it is natural to expect that the American option should have a higher market value than the corresponding European one because of the additional right to choose the exercise time. Surprisingly, this is not the case for call options on stocks that do not pay dividends:

Theorem 6.7. *Consider a market $S = (S^0, S^1, S^2, S^3)$ where S^2, S^3 denote the price processes of a European and an American call option on S^1 with strike price K and maturity N (i.e. $X_N = (S_N^1 - K)^+$ is the terminal payoff of S^2, and $X = (S^1 - K)^+$ is the exercise process of S^3). If the numeraire S^0 is increasing and the market does not allow for $(0,0,0,1_{\{S_-^3 = X_-\}})$-arbitrage, then $S^2 = S^3$.*

Proof. By Corollary 6.6 there is some $Q \sim P$ such that $\hat{S}^2$ is a Q-martingale and $\hat{S}^3$ is the Q-Snell envelope of $\hat{X}$. Now, note first that $f(x) = (x-K)^+$ satisfies $cf(x) \geq f(cx)$ for every $0 \leq c \leq 1$. Therefore, since S^0 is increasing, we obtain

$$\begin{aligned} \hat{S}_n^2 &= E_Q\Big(f(S_N^1)/S_N^0 \Big| \mathscr{F}_n\Big) \\ &= E_Q\Big(f(S_N^1)S_n^0/S_N^0 \Big| \mathscr{F}_n\Big)/S_n^0 \\ &\geq E_Q\Big(f(S_n^0 \hat{S}_N^1) \Big| \mathscr{F}_n\Big)/S_n^0. \end{aligned}$$

Secondly, we use Jensen's inequality for conditional expectations which states that $E(g(X)|\mathcal{F}) \geq g(E(X|\mathcal{F}))$ for random variables X, convex functions g and any σ-field $\mathcal{F}$. Since f as above is a convex function, Jensen's inequality yields

$$\begin{aligned} \hat{S}_n^2 &\geq E_Q\Big(f(S_n^0 \hat{S}_N^1) \Big| \mathscr{F}_n\Big)/S_n^0 \\ &\geq f\Big(S_n^0 E_Q(\hat{S}_N^1 | \mathscr{F}_n)\Big)/S_n^0 \\ &= f(S_n^1)/S_n^0 \\ &= \hat{X}_n. \end{aligned}$$

Together with the Q-martingale property of $\hat{S}^2$ we have

$$\begin{aligned}
\hat{S}_n^3 &= \max_{\tau \in \mathscr{T}_n} E_Q(\hat{X}_\tau | \mathscr{F}_n) \\
&\leq \max_{\tau \in \mathscr{T}_n} E_Q(\hat{S}_\tau^2 | \mathscr{F}_n) \\
&= \hat{S}_n^2.
\end{aligned}$$

In view of

$$\hat{S}_n^3 = \max_{\tau \in \mathscr{T}_n} E_Q(\hat{X}_\tau | \mathscr{F}_n) \geq E_Q(\hat{X}_N | \mathscr{F}_n) = \hat{S}_n^2,$$

the assertion follows. The idea of the proof is that early exercise never pays because the market price of the European call always exceeds the payoff of the American option. □

If the numeraire price process S^0 is constant or decreasing (i.e. the numeraire has non-positive return), the statement in Theorem 6.7 holds for put options as well.

American put options typically do not allow for a simple pricing formula, not even in the simple Cox-Ross-Rubinstein model. However, the fair price can be computed recursively. Note that the situation is simpler for the American call because its price coincides with the European call by the previous theorem.

Example 6.8. We consider the binomial model from Section 5.4 with a riskless asset S^0 and a risky stock S^1. Let S^2 denote the fair price of an American put option on S^1 with strike price $K \in \mathbb{R}$, i.e. with payoff process $X = (K - S^1)^+$. We want to determine S_n^2 recursively for $n = N, N-1, \dots, 0$. To this end, denote by Q the unique equivalent martingale measure in the CRR model. Obviously, we have $S_N^2 = (K - S_N^1)^+$. Since $\hat{S}^2$ is the Q-Snell envelope of X/S^0, we have

$$\hat{S}_n^2 = \max\left\{ \left(\frac{K}{S_n^0} - \hat{S}_n^1 \right)^+, E_Q(\hat{S}_{n+1}^2 | \mathscr{F}_n) \right\}. \tag{6.1}$$

One can show that

1. $\hat{S}_n^2 = g_n(\hat{S}_n^1)$ for some function $g_n : \mathbb{R} \to \mathbb{R}$, i.e. S_n^2 depends only on the present stock price S_n^1 but not on the whole past $S_0^1, \dots, S_n^1$,
2. $g_n(\hat{S}_n^1)$ is decreasing (and convex) in $\hat{S}_n^1$,
3. there is some threshold x_n such that

$$\hat{S}_n^2 = \begin{cases} E_Q(\hat{S}_{n+1}^2 | \mathscr{F}_n) & \text{for } \hat{S}_n^1 > x_n, \\ \left(\frac{K}{S_n^0} - \hat{S}_n^1 \right)^+ & \text{for } \hat{S}_n^1 \leq x_n. \end{cases}$$

   The interpretation is that the American option price coincides with the payoff if the stock price is below some threshold.

The expectation in (6.1) can be expressed more explicitly. We have

$$\begin{aligned}
\hat{S}_n^2 &= \max\left\{ (\tfrac{K}{S_n^0} - \hat{S}_n^1)^+, E_Q(g_{n+1}(\hat{S}_{n+1}^1) | \mathscr{F}_n) \right\} \\
&= \max\left\{ (\tfrac{K}{S_n^0} - \hat{S}_n^1)^+, E_Q(g_{n+1}(\hat{S}_n^1(1 + \Delta \hat{X}_{n+1})) | \mathscr{F}_n) \right\} \\
&= \max\left\{ (\tfrac{K}{S_n^0} - \hat{S}_n^1)^+, q g_{n+1}(\hat{S}_n^1 \tfrac{u}{1+\tilde{r}}) + (1-q) g_{n+1}(\hat{S}_n^1 \tfrac{d}{1+\tilde{r}}) \right\},
\end{aligned}$$

which means that the functions $g_N, \ldots, g_0$ can be determined recursively by $g_N(x) = (K/S_N^0 - x)^+$ and

$$g_n(x) = \max\left\{(\tfrac{K}{S_n^0} - x)^+, qg_{n+1}(x\tfrac{u}{1+\tilde{r}}) + (1-q)g_{n+1}(x\tfrac{d}{1+\tilde{r}})\right\}.$$

Appendix QF

6.A An illustration and a second perspective

The individual perspective on American options

Finally, let us consider American options from the individual perspective of Chapter 5.3, where over-the-counter deals are considered. In this case the option is not traded liquidly but instead offered to the potential buyer by a bank. We assume that the market $(S^0, \dots, S^d)$ is complete, i.e. there exists a unique EMM Q. Above we have seen that if the options were liquidly traded,

$$\pi := S_0^0 \max\left\{E_Q(\hat{X}_\tau) : \tau \text{ stopping time}\right\}$$

would be the only initial option price that does not lead to arbitrage. The following result shows that this price π is reasonable from the point of view of OTC deals as well.

Lemma 6.A.1. *At price π the bank can purchase a self-financing portfolio φ that warrants perfect protection against losses in the sense that $V(\varphi) \geq X$. If, however, the bank sells the option for a premium $x < \pi$, there is an arbitrage opportunity for the buyer.*

Proof. We have $\pi/S_0^0 = U_0$, where U denotes the Q-Snell envelope of $\hat{X}$. We write $U = U_0 + M + A$ for the Q-Doob decomposition of U, i.e. M is a Q-martingale and A a predictable decreasing process with $M_0 = 0 = A_0$.

By market completeness there exists a self-financing strategy φ with $\hat{V}_N(\varphi) = U_0 + M_N$. The martingale property of M implies $\hat{V}(\varphi) = U_0 + M$ and in particular $\hat{V}_0(\varphi) = U_0$, i.e. the bank can afford φ. If the holder of the option exercises at time n, she obtains the discounted payoff

$$\hat{X}_n \leq U_n \leq U_0 + M_n = \hat{V}_n(\varphi),$$

i.e. the value of the hedge portfolio φ is large enough to cover this obligation.

We turn now to the second claim. Since $\pi = S_0^0 E_Q(\hat{X}_{\tau_f})$ for an appropriately chosen stopping time τ_f, we can consider π as unique fair price of a European contingent claim with discounted payoff $\hat{X}_{\tau_f}$. Suppose that the investor shorts the perfect hedge of this option and consequently obtains π. For the smaller amount x she buys the American option. The difference is invested in the numeraire asset. At time τ_f the investor exercises the American option and invests the payoff X_{τ_f} in the numeriare asset. The discounted terminal value of this investment coincides with and hence covers the liabilities from

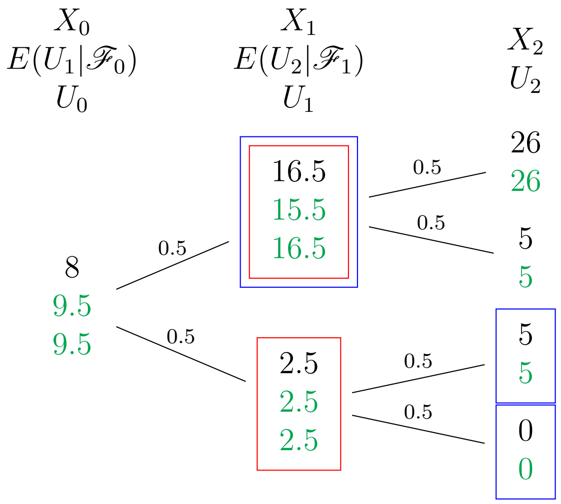

Figure 6.A.1: Snell envelope U of process X and stopping times τ_f (red) and τ_s (blue)

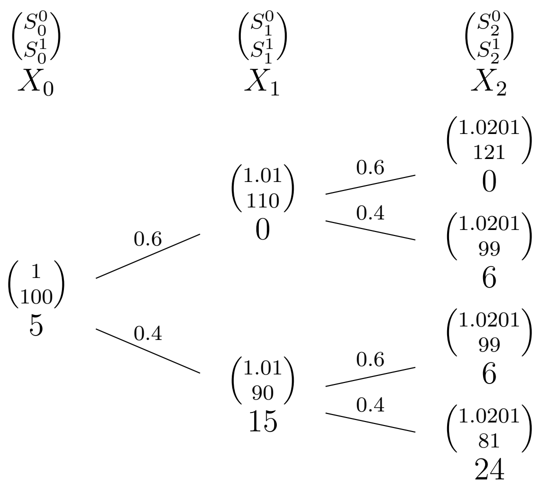

Figure 6.A.2: An American contingent claim X (namely a put) in the market of Figure 4.1

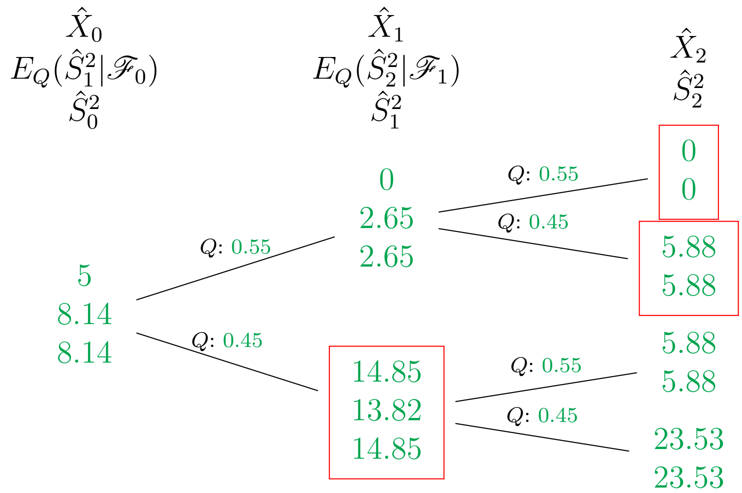

Figure 6.A.3: The discounted exercise process $\hat{X}$, martingale measure probabilities and the American option's fair price process $\hat{S}^2$ in Figure 6.A.2

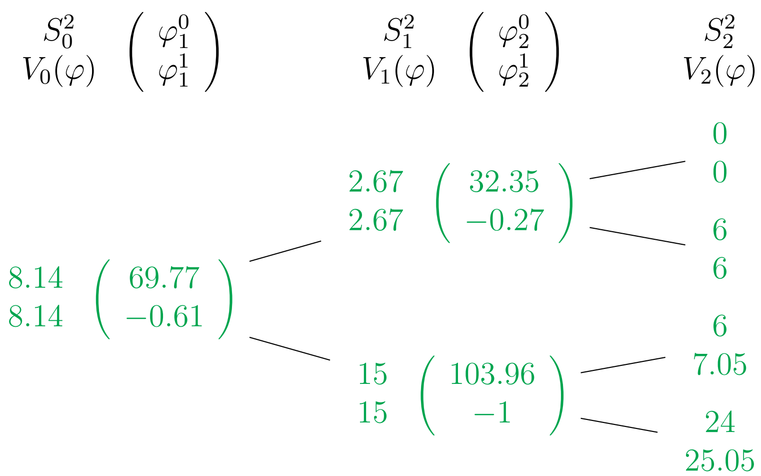

Figure 6.A.4: The fair price S^2 of the American claim in Figures 6.A.2, 6.A.3, its perfect hedge φ and the value process of φ

the shorted hedge of the European option. Hence the investor profits from the riskless discounted gain of $\pi - x$ stemming from the initial trade. $\square$

Appendix MF

6.B Snell-envelope theorem and applications

An elementary approach to certain optimal stopping problems

Usually, optimal stopping problems – i.e. the problems of maximizing $E(X_\tau)$ over all stopping times τ – are difficult to solve.

A reasonable stopping rule is the following: Comparing the current gain with what to expect in the next step leads to the stopping time

$$\tau^* = \inf\{n \geq 0 : X_n \geq E(X_{n+1}|\mathscr{F}_n)\}.$$

It is called the *one-step look ahead rule* or, as we will use here, the *myopic rule*. In general, there is no reason to assume that such a rule is optimal.

In this section, we consider an interesting class of problems in which optimality can be obtained. The key for this is the notion of *monotone stopping problems*. Here, we call the above problem a *monotone case problem* iff for all n it holds that

$$E(X_{n+1}|\mathscr{F}_n) \leq X_n \quad \Longrightarrow \quad E(X_{n+2}|\mathscr{F}_{n+1}) \leq X_{n+1}.$$

Using sets in the notation this may be written as

$$\{E(X_{n+1}|\mathscr{F}_n) \leq X_n\} \subseteq \{E(X_{n+2}|\mathscr{F}_{n+1}) \leq X_{n+1}\} \text{ for all } n.$$

We now provide a short review using the Doob decomposition, which leads to a shortcut to optimality results without the usual machinery of optimal stopping theory. For every n let

$$\begin{aligned} M_n &= \sum_{k=0}^{n-1} (X_{k+1} - E(X_{k+1}|\mathscr{F}_k)), \ M_0 = 0, \\ A_n &= \sum_{k=0}^{n-1} (E(X_{k+1}|\mathscr{F}_k) - X_k) = \sum_{k=0}^{n-1} Y_k, \ A_0 = 0, \end{aligned}$$

so that by telescoping expectation terms we have the Doob decomposition

$$X_n = X_0 + M_n + A_n$$

with a zero mean martingale $(M_n)_n$. For the myopic stopping time τ^*

$$Y_k := E(X_{k+1}|\mathscr{F}_k) - X_k > 0 \text{ for } k = 0, \ldots, \tau^* - 1$$

– valid for all k if $\tau^* = \infty$ – and in the monotone case

$$E(X_{k+1}|\mathscr{F}_k) - X_k \leq 0 \text{ for } k = \tau^*, \tau^* + 1, \dots.$$

Thus we have

$$A_{\tau^*} = \sup_n A_n$$

and, using $\tau_L^* = \min\{\tau^*, L\}$ for some fixed L,

$$A_{\tau_L^*} = \sup_{n \leq L} A_n.$$

So, for any stopping time τ, not necessarily finite a.s.,

$$A_\tau \leq A_{\tau^*}, \quad A_{\min\{\tau, L\}} \leq A_{\tau_L^*}.$$

This yields the following

Lemma 6.B.1. *In the monotone case optimal stopping problem fix $L \in \mathbb{N}$. Then for all $\tau \leq L$*

$$EX_\tau \leq EX_{\tau_L^*},$$

i.e. τ_L^* is optimal in the class of all stopping times bounded by L.

Proof.

$$EX_\tau = EX_0 + EA_\tau \leq EX_0 + EA_{\tau_L^*} = EX_{\tau_L^*}$$

using Theorem 3.17. □

The extension from the finite to the infinite case uses the approximation $\tau = \lim_{L\to\infty} \tau_L$, $\tau_L = \min\{\tau, L\}$, so that $X_\tau = \lim_{L\to\infty} X_{\tau_L}$ on $\{\tau < \infty\}$, but on $\{\tau = \infty\}$ we need a specific definition of X_∞. Here, we use

$$X_\infty = \liminf_{n\to\infty} X_n,$$

so that $X_\tau = \liminf_{L\to\infty} X_{\tau_L}$ always. We introduce the following conditions:

$$\lim_{L\to\infty} EX_{\tau_L^*} \leq EX_{\tau^*} \tag{V1}$$

$$\sup\{EX_\tau : \tau \text{ bounded}\} = \sup_\tau EX_\tau \tag{V2}$$

Note for (V1) that $EX_{\tau_L^*}$ is increasing in L.

Theorem 6.B.2. *Under* (V1) *and* (V2), *τ^* is optimal, i.e.*

$$EX_{\tau^*} = \sup_\tau EX_\tau.$$

Proof. We have from Lemma 6.B.1 and (V1) that

$$EX_{\tau^*} \geq \lim_{L\to\infty} \sup_{\tau \leq L} EX_\tau = \sup\{EX_\tau : \tau \text{ bounded}\},$$

proving the claim by (V2). □

Example 6.B.3. [House selling problem] Assume we have a house to sell and the potential buyers come one after the other and make offers, which we model as i.i.d. integrable random variables $(Z_n)_{n\in\mathbb{N}}$. For each observation, we have fixed costs of $c>0$. We assume a situation with recall, that is, even if they are gone, we can later on access the best offer (seller's market). Therefore, our net payoff is given by

$$X_n = \max\{Z_1,\dots,Z_n\} - cn.$$

The filtration is of course the one generated by $X_1, X_2, \ldots$.
We have – using Lemma 2.B.12 –

$$\begin{aligned} Y_k &:= E(X_{k+1}|\mathscr{F}_k) - X_k = E(\max\{Z_1,\dots,Z_k\}\vee Z_{k+1}|\mathscr{F}_k) - \max\{Z_1,\dots,Z_k\} - c \\ &= f(S_k) - c \end{aligned}$$

where

$$S_k = \max\{Z_1,\dots,Z_k\},\ f(z) = E((Z_1 - z)^+).$$

Since f is non-increasing in z and the S_k are non-decreasing in k, the process $(Y_k)_{k\in\mathbb{N}}$ is non-increasing in k. Hence, the problem is obviously monotone. The myopic stopping time

$$\tau^* = \inf\{k\ge 0 : f(S_k)\le c\}.$$

is optimal if we can verify (V1) and (V2). This can be done under the additional condition of finite variance for Z_n.

Example 6.B.4. [Burglar's problem] Here, we again have an i.i.d. sequence $(Z_n)_{n\in\mathbb{N}}$ and an independent i.i.d. sequence $(\delta_n)_{n\in\mathbb{N}}$, where $Z_n\ge 0$ describes the burglar's gain and $\delta_n = 0$ or $=1$ when getting caught or not caught, resp. Then, we look at

$$X_n = \left(\sum_{j=1}^n Z_j\right)\prod_{j=1}^n \delta_j$$

with the obvious interpretation. This problem is a monotone case problem. Indeed, writing $p = E\delta,\ a = EZ_1$ it holds that

$$\begin{aligned} Y_k &= E\left(\left(\sum_{j=1}^k Z_j + Z_{k+1}\right)\prod_{j=1}^k \delta_j\delta_{k+1}\middle|\mathscr{F}_k\right) - X_k \\ &= X_k p + \left(\prod_{j=1}^k \delta_j\right) ap - X_k = X_k(p-1) + \left(\prod_{j=1}^k \delta_j\right) ap, \end{aligned}$$

which obviously changes sign just once. Hence the myopic rule is optimal and stops iff

$$\prod_{j=1}^k \delta_j = 0 \text{ or } \sum_{j=1}^k Z_j \ge \frac{ap}{1-p}.$$

Proof of the Snell-envelope theorem

Proof of Theorem 6.2. 1. By definition we have

$$U_n = \max\{X_n, E(U_{n+1}|\mathscr{F}_n)\} \geq E(U_{n+1}|\mathscr{F}_n)$$

for $n = 0, \dots, N-1$, which implies the supermartingale property of U. Moreover, $U \geq X$ is obvious.

Consider now an arbitrary supermartingale $V \geq X$. We show recursively that $V_n \geq U_n$ for $n = N, \dots, 0$. We start with $V_N \geq X_N = U_N$. For $n = N-1$ the supermartingale property yields

$$V_{N-1} \geq E(V_N|\mathscr{F}_{N-1}) \geq E(U_N|\mathscr{F}_{N-1}).$$

Since we also have $V_{N-1} \geq X_{N-1}$, we get

$$V_{N-1} \geq \max\{X_{N-1}, E(U_N|\mathscr{F}_{N-1})\} = U_{N-1}.$$

Proceeding in the same way for $n = N-2, \dots, 0$ yields $V \geq U$.

2. For any stopping time $\tau \in \mathscr{T}_n$ we have

$$E(X_\tau|\mathscr{F}_n) \leq E(U_\tau|\mathscr{F}_n) = E(U_N^\tau|\mathscr{F}_n) \leq U_n^\tau = U_n,$$

because U^τ is a supermartingale by Lemma 3.16. By $\tau_f, \tau_s \in \mathscr{T}_n$ it remains to be shown that $U_n = E(X_{\tau_f}|\mathscr{F}_n) = E(X_{\tau_s}|\mathscr{F}_n)$.

Recursively we show that

$$E(U_{\max\{\tau_f, m\}}|\mathscr{F}_m) = U_m \tag{6.2}$$

for $m = N, \dots, n$. For $m = N$ this is obvious. For $m = N-1$ let $A := \{\tau_f \leq N-1\} \in \mathscr{F}_{N-1}$. Observe that $E(1_A U_{\max\{\tau_f, N-1\}}|\mathscr{F}_{N-1}) = E(1_A U_{N-1}|\mathscr{F}_{N-1}) = 1_A U_{N-1}$. Also

$$\begin{aligned}
E(1_{A^C} U_{\max\{\tau_f, N-1\}}|\mathscr{F}_{N-1}) &= E(1_{A^C} U_{\max\{\tau_f, N\}}|\mathscr{F}_{N-1}) \\
&= 1_{A^C} E(E(U_{\max\{\tau_f, N\}}|\mathscr{F}_N)|\mathscr{F}_{N-1}) \\
&= 1_{A^C} E(U_N|\mathscr{F}_{N-1}) \\
&= 1_{A^C} U_{N-1},
\end{aligned}$$

because $U_{N-1} = E(U_N|\mathscr{F}_N - 1)$ if $N-1 < \tau_f$. Together we obtain $E(U_{\max\{\tau_f, N-1\}}|\mathscr{F}_{N-1}) = U_{N-1}$. Repeating the argument for $m = N-2, \dots, 0$ yields (6.2).

For $m = n$ we obtain $U_n = E(U_{\tau_f}|\mathscr{F}_n) = E(X_{\tau_f}|\mathscr{F}_n)$. The assertion for τ_s follows along the same lines. □

Proof of Lemma 6.3. $\Rightarrow$: According to Theorem 6.2 it remains to be shown that $1_{\{U_- \neq X_-\}} \bullet U$ is a martingale. By definition of the Snell envelope we have

$$E(1_{\{U_{n-1} \neq X_{n-1}\}} \Delta U_n|\mathscr{F}_{n-1}) = 1_{\{U_{n-1} \neq X_{n-1}\}}(E(U_n|\mathscr{F}_{n-1}) - U_{n-1}) = 0$$

for $n = 1, \dots, N$, which implies the desired martingale property.

$\Leftarrow$: Again by Theorem 6.2 it suffices to show that $V \geq U$ holds for any supermartingale $V \geq X$. Obviously, we have $V_N \geq U_N$. If $U_{N-1} = X_{N-1}$, then $V_{N-1} \geq U_{N-1}$. Moreover, since V is a supermartingale and $1_{\{U_- \neq X_-\}} \bullet U$ is a martingale, we have

$$\begin{aligned}
1_{\{U_{N-1} \neq X_{N-1}\}}(V_{N-1} - U_{N-1}) &\geq 1_{\{U_{N-1} \neq X_{N-1}\}}(E(V_N|\mathscr{F}_{N-1}) - U_{N-1}) \\
&\geq 1_{\{U_{N-1} \neq X_{N-1}\}}(E(U_N|\mathscr{F}_{N-1}) - U_{N-1}) \\
&= E(1_{\{U_{N-1} \neq_{N-1}\}} \Delta U_N|\mathscr{F}_{N-1}) = 0
\end{aligned}$$

Together, this implies $V_{N-1} \geq U_{N-1}$. Repeating the argument for $N-2, \ldots, 0$, it follows that $V_n \geq U_n$ for $n = N, \ldots, 0$. □

Best Choice Problem using the Snell-envelope

A manager has to select one of a group of N applicants for a job one after the other, where each applicant is interviewed and immediately after each interview it has to be decided whether the job seeker is hired or not. It is not possible to make use of previously rejected applicants, and a decision rule is sought that selects the best candidate with as high a probability as possible.

Let us assume that the N applicants can be assessed in terms of their qualifications for the advertised position with the absolute rankings $1, \ldots, N$ (1 is the rank of the best candidate). It is assumed that the $N!$ permutations are equally likely to occur.

At the interview of the k-th candidate, the manager knows the rank of his interview partner only in comparison to the previous applicants, so that a decision on hiring the k-th candidate is only based on the relative ranks $R_1, \ldots, R_k$. R_i denotes the rank of the i-th candidate among the first i applicants.

A formal description can be given as follows:

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_N) : \omega \text{ permutation of } \{1, \ldots, N\}\},$$

$\mathscr{F}$ is the power set and P denotes the uniform distribution on $(\Omega, \mathscr{F})$.
Furthermore, $X_i(\omega) = \omega_i$ denotes the absolute rank of the i-th applicant.
The relative ranks R_i are defined via

$$R_i(\omega) = |\{k \leq i : \omega_k \leq \omega_i\}| = |\{k \leq i : X_k(\omega) \leq X_i(\omega)\}|.$$

At time i, i.e. at the i-th interview, the information is given by

$$\mathscr{F}_i = \sigma(R_1, \ldots, R_i).$$

We try to find a stopping time τ^* w.r.t. $(\mathscr{F}_i)_{i=1,\ldots,n}$ such that the probability to choose the best candidate, i.e. the candidate with absolute rank 1, is maximized by τ^*. We use backward induction with time parameter $\mathscr{T} = \{1, \ldots, N\}$, which is of course similar to using $\mathscr{T} = \{0, 1, \ldots, N\}$.
It holds

$$\begin{aligned}
P(X_\tau = 1) &= \sum_{i=1}^N P(\tau = i, X_i = 1) \\
&= \sum_{i=1}^N \int_{\{\tau = i\}} 1_{\{X_i = 1\}} dP = E\hat{Z}_\tau,
\end{aligned}$$

where $\hat{Z}_i = 1_{\{X_i=1\}},\ i = 1, \ldots, N$.
Note that $\hat{Z}_i$ is not $\mathscr{F}_i$-measurable ($i < N$), since at time i we only know the relative ranks. The definition of $\hat{Z}_i$ uses, however, the knowledge of all ranks.
To circumvent this problem, we define

$$Z_i = E(\hat{Z}_i|\mathscr{F}_i) = P(X_i = 1|R_1, \ldots, R_i),$$

so that for each stopping time τ we obtain using $\{\tau = i\} \in \mathscr{F}_i$ that

$$\begin{aligned} EZ_\tau &= \sum_{i=1}^N \int\limits_{\{\tau=i\}} Z_i dP = \sum_{i=1}^N \int\limits_{\{\tau=i\}} \hat{Z}_i dP \\ &= E\hat{Z}_\tau = P(X_\tau = 1). \end{aligned}$$

To apply the backward induction principle to the best choice problem, we first have to find the underlying distributions.

Using elementary combinatorics, we obtain that $R_1, \ldots, R_n$ are independent with[1]

$$P(R_i = k) = \frac{1}{i} \text{ for } k = 1, \ldots, i,\ i = 1, \ldots, N,$$

and for Z_i it holds

$$Z_i = P(X_i = 1|R_1, \ldots, R_i) = \frac{i}{N} 1_{\{R_i=1\}},\ i = 1, \ldots, N.$$

Using this, we can find a simple expression for U_i. Indeed,

$$U_i = \max\{Z_i, v_{i+1}\} \text{ for } i = 1, \ldots, N,$$

where $v_{N+1} := 0$ and $v_{i+1} := E(U_{i+1})$ otherwise.
This is clear for $i = N$ as $U_N = Z_N = \max\{Z_N, 0\}$.
Assume by induction that

$$U_i = \max\{Z_i, v_{i+1}\}.$$

Then U_i is measurable w.r.t. $\sigma(R_i)$, as Z_i is measurable w.r.t. $\sigma(R_i)$.
As R_i and $(R_1, \ldots, R_{i-1})$ are independent, we obtain

$$\begin{aligned} U_{i-1} &= \max\{Z_{i-1}, E(U_i|R_1, \ldots, R_{i-1})\} \\ &= \max\{Z_{i-1}, EU_i\} = \max\{Z_{i-1}, v_i\}. \end{aligned}$$

[1] The formal argument is as follows:
We first prove the second statement. Denoting the set of permutations of $1, ..., k$ by S_k, we have

$$P(R_i = k) = P(\{\omega : |\{l \le i : \omega_l \le \omega_i\}| = k\}) = \frac{|\{\tilde{\omega} \in S_i : \tilde{\omega}_i = k\}|}{|S_i|} = \frac{(i-1)!}{i!} = \frac{1}{i}.$$

For the first claim, note that $R_1(\omega), \ldots, R_N(\omega)$ totally determines ω, so that for all $r_i \in \{1, ...i\}$, $i = 1, ..., N$

$$P(R_1 = r_1, \ldots, R_N = r_N) = \frac{1}{N!} = \frac{1}{1} \cdots \frac{1}{N} = P(R_1 = r_1) \cdots P(R_N = r_N),$$

where we used the second claim.

The optimal stopping time is of the form

$$\tau^* = \inf\{i : Z_i = U_i\} = \inf\{i : Z_i \geq v_{i+1}\}.$$

Note that $v_1 \geq v_2 \geq \ldots \geq v_N = EZ_N = \frac{1}{N} > 0$, and as furthermore for $i < N$

$$Z_i \geq v_{i+1} \text{ iff } R_i = 1 \text{ and } \frac{i}{N} \geq v_{i+1},$$

we obtain

$$\tau^* = \inf\{i \geq r^* : R_i = 1 \text{ or } i = N\}$$

with

$$r^* = \inf\{i : \frac{i}{N} \geq v_{i+1}\}.$$

The optimal stopping problem is therefore reduced to the problem of finding r^* such that for the stopping time

$$\sigma_r = \inf\{i \geq r : R_i = 1 \text{ or } i = N\}$$

it holds that

$$EZ_{\sigma_{r^*}} = \max_{r \in \mathscr{T}} EZ_{\sigma_r}.$$

As the first applicant always has relative rank 1, we obtain

$$EZ_{\sigma_1} = EZ_1 = \frac{1}{N}.$$

For $r = 2, \ldots, N$ it holds that

$$\begin{aligned}
EZ_{\sigma_r} &= \sum_{i=r}^{N} \int_{\{\sigma_r = i\}} Z_i dP = \sum_{i=r}^{N} \frac{i}{N} P(\sigma_r = i, R_i = 1) \\
&= \frac{r}{N} P(R_r = 1) + \sum_{i=r+1}^{N} \frac{i}{N} P(R_r > 1, \ldots, R_{i-1} > 1, R_i = 1) \\
&= \frac{1}{N} + \sum_{i=r+1}^{N} \frac{i}{N} \frac{r-1}{r} \frac{r}{r+1} \cdots \frac{i-2}{i-1} \frac{1}{i} \\
&= \frac{r-1}{N} \sum_{i=r}^{N} \frac{1}{i-1}.
\end{aligned}$$

Write $h(r) = EZ_{\sigma_r}$, it holds that $h(1) \leq h(2)$ and

$$h(r+1) - h(r) = \frac{1}{N} \left((\sum_{i=r}^{N-1} \frac{1}{i}) - 1 \right).$$

If we define

$$r^* = \inf\{r : \sum_{i=r}^{N-1} \frac{1}{i} \leq 1\},$$

then h is increasing for $r = 1, \ldots, r^*$ and decreasing for $r = r^* + 1, \ldots, N$. Hence, σ_{r^*} is the optimal stopping time.

Last, we consider how $r^* = r^*(N)$ and $v = v(N)$ depend on N; more precisely, we show

$$\lim_{N \to \infty} \frac{r^*(N)}{N} = \lim_{N \to \infty} v(N) = \frac{1}{e}.$$

By definition of $r^*(N)$ it holds that

$$\sum_{i=r^*(N)}^{N-1} \frac{1}{i} \leq 1 < \sum_{i=r^*(N)-1}^{N-1} \frac{1}{i},$$

hence

$$\int_{r^*(N)}^{N} \frac{1}{y} dy \leq 1 < \int_{r^*(N)-2}^{N-1} \frac{1}{y} dy.$$

We obtain

$$\log\left(\frac{N}{r^*(N)}\right) \leq 1 < \log\left(\frac{N-1}{r^*(N)-2}\right),$$

yielding

$$\lim_{N\to\infty} \frac{r^*(N)}{N} = \frac{1}{e},$$

and analogously

$$\lim_{N\to\infty} v(N) = \frac{1}{e}.$$

To summarize, the best strategy in the best choice problem is to not select the first $r^* \approx \frac{1}{e}N$ candidates and then accept the first candidate who is better then all other candidates seen before. Using this rule, the best candidate is selected with probability $v \approx \frac{1}{e}$.

Markovian stopping situation

Stopping problems simplify dramatically in the case that the underlying process has a Markovian structure.

Definition 6.B.5. Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a stochastic process with values in some measurable space $(E, \mathcal{E})$ and adapted to the filtration $(\mathscr{F}_n)_{n\in\mathbb{N}_0}$ with $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

X is called *stationary Markov sequence* w.r.t. $\mathscr{F}$ with initial state $z \in E$ if

(i) $X_0 = z$.

(ii) There exists a mapping $Q : \mathcal{E} \times E \to [0,1]$ such that

$Q(\cdot, x)$ is a probability measure for each x,

$Q(B, \cdot)$ is measurable for all $B \in \mathcal{E}$

and for all n it holds that

$$P(X_{n+1} \in B \mid \mathscr{F}_n) = Q(B, X_n).$$

Q is called *transition probability* and E is called *state space*.

From the definition, we see that

$$E(h(X_{n+1}) \mid \mathscr{F}_n) = E(h(X_{n+1}) \mid X_n),$$

$$E(h(X_{n+1}) \mid X_n = x) = \int h(y)Q(dy, x)$$

for all measurable $h : E \to \mathbb{R}$ with existing expectation $Eh(X_{n+1})$.

Example 6.B.6. Let $Y_1, Y_2, ...$ be i.i.d. with values in some space $\mathcal{Y}$. Let $(E, \mathcal{E})$ be another measurable space and $h : E \times \mathcal{Y} \to E$ be measurable. For $z \in E$ we define $X_0 = X_0^z = z$ and recursively for all $n \geq 1$

$$X_n = X_n^z = h(X_{n-1}^z, Y_n).$$

We obtain a Markov sequence w.r.t. $\mathscr{F}_n = \sigma(Y_1, .., Y_n)$.
The transition probabilities read as

$$\begin{aligned} P(X_{n+1} \in B \mid Y_1, \dots, Y_n) &= P(h(X_n, Y_{n+1}) \in B \mid Y_1, \dots, Y_n) \\ &= P^{Y_1}(\{y : h(X_n, y) \in B\}) \\ &= Q(B, X_n). \end{aligned}$$

The transition probability

$$Q(B, x) = P(h(x, Y_1) \in B)$$

is here independent of the initial state z. One example for such a construction is the asset price process in the CRR model.

Now, we combine the Markov sequences introduced above for different initial states:

Definition 6.B.7. Let $(E, \mathcal{E})$ be a measurable space. For all $z \in E$ let $X^z = (X_n^z)_{n \in \mathbb{N}_0}$ be a stationary Markov sequence with state space E, initial state z and transition probability Q independent of z.
Then, $(X^z)_{z \in E}$ is called a *stationary Markov system*.

For a stationary Markov system $(X^z)_{z \in E}$ and all $i \in \mathbb{N}_0$ let $h_i : E \to \mathbb{R}$ be measurable and $Z_i^z = h_i(X_i^z)$. Assume $E|Z_i^z| < \infty$ for all i, z.
For each $z \in E$ consider the problem of optimal stopping for $(Z_i^z)_{i=0,\dots,N}$. First, we give a heuristic explanation: A stationary Markov sequence $X_{i+1}^z, \dots, X_N^z$ evolves for given $X_i^z(\omega) = x$ the same way as $X_1^x, \dots, X_{N-i}^x$, and hence $h_{i+1}(X_{i+1}^z), \dots, h_N(X_N^z)$ as $h_{i+1}(X_1^x), \dots, h_N(X_{N-i}^x)$. Hence it is plausible to assume the the maximal value we get when stopping $h_{i+1}(X_{i+1}^z), \dots, h_N(X_N^z)$ for given $X_i^z(\omega) = x$ is the same as the maximal we can get by stopping $h_{i+1}(X_1^x), \dots, h_N(X_{N-i}^x)$.
For the exact formulation and for the proof we define for all $i = 0, \dots, N-1$, $k = 1, \dots, N-i$ and $z \in E$

$$w_i^k(z) = \sup_{\tau \in \mathscr{T}_1^k} E h_{i+\tau}(X_\tau^z).$$

Theorem 6.B.8. *In the Markovian stopping situation, the following holds true:*

(i) *For each* $i = 0, 1, \ldots, N-1$ *and* $z \in E$

$$E({}^zU_{i+1} \mid \mathscr{F}_i) = w_i^{N-i}(X_i^z),$$

where the Snell envelope ${}^zU$ *is given by*

$$\begin{aligned} {}^zU_N &= Z_N^z, \\ {}^zU_i &= \max\{Z_i^z, E({}^zU_{i+1} \mid \mathscr{F}_i)\},\ i = N-1, \ldots, 0. \end{aligned}$$

In particular,

$$v(z) = v_0^N(z) = \max\{h_0(z), w_0^N(z)\},\ \ v_1^N(z) = w_0^N(z).$$

(ii) *Write*

$$B_k^N = \{x \in E : h_k(x) \geq w_k^{N-k}(x)\},\ k = 0, \ldots, N-1,\ B_N^N = E.$$

Then, for each $i = 0, 1, \ldots, N$ *and* $z \in E$ *the stopping time*

$$\sigma_i^z = \inf\{k \geq i : X_k^z \in B_k^N\}$$

is optimal in $\mathscr{T}_i^N$*, hence*

$$EZ^z_{\sigma_i^z} = v_i^N(z).$$

The main message of the theorem is that in the Markovian framework we can solve the stopping problem just by calculating certain functions (which seems to be much easier than finding general conditional expectations). *Proof.* As a main tool, we use backwards induction in the proof.

(a) We first show that ${}^zU_i$ is $\sigma(X_i^z)$-measurable.
In the case $i = N$ it obviously holds that ${}^zU_N = Z_N^z = h_N(X_N^z)$.
So, assume that $1 \leq i \leq N$ is such that the claim holds true. By the Markov property

$$\begin{aligned} {}^zU_{i-1} &= \max\{h_{i-1}(X_{i-1}^z), E({}^zU_i \mid \mathscr{F}_{i-1})\} \\ &= \max\{h_{i-1}(X_{i-1}^z), E({}^zU_i \mid X_{i-1}^z)\}. \end{aligned}$$

(b) Next, we obtain the representation $E({}^zU_{i+1} \mid \mathscr{F}_i) = f_i^N(X_i^z)$, where $f_i^N : E \to \mathbb{R}$ is measurable and does not depend on z. For the backwards induction, we first consider $i = N-1$. Then

$$\begin{aligned} E({}^zU_N \mid \mathscr{F}_{N-1}) &= E(h_N(X_N^z) \mid \mathscr{F}_{N-1}) \\ &= \int h_N(y) Q(dy, X_{N-1}^z) \\ &= f_{N-1}^N(X_{N-1}^z) \end{aligned}$$

where $f_{N-1}^N(z) = \int h_N(y) Q(dy, z)$.

Now let $1 \leq i \leq n$ be such that the claim holds true. Then

$$\begin{aligned} E({}^zU_i \mid \mathscr{F}_{i-1}) &= E(\max\{h_i(X_i^z), f_i^N(X_i^z)\} \mid X_{i-1}^z) \\ &= \int \max\{h_i(y), f_i^N(y)\} Q(dy, X_{i-1}^z) \\ &= f_{i-1}^N(X_{i-1}^z) \end{aligned}$$

where $f_{i-1}^N(z) = \int \max\{h_i(y), f_i^N(y)\} Q(dy, z)$.

(c) For given z, i let $N \geq i+1$. Consider $\tau \in \mathscr{T}_{i+1}^n$ given by

$$\tau = \inf\{k \geq i+1 : X_k^z \in B_k\},$$

where $B_{i+1}, ..., B_n$ are measurable subsets of E.
Then, we have the representation

$$Z_\tau^z = h(X_{i+1}^z, \ldots, X_n^z),$$

where h is given by

$$h(y_1, \ldots, y_{n-i}) = h_{k+i}(y_k) \text{ on } B_{i+1}^c \times \ldots \times B_{i+k-1}^c \times B_{i+k}.$$

Hence,

$$\begin{aligned} E(Z_\tau^z \mid \mathscr{F}_i) &= E(h(X_{i+1}^z, \ldots, X_N^z) \mid \mathscr{F}_i) \\ &= Eh(X_1^x, \ldots, X_{N-i}^x) \text{ with } x = X_i^z. \end{aligned}$$

Let

$$\sigma = \inf\{k \geq 1 : X_k^x \in B_{i+k}\}.$$

Then

$$h(X_1^x, \ldots, X_{N-i}^x) = h_{i+\sigma}(X_\sigma^x),$$

which implies

$$E(Z_\tau^z \mid \mathscr{F}_i) = Eh_{i+\sigma}(X_\sigma^x) \text{ with } x = X_i^z.$$

(d) Now, consider the stopping time

$$\begin{aligned} {}^z\tau_{i+1}^N &= \inf\{k \geq i+1 : Z_k^z = {}^zU_k\} \\ &= \inf\{k \geq i+1 : h_k(X_k^z) \geq E({}^zU_{k+1} \mid \mathscr{F}_k)\} \\ &= \inf\{k \geq i+1 : X_k^z \in E_k\} \end{aligned}$$

where $E_k = \{z : h_k(z) \geq f_{k+1}^N(z)\}$.
We define σ by ${}^z\tau_{i+1}^N$ and obtain as in (c) that

$$\begin{aligned} E({}^zU_{i+1} \mid \mathscr{F}_i) &= E(Z^z_{{}^z\tau_{i+1}^n} \mid \mathscr{F}_i) \\ &= Eh_{i+\sigma}(X_\sigma^x) \leq w_i^{N-i}(X_i^z). \end{aligned}$$

It remains to prove the other implication.

(e) For given $y \in E$ we consider the stopping problem

$$h_{i+1}(X_1^y), \ldots, h_n(X_{n-i}^y).$$

The previous considerations show the the optimal stopping time σ^* is of the form

$$\sigma^* = \inf\{i \leq n-i : X_i^z \in D_i\}$$

for suitable $D_1, \ldots, D_{n-i} \subset E$. We define

$$\tau = \inf\{k \geq i+1 : X_k^x \in D_{k-i}\} \text{ with } x = X_i^z(\omega).$$

Then, we obtain from (c):

$$\begin{aligned} E(Z_\tau^z \mid \mathscr{F}_i) &= E(h_{i+\sigma^*}(X_{\sigma^*}^x)) \\ &= \omega_i^{N-i}(X_i^z). \end{aligned}$$

Backwards induction yields

$$E({}^zU_{i+1} \mid \mathscr{F}_i) \geq E(Z_\tau^z \mid \mathscr{F}_i).$$

This gives the other inequality.

(f) Taking expectations, we obtain using backwards induction

$$v(z) = v_0^N(z) = \max\{h_0(z), w_0^N(z)\},\ v_1^N(z) = w_0^N(z).$$

Part (ii) can now be inferred from (i). □

Chapter 7

Incomplete markets

In the preceding chapters we have seen that the modest assumption of absence of arbitrage may have far-reaching consequences for valuation and hedging of derivatives. In other cases arbitrage theory does not provide much useful information. Recall that the market price of a European call option in the standard model may be arbitrarily close to the current stock price, even if the strike is very high. Such extreme prices are rather unrealistic because they would turn the option into an exceptionally attractive asset. Similarly, superhedging strategies as in Chapter 5.3 are not worth considering in the standard model. Indeed, the seller will hardly receive any premium close to the upper price, which is needed to afford the cheapest superhedge but which coincides with the stock price in the standard model. In this chapter we consider the question how to model and hedge non-attainable contingent claims. As before, we distinguish exchange-traded options and OTC contracts.

Martingale modelling

Recall from Section 5.2 that the only reasonable derivative price processes are obtained via conditional expectation relative to some equivalent martingale measure Q. But who chooses Q? The market does as a result of supply and demand! Arbitrage theory only limits its choice by imposing the constraint that the prices need to be consistent with some EMM. But how do we obtain a mathematical model for derivative price processes, without knowing the manifold preferences of the agents and their effects on market prices?

We consider here an approach which relies on ideas from statistics and which is related to what is done in practice. The idea is to use observed quotes of liquidly traded options—so-called *vanilla options*—in order to obtain information on the market's pricing measure Q.

We proceed as follows. Denote by $\hat{S}^1$ resp. $\hat{S}^2, \dots, \hat{S}^d$ the discounted price processes of the stock and of $d-1$ options on the stock. By the fundamental theorem 4.10 there is a probability measure $Q \sim P$ such that all discounted price processes—of the stock $\hat{S}^1$ as well as of all derivatives $\hat{S}^2, \dots, \hat{S}^d$—are Q-martingales. Unfortunately, Q cannot be determined by statistical methods directly because the observed frequencies are subject to the physical probability measure P rather than the pricing measure Q.

But similarly as in statistics we may assume that Q belongs to a certain parametric (or possibly nonparametric) set; cf. Example 7.1 below). More precisely, we express the dynamics of the stock relative to Q in terms of a parametric model. The parameter vector ϑ_Q must be chosen such that $\hat{S}^1$ is a Q-martingale. Since Q is supposed to be a martingale measure for the whole market, the price processes $\hat{S}^2, \dots, \hat{S}^d$ and in particular the initial prices $\hat{S}^2_0, \dots, \hat{S}^d_0$ are obtained by the martingale conditions

$$\hat{S}^i_n = E_Q(\hat{S}^i_N | \mathscr{F}_n)$$

and in particular

$$\hat{S}^i_0 = E_Q(\hat{S}^i_N). \tag{7.1}$$

If the options are liquidly traded, the parameter vector ϑ_Q must be chosen such that the theoretical option prices (7.1) coincide—at least approximately— with observed market quotes at time 0. In mathematical terms, this parallels computing a moment estimator in statistics. "Estimating" ϑ_Q by equating theoretical and observed option prices is called *calibration*. If the dimension of the parameter vector is smaller than the number of observed options, a perfect fit is typically impossible. Instead, one needs to use an approximation, e.g. by least squares. In the non-parametric case, on the other hand, it makes sense to apply methods from nonparametric statistics.

Strictly speaking, two situations should be distinguished. If one only wants to express option prices in terms of the underlying stock, it suffices to consider the pricing measure Q. There is no real need to determine the market's dynamics relative to the real-world measure P as well. If, on the other hand, one wants to make statements on physical probabilities, quantiles, expectations etc., P must be considered as well.

Often one chooses the same parametric class of models for both the P- and the Q-dynamics of the stock. The parameter vector ϑ_P is obtained by statistical estimation based on past data of the stock price S^1, whereas the corresponding Q-parameters are obtained by calibration as stated above. Note, however, that the parameter vectors ϑ_P and ϑ_Q are possibly linked by the fact that measures Q and P need to be equivalent, i.e. the set of possible and impossible scenarios must be the same. This constraint implies that statistical estimation of P-parameters may in fact yield some information on Q-parameters.

How can we assess whether the more or less arbitrarily chosen class of possible Q-dynamics is appropriate? In contrast to statistical estimation of P-models, this decision cannot be based on statistical tests because observed frequencies are not subject to Q-probabilities. However, some evidence can be obtained from observed option prices. If no parameter vector ϑ_Q leads to theoretical option prices that fit observed market quotes reasonably well, then the chosen parametric class of models obviously fails to explain the market under consideration. But as in statistics, one should beware of choosing a high-dimensional class which allows to fit almost any conceivable set of market quotes. Such an *overfitting* may manifest itself after a few days in terms of a growing discrepancy between theoretical and observed option prices. One may be tempted to solve this problem by frequent *recalibration*, i.e. by repeatedly choosing new parameter vectors ϑ_Q. This common practice, however, contradicts our general paradigm. Indeed, the martingale measure Q in the Fundamental Theorem 4.10 is supposed to be fixed once and for all. In other words, it depends neither on time nor on the options under consideration.

Example 7.1. (Calibration in the standard model using implied volatility) We consider the standard market model of Section 5.5. Denote by $Q_{\widetilde{\sigma}}$, $\widetilde{\sigma} > 0$, the equivalent martingale measures which were derived there. We suppose that all market prices are consistent (in the sense of the FTAP 4.10) with some EMM from the corresponding parametric family $\{Q_{\widetilde{\sigma}} : \widetilde{\sigma} > 0\}$. The market's $Q_{\widetilde{\sigma}}$ can be inferred from the observed price of a call by solving

$$\pi_Q = S_0^1 \Phi(d_1) - K e^{-rN} \Phi(d_2)$$

for $\widetilde{\sigma}$, cf. (5.6). The solution $\widetilde{\sigma}$ is called the option's *implied volatility*. If several options are observed, all such equations must be solved by the same value $\widetilde{\sigma}$. If this cannot be achieved in reasonable approximation, the parametric family of martingale measure does not seem to be well suited for explaining the market.

The calibrated model can now be used to price further, not yet liquidly traded options. Indeed, if we assume that Q is a pricing measure as in the first fundamental theorem of asset pricing, we obtain any initial option price as usual by computing Q-expectations of the discounted payoff.

However, one should be careful for at least two reasons. Firstly, applying the calibrated model to new options amounts to extrapolation. It may happen that two different parametric families that have been calibrated successfully to a set of derivatives yield very different prices for new contingent claims. Secondly, it is not clear what the computed option prices mean to the single investor. Absence of arbitrage does not imply that the option can be hedged perfectly or at least reasonably well.

7.1 Variance-optimal hedging

We turn back to the questions raised in Section 5.3. We already discussed superreplication, which extends the idea of perfect hedging to non-replicable claims. However, the necessary initial endowment, i.e. the upper price of the claim often turns out to be excessively high in concrete models. As derived in Section 5.5, the cheapest superhedge of an arbitrary call in the standard model is to buy the stock.

The literature provides a number of suggestions how to hedge at least partially against the risk of losses. We give a brief introduction to the concept of variance-optimal hedging. The idea is to approximate the contingent claims as well as possible by a replicable payoff. Hedging instruments may be the underlying stock but also liquid derivatives if they are available on the market.

The general setup is as in Section 5.3, but we work under P and focus for simplicity on the case that the discounted price process $\hat{S}$ is a martingale. This assumption may appear to be somewhat rough, but probably acceptable for hedging purposes. In the general case one can derive similar, but considerably more involved results. As in Section 5 we consider a contingent claim X whose discounted payoff is denoted as $\hat{X} := X/S_N^0$. In the present section we consider only discounted quantities, as usual expressed by the ˆ notation.

The following definition specifies the idea of a best approximation:

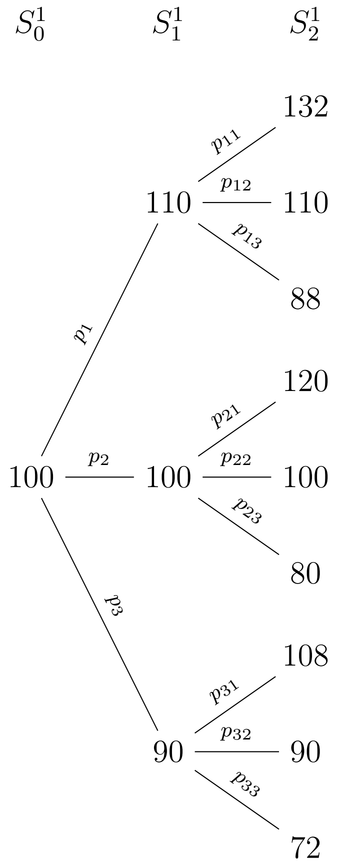

Figure 7.0.1: An (incomplete) trinomial tree model ($S^0 = 1$) where we assume that $q_1 = q_{11} = q_{21} = q_{31}$. It can be made complete if we know that the price of a European call with strike $K = 100$ is 7.5.

Definition 7.1. We call φ *variance-optimal hedging strategy* for X if $\varphi = \vartheta$ minimizes the expected squared hedging error

$$E\Big((\hat{V}_N(\vartheta) - \hat{X})^2\Big)$$

over all self-financing trading strategies ϑ.

The value process of such a strategy φ is unique:

Lemma 7.2. *Any two variance-optimal hedging strategies have the same discounted wealth process $\hat{V}(\varphi)$.*

Proof. For two variance-optimal hedging strategies φ and $\widetilde{\varphi}$ define $\psi := (\varphi + \widetilde{\varphi})/2$. If $\hat{V}_N(\varphi) \neq \hat{V}_N(\widetilde{\varphi})$, we have

$$\begin{aligned} E\Big((\hat{V}_N(\psi) - \hat{X})^2\Big) &= E\Big((\tfrac{1}{2}(\hat{V}_N(\varphi) - \hat{X}) + \tfrac{1}{2}(\hat{V}_N(\widetilde{\varphi}) - \hat{X}))^2\Big) \\ &< \tfrac{1}{2}E((\hat{V}_N(\varphi) - \hat{X})^2) + \tfrac{1}{2}E((\hat{V}_N(\widetilde{\varphi}) - \hat{X})^2) \\ &= E((\hat{V}_N(\varphi) - \hat{X})^2) \end{aligned}$$

which contradicts the optimality of φ. Hence we have $\hat{V}_N(\varphi) = \hat{V}_N(\widetilde{\varphi})$. By Lemma 4.9 we even have $\hat{V}(\varphi) = \hat{V}(\widetilde{\varphi})$. $\square$

In the language of functional analysis, the discounted terminal value $\hat{V}_N(\varphi)$ of a variance-optimal hedge is the orthogonal projection of the discounted payoff $\hat{X}$ on the subspace of stochastic integrals relative to $\hat{S}$, shifted by the initial endowment. This fact plays a role in the following martingale decomposition, which in turn will be the key to variance-optimal hedging.

Theorem 7.3 (Galchouk-Kunita-Watanabe-decomposition)**.** *Any martingale $\hat{V}$ allows for a decomposition*

$$\hat{V} = \hat{V}_0 + \varphi \bullet \hat{S} + M, \tag{7.2}$$

where φ is some predictable, $\mathbb{R}^d$-valued process and M a martingale with $M_0 = 0$ which is orthogonal *to $\hat{S}$ in the sense that $M\hat{S}^i$ is a martingale for $i = 1, \dots, d$. The martingales $\varphi \bullet \hat{S}$ and M in this decomposition are unique. The integrand φ solves the vector equation*

$$\Delta\langle \hat{V}, \hat{S}\rangle = \varphi^\top \Delta\langle \hat{S}, \hat{S}\rangle,$$

or, more precisely,

$$\Delta\langle \hat{V}, \hat{S}^i\rangle_n = \sum_{j=1}^{d} \varphi_n^j \Delta\langle \hat{S}^j, \hat{S}^i\rangle_n \tag{7.3}$$

for $n = 1, \dots, N$ und $i = 1, \dots, d$.

Proof. The set $\{\hat{V}_N(\varphi) : \varphi \text{ self-financing strategy}\}$ is a subspace of the finite-dimensional space of random variables. Denote by $Y = \hat{V}_N(\varphi)$ the orthogonal projection in L^2 of $\hat{V}_N$ on that space. Moreover, let U be the martingale generated by Y, i.e. $U_n = E(Y|\mathscr{F}_n)$. In view of Lemma 3.16, using that $\hat{S}$ is a martingale, we have $U = \hat{V}(\varphi)$. Define $M := \hat{V} - U = \hat{V} - \hat{V}(\varphi)$. As a difference of martingales, M is a martingale as well.

Since Y is the orthogonal projection, $E((\hat{V}_N - Y)\hat{V}_N(\vartheta)) = 0$ holds for any self-financing strategy ϑ. In particular, $\hat{V}_0 - \hat{V}_0(\varphi) = E(\hat{V}_N - \hat{V}_N(\varphi)) = 0$. Indeed, this follows from considering the self-financing strategy ϑ with $\hat{V}_n(\vartheta) = 1$ for any n. This yields (7.2).

Let $n \in \{1, \dots, N\}$, $i \in \{1, \dots, d\}$ and $A \in \mathscr{F}_{n-1}$. If we define a predictable process

$$\vartheta^j_m = \begin{cases} 1_A & \text{if } j = i \text{ and } m \geq n, \\ 0 & \text{otherwise,} \end{cases}$$

we have

$$\begin{aligned} 0 &= E((\hat{V}_N - Y)(\vartheta \bullet \hat{S}^i_N)) \\ &= E(M_N 1_A(\hat{S}^i_N - \hat{S}^i_{n-1})) \\ &= E(M_N 1_A \hat{S}^i_N) - E(E(M_N 1_A \hat{S}^i_{n-1} | \mathscr{F}_{n-1})) \\ &= E(M_N 1_A \hat{S}^i_N) - E(E(M_N | \mathscr{F}_{n-1}) 1_A \hat{S}^i_{n-1}) \\ &= E(1_A M_N \hat{S}^i_N) - E(1_A M_{n-1} \hat{S}^i_{n-1}), \end{aligned}$$

i.e. $M\hat{S}^i$ is a martingale.

Since $[M, \hat{S}^i] = M\hat{S}^i - M_0\hat{S}^i_0 - M_- \bullet \hat{S}^i - \hat{S}^i_- \bullet M$ is a martingale, $\langle M, \hat{S} \rangle = 0$. Equation (7.2) yields $\langle \hat{V}, \hat{S}^i \rangle = \varphi \bullet \langle \hat{S}, \hat{S}^i \rangle + \langle M, \hat{S}^i \rangle = \varphi \bullet \langle \hat{S}, \hat{S}^i \rangle$, which implies (7.3).

For the uniqueness proof consider another decomposition $\hat{V} = \hat{V}_0 + \widetilde{\varphi} \bullet \hat{S} + \widetilde{M}$. Then $\langle M - \widetilde{M}, M - \widetilde{M} \rangle = \langle (\widetilde{\varphi} - \varphi) \bullet \hat{S}, M - \widetilde{M} \rangle = (\widetilde{\varphi} - \varphi) \bullet \langle \hat{S}, M - \widetilde{M} \rangle = 0$ and hence $M - \widetilde{M} = 0$ (e.g. by the remark following Definition 3.B.3). ☐

Corollary 7.4. *Denote by*

$$\hat{V} = \hat{V}_0 + (\varphi^1, \dots, \varphi^d) \bullet \hat{S} + M$$

the Galchouk-Kunita-Watanabe decomposition of the martingale $\hat{V}_n := E(\hat{X} | \mathscr{F}_n)$ *relative to* $\hat{S} = (\hat{S}^1, \dots, \hat{S}^d)$*. Moreover, let* φ *be the self-financing trading strategy corresponding to* $(\varphi^1, \dots, \varphi^d)$ *and initial discounted wealth* $\hat{V}_0$ *(cf. Lemma 4.7). Then* φ *is variance-optimal. It solves*

$$\Delta \langle \hat{V}, \hat{S} \rangle = \varphi^\top \Delta \langle \hat{S}, \hat{S} \rangle \tag{7.4}$$

in the sense of (7.3). The expected quadratic hedging error amounts to

$$\begin{aligned} \varepsilon^2 &:= E((\hat{V}_N(\varphi) - \hat{X})^2) \\ &= E\Big(\langle \hat{V} - \varphi \bullet \hat{S}, \hat{V} - \varphi \bullet \hat{S} \rangle_N\Big) &(7.5) \\ &= E\left(\langle \hat{V}, \hat{V} \rangle_N - \sum_{i,j=1}^d (\varphi^i \varphi^j) \bullet \langle \hat{S}^i, \hat{S}^j \rangle_N \right). &(7.6) \end{aligned}$$

Proof. In the proof of Theorem 7.3 it is shown that $\hat{V}_N(\varphi) = \hat{V}_0 + \varphi \bullet \hat{S}_N$ is the orthogonal projection of $\hat{X}$ on $\{\hat{V}_N(\vartheta) \in L^2 : \vartheta \text{ self-financing strategy}\}$. Hence φ is variance-optimal by the definition of the orthogonal projection.

Since M is a martingale starting at zero,

$$M^2 - \langle M, M \rangle = 2M_- \bullet M + [M, M] - \langle M, M \rangle$$

is a martingale starting at zero as well. Hence we obtain $E(M_N^2) = E(\langle M, M\rangle_N)$, which together with (7.2) and $\langle M + \hat{V}_0, M + \hat{V}_0\rangle = \langle M, M\rangle$ yields (7.5). (7.6) follows from

$$\begin{aligned}
\langle \hat{V}, \varphi \bullet \hat{S}\rangle_N &= \sum_{n=1}^{N} \Delta\langle \hat{V}, \varphi \bullet \hat{S}\rangle_n \\
&= \sum_{n=1}^{N}\sum_{i=1}^{d} \varphi_n^i \Delta\langle \hat{V}, \hat{S}^i\rangle_n \\
&= \sum_{n=1}^{N}\sum_{i,j=1}^{d} \varphi_n^i \varphi_n^j \Delta\langle \hat{S}^j, \hat{S}^i\rangle_n \\
&= \sum_{i,j=1}^{d} (\varphi^i\varphi^j) \bullet \langle \hat{S}^i, \hat{S}^j\rangle_N,
\end{aligned}$$

where we used (7.4) in the second but last equation. □

In the univariate case $d = 1$ the variance-optimal hedge is obtained from

$$\varphi_n^1 = \frac{\Delta\langle \hat{V}, \hat{S}^1\rangle_n}{\Delta\langle \hat{S}^1, \hat{S}^1\rangle_n}. \tag{7.7}$$

The numeraire component φ^0 is obtained from the self-financing condition as usual.

Remark. The optimisation problem in Definition 7.1 can be considered subject to the constraint that the initial wealth $\hat{V}_0(\vartheta)$ is fixed. Since $\hat{S}$ is a martingale and $E(\hat{X}) = \hat{V}_0$, we have

$$\begin{aligned}
&E\Big((\hat{V}_N(\vartheta) - \hat{X})^2\Big) \\
&\quad= E\Big((\hat{V}_0(\vartheta) - \hat{V}_0 + \vartheta \bullet \hat{S}_N - \hat{X} + \hat{V}_0)^2\Big) \\
&\quad= (\hat{V}_0(\vartheta) - \hat{V}_0)^2 + 2(\hat{V}_0(\vartheta) - \hat{V}_0)E(\vartheta \bullet \hat{S}_N) + E\Big((E(\hat{X}) + \vartheta \bullet \hat{S}_N - \hat{X})^2\Big) \\
&\quad= (\hat{V}_0(\vartheta) - E(\hat{X}))^2 + E\Big((E(\hat{X}) + \vartheta \bullet \hat{S}_N - \hat{X})^2\Big).
\end{aligned}$$

This implies that the strategy ϑ minimizing the hedging error $E((\hat{V}_N(\vartheta) - \hat{X})^2)$ depends on the fixed initial endowment $\hat{V}_0(\vartheta)$ only through its numeraire component.

For the variance-optimal strategy in Corollary 7.4 we have

$$\hat{V}_0(\varphi) - E(\hat{X}) = 0.$$

If $\hat{V}_0(\vartheta)$ is fixed instead, this non-optimal initial endowment raises the expected squared error by $(\hat{V}_0(\vartheta) - E(\hat{X}))^2$.

A concrete answer to the questions raised in Section 5.3 could be to charge the initial value of the variance-optimal hedge φ, raised by some risk premium which depends (e.g. linearly) on the expected squared hedging error. By investing in the variance-optimal hedge the bank can remove its risk at least partially. A weakness of a symmetric criterion as the quadratic one is that profits and losses are penalized in the same way. In the academic literature alternative concepts have been suggested as e.g. the utility indifference criterion discussed in Section 9.2. In practice heuristic approaches seem to dominate, which are not easily justified from a formal theoretical point of view.

Chapter 8

Elements of continuous-time finance

In this course we consider mainly discrete-time models with time set $\mathbb{N}$ resp. $\{0, \ldots, N\}$. Alternatively, mathematical finance can be based on continuous-time models with time set $\mathbb{R}_+$ or $[0, t]$, respectively. Many concepts and results of the preceding chapters can be carried over to this case. The mathematical theory, however, is too involved to be covered in this introductory course. Nevertheless, we want to give a short overview. In particular, we briefly address the Black-Scholes model as a main cornerstone of mathematical finance.

Continuous-time theory

The majority of definitions and results from Chapters 3 and 6 allows for a continuous-time counterpart. This includes filtrations, filtered probability spaces, stochastic processes, adaptedness, predictability, generated filtrations, stopping times, stopped processes, martingales, sub- and supermartingales, generated martingales, density processes, the generalized Bayes' fomula, the Doob decomposition, compensators, stochastic integrals, covariation and predictable covariation, integration by parts and other rules, Itô's formula, stochastic exponential, Yor's formula, Girsanov's theorem, martingale representation, the Snell envelope. Some notions require a more refined theory as e.g. predictability and stochastic integration. In this chapter we liberally use notions and results from Chapters 3 and 6 in continuous time without precise definitions, in particular related to stochastic integration. The reader should regard them simply as counterparts or limits of the familiar discrete objects.

Sums of independent, identically distributed random variables play an important role in discrete time. We applied them e.g. in the market models of Section 4. Their continous-time counterpart is called *Lévy process*.

Definition 8.1. A *Lévy process* (*process with stationary and independent increments*) is an adapted process $X = (X_t)_{t \geq 0}$ such that

1. $X_0 = 0$
2. $X_t - X_s$ is stochastically independent of $\mathscr{F}_s$ for $s \leq t$.
3. The law of $X_t - X_s$ depends only on $t - s$.

4. As a function of t, the process $(X_t)_{t\geq 0}$ is continuous from the right and has left-hand limits.

In some sense, Lévy processes play a similar role as linear functions in analysis. Firstly, they represent constant growth, where *constant* here is to be interpreted in a stochastic sense. Secondly, they can be characterized by relatively few parameters. Finally, a large class of more general processes resembles Lévy processes on a local scale, similarly as differentiable functions in analysis locally look like linear functions.

For simplicity we focus on univariate processes. The most important Lévy process next to deterministic linear functions is standard Brownian motion.

Definition 8.2. A Lévy process is called *standard Brownian motion* if X_1 is a standard normal random variable, i.e. with mean 0 and variance 1.

Remark. The law of standard Brownian motion is uniquely determined. We have that X_t is normally distributed with mean 0 and variance t. Moreover, the definition easily yields that standard Brownian motion is a martingale.

From now on we consider only *continuous* processes, i.e. $(X_t)_{t\geq 0}$ is supposed to be continuous in time t. The following theorem states the surprising and deep fact that any continuous process exhibiting constant growth in the sense of Definition 8.2 is a linear combination of a linear function and standard Brownian motion. This underlines the importance of Brownian motion for stochastic calculus. Moreover, we see that—as in deterministic calculus—objects of constant growth are determined by relatively few parameters, namely μ and σ compared to the slope μ in the deterministic case.

Theorem 8.3. *A real-valued Lévy process X is continuous if and only if it can be written as*

$$X_t = \mu t + \sigma W_t$$

for some standard Brownian motion W and constants $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$.

Definition 8.4. In view of the previous theorem we call continuous Lévy processes *Brownian motion with drift.*

Notation. By I we denote the *identity process* $I_t := t$, i.e. a simple linear function.

In calculus, differentiable functions can be viewed as approximately linear on a local scale. E.g. they can be written as $f(t) = \int_0^t \mu(s)ds = \mu \bullet I_t$, where the derivative $\mu(s)$ equals the slope of the linear function whose growth resembles that of f around s. Similarly, one can consider processes that resemble Brownian motion with parameters μ, σ on a local scale. They are called Itô processes.

Definition 8.5. Let W be a standard Brownian motion. Processes of the form

$$X = X_0 + \mu \bullet I + \sigma \bullet W \tag{8.1}$$

are called *Itô process* where μ, σ denote predictable processes μ, σ which are integrable with respect to I resp. W.

Remark. If (8.1) referred to discrete-time integrals, we could write it as

$$X_t = X_0 + \sum_{s=1}^{t} \mu_s \Delta I_s + \sum_{s=1}^{t} \sigma_s \Delta W_s$$

or

$$\Delta X_t = \mu_t \Delta I_t + \sigma_t \Delta W_t.$$

If we replace $\sum$ by the integral sign $\int$ (which is a stylized "S" for sum) and also Δ by d to indicate a limit, we obtain the two more common notations for (8.1) in the literature, namely

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

and

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

Covariation is defined in continuous time as well. For Itô processes $X = X_0 + \mu \bullet I + \sigma \bullet W$ and $\widetilde{X} = \widetilde{X}_0 + \widetilde{\mu} \bullet I + \widetilde{\sigma} \bullet W$ it is given by

$$[X, \widetilde{X}] = (\sigma\widetilde{\sigma}) \bullet I. \tag{8.2}$$

It is needed in Itô's formula, which plays a key role in stochastic calculus.

Theorem 8.6 (Itô's formula)**.** *If $X = X_0 + \mu \bullet I + \sigma \bullet W$ is an Itô process and $f : \mathbb{R} \to \mathbb{R}$ a twice continuously differentiable function, we have*

$$\begin{aligned} f(X_t) &= f(X_0) + f'(X) \bullet X_t + \frac{1}{2} f''(X) \bullet [X, X]_t \qquad (8.3) \\ &= f(X_0) + \left(f'(X)\mu + \frac{1}{2} f''(X)\sigma^2 \right) \bullet I_t + (f'(X)\sigma) \bullet W_t. \end{aligned}$$

More generally, if $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable, then

$$f(t, X_t) = f(0, X_0) + \left(\dot{f}(I, X) + f'(I, X)\mu + \frac{1}{2} f''(I, X)\sigma^2 \right) \bullet I_t + (f'(I, X)\sigma) \bullet W_t, \tag{8.4}$$

where dot and prime represent the derivatives with respect to the first and second argument of f, i.e. $\dot{f} = D_1 f$, $f' = D_2 f$ etc.

Interestingly, the approximation (3.6) holds exactly for Itô processes. On closer look, however, a difference can be detected. Instead of X_- we have written X in (8.3). But the previous value $X_{n-} = X_{n-1}$ is naturally replaced by a left-hand limit $X_{t-} := \lim_{s \uparrow t} X_s$ in continuous time. Since we consider only continuous processes here, we have $X = X_-$, i.e. the two formulas coincide when interpreted properly.

The *integration by parts* rule reads as

$$XY = X_0 Y_0 + X \bullet Y + Y \bullet X + [X, Y],$$

in continuous time. Note that this corresponds to (3.5), but not to (3.4).

As in discrete time, the *stochastic exponential* $Z = \mathscr{E}(X)$ of a process X is defined as unique solution to the equation $Z = 1 + Z_- \bullet X$. In the present case of continuous processes we can write again Z for Z_-.

Theorem 8.7 (Stochastic exponential)**.** *For Itô processes X we have*

$$\mathscr{E}(X)_t = \exp\left(X_t - X_0 - \frac{1}{2}[X,X]_t\right). \tag{8.5}$$

Proof. The Itô process $Y := X - X_0 - \frac{1}{2}[X,X]$ satisfies $[Y,Y] = [X,X]$ because the constant $X(0)$ and integrals relative to I do not change the covariation by (8.2). Using Itô's formula (8.3) we obtain for $Z := \exp(Y)$:

$$\begin{aligned}
Z &= \exp(Y) \\
&= \exp(Y_0) + e^Y \bullet Y + \frac{1}{2} e^Y \bullet [Y,Y] \\
&= 1 + e^Y \bullet X - \frac{1}{2} e^Y \bullet [X,X] + \frac{1}{2} e^Y \bullet [X,X] \\
&= 1 + Z \bullet X
\end{aligned}$$

as desired. □

Hence the approximation (3.7) is exact as well in continuous time.

Transferring the results from discrete-time mathematical finance to continuous time is less obvious than for the theory of stochastic processes. Continuous-time counterparts exist for price processes, trading strategies, value processes, self-financing strategies, discounted processes, arbitrage, the first fundamental theorem of asset pricing, equivalent martingale measures, dividend processes, market valuation, attainability, completeness, the second fundamental theorem of asset pricing, OTC valuation, upper and lower price, superhedging, American options and their relation to stopping problems and the Snell envelope, martingale modelling, variance-optimal hedging.The validity of the corresponding results, however, depends sensitively on the precise definitions of the set of admissible trading strategies etc. These definitions differ in the literature depending on the setup and on the problem under consideration. The continuous-time analogues have not yet always been stated in a clear and convincing fashion, at least compared to the situation in the theory of stochastic processes.

Black-Scholes model

Having finished the general overview we turn now to a concrete market model with two assets. This so-called Black-Scholes model goes back to Osborne and Samuelson and corresponds directly to the its discrete-time analogue in Section 4. Again we assume that relative price movements evolve homogeneously in time and independently of the past. In addition, we suppose that the price process is continuous. These assumptions already determine the model uniquely up to few parameters, in contrast to discrete time, where the law of daily relative time changes could be freely chosen. This follows from the surprising observation from Theorem 8.3, namely that any continuous process with stationary, independent increments is a Brownian motion with drift. In this sense, the process below can indeed be viewed as a *standard market model* even if the reservations of Section 4 apply here as well.

The *money market account* or *bond* is modelled as

$$S_t^0 = S_0^0 \exp(rt) = S_0^0 \mathscr{E}(rI)_t \tag{8.6}$$

with constant interest rate $r \in \mathbb{R}$. Moreover, we consider a *stock* or *foreign currency* whose price evolves in the following form:

$$S_t^1 = S_0^1 \exp(\mu t + \sigma W_t) = S_0^1 \mathscr{E}(\widetilde{\mu} I + \sigma W)_t, \tag{8.7}$$

with $\mu \in \mathbb{R}$, $\sigma > 0$, $\widetilde{\mu} = \mu + \frac{\sigma^2}{2}$ and some standard Brownian motion W. A process as in (8.7) is called *geometric Brownian motion.* The Itô-process representations of S^0 and S^1 are

$$\begin{aligned} S_t^0 &= S_0^0 + (S^0 r) \bullet I_t, \\ S_t^1 &= S_0^1 + (S^1 \widetilde{\mu}) \bullet I_t + (S^1 \sigma) \bullet W_t. \end{aligned}$$

One can show that this *Black-Scholes model* is complete. Relative to the corresponding equivalent martingale measure Q the process $\widetilde{W}_t := W_t + \frac{\widetilde{\mu}-r}{\sigma} t$ is a standard Brownian motion. Obviously we have

$$S_t^1 = S_0^1 \exp\left(rt - \frac{\sigma^2}{2} t + \sigma \widetilde{W}_t \right) = S_0^1 \mathscr{E}(rI + \sigma \widetilde{W})_t.$$

For the *return process* $X := \log S^1$ we have that the increments $X_t - X_{t-1}$, $t = 1, 2, \ldots$ are independent and normally distributed with mean $r - \frac{\sigma^2}{2}$ and variance σ^2. Restricted to integer times, the process X has the same law as in Lemma 5.6 for $\widetilde{\sigma} = \sigma$. This implies that we obtain the option prices from (5.6) when we are interested in the European call.

In the remainder of this section we take a different approach. We show that the European call $(S_T^1 - K)^+$ is replicable and we determine the corresponding perfect hedging strategy. Put differently, we look for a self-financing strategy φ satisfying $V_T(\varphi) = (S_T^1 - K)^+$. We make the natural ansatz that the value of this portfolio is a deterministic function of the stock price at any time, i.e.

$$V_t(\varphi) = f(t, S_t^1)$$

for some deterministic, twice continuously differentiable function $f : [0, T] \times \mathbb{R}_+ \to \mathbb{R}$. Itô's formula implies that

$$\begin{aligned} V_t(\varphi) &= f(t, S_t^1) \\ &= V_0(\varphi) + \left(\dot{f}(I, S^1) + f'(I, S^1) S^1 \widetilde{\mu} + \frac{1}{2} f''(I, S^1)(S^1 \sigma)^2 \right) \bullet I_t \\ &\quad + \left(f'(I, S^1) S^1 \sigma \right) \bullet W_t. \end{aligned} \tag{8.8}$$

On the other hand, we obtain from the self-financing condition

$$V_t(\varphi) = V_0(\varphi) + \varphi^0 \bullet S_t^0 + \varphi^1 \bullet S_t^1 \tag{8.9}$$

$$\quad = V_0(\varphi) + \left(\varphi^0 S^0 r + \varphi^1 S^1 \widetilde{\mu} \right) \bullet I_t + (\varphi^1 S^1 \sigma) \bullet W_t \tag{8.10}$$

Since the decomposition of an Itô process in drift and diffusion part, i.e. in integrals with respect to time and W is essentially unique, we expect from comparing (8.8) and (8.10) that

$$\varphi^1 = f'(I, S^1).$$

Since the value of a portfolio is

$$V(\varphi) = \varphi^0 S^0 + \varphi^1 S^1,$$

the previous equation implies

$$\varphi^0 S^0 = f(I, S^1) - f'(I, S^1) S^1.$$

Equating the drift terms in (8.8, 8.10) we get

$$\begin{aligned} & \dot{f}(I, S^1) + f'(I, S^1) S^1 \widetilde{\mu} + \frac{1}{2} f''(I, S^1)(S^1 \sigma)^2 \\ = \;& \varphi^0 S^0 r + \varphi^1 S^1 \widetilde{\mu} \\ = \;& f(I, S^1) r - f'(I, S^1) S^1 r + f'(I, S^1) S^1 \widetilde{\mu}. \end{aligned}$$

Since this is supposed to hold for arbitrary values of $I_t = t$ and S_t^1, we obtain the following partial differential equation (PDE) for f:

$$\dot{f}(t, x) = -\frac{1}{2} f''(t, x)(x\sigma)^2 - f'(t, x) x r + f(t, x) r$$

together with the terminal condition

$$f(T, x) = (x - K)^+,$$

which is obtained from $V_T(\varphi) = (S_T^1 - K)^+$.

Since the setup is related to Section 5.5, we can try and use the pricing formula from there in order to guess a solution. Indeed, one easily verifies that the above PDE is solved by the function

$$f(t, x) = x \Phi(d_1(t, x)) - K e^{-r(T-t)} \Phi(d_2(t, x))$$

with

$$\begin{aligned} d_1(t, x) &:= \frac{\log \frac{x}{K} + r(T-t) + \frac{\sigma^2}{2}(T-t)}{\sigma \sqrt{T-t}} \\ d_2(t, x) &:= \frac{\log \frac{x}{K} + r(T-t) - \frac{\sigma^2}{2}(T-t)}{\sigma \sqrt{T-t}}, \end{aligned}$$

which is known from (5.6). The corresponding portfolio equals

$$\varphi_t^1 = f'(t, S_t^1) = \Phi(d_1(t, S_t^1))$$

and

$$\varphi_t^0 = \frac{f(t, S_t^1) - f'(t, S_t^1) S_t^1}{S_t^0} = -K e^{-rT} \Phi(d_2(t, S_t^1)).$$

Reversing the above computations yields that this strategy really has the value

$$V_t(\varphi) = f(t, S_t^1) = S_t^1 \Phi(d_1(t, S_t^1)) - K e^{-r(T-t)} \Phi(d_2(t, S_t^1)) \tag{8.11}$$

and that its satisfies the self-financing condition (8.9). Since in addition $V_T(\varphi) = (S_T^1 - K)^+$, we have found a replicating strategy for the European call.

Note that the fair call price in this *Black Scholes formula* (8.11) depends only on the volatility parameter σ but non on the drift parameter μ, which is notoriously hard to estimate. It may appear as very surprising or even paradox that the very parameter is missing which crucially affects the probability of receiving a non-zero payoff at all.

Moreover, one may observe that the number of shares in the hedge is obtained by differentiating the pricing function relative to the stock price. This derivative is called *Delta* of the option.

Finally we mention that the Black-Scholes model along with pricing formula (8.11) can be obtained as a limit of Cox-Ross-Rubinstein models. From the point of view of the standard model in Section 5.5, however, the completeness of the Black-Scholes model may seem rather surprising. Indeed, no matter how fine we choose the mesh size of a grid discretizing proceses (8.6, 8.7) in time, we obtain the trivial price bounds for the European call. In the limiting continuous-time model, however, only one price is consistent with absence of arbitrage.

The Black-Scholes formula and the arbitrage reasoning are of utmost importance in practice. Indeed, tremendous sums are turned over in the global derivatives market. This was acknowleged by awarding the so-called Nobel price in economics to M. Scholes und R. Merton; F. Black had already passed away earlier. On the other hand, the theory has been critized to have contributed to the global financial crisis because it gives the impression of perfect control of financial risks which in fact one cannot handle. As can be seen from the previous sections, unique pricing formuals and perfect hedging strategies are the exception rather than the rule, even if we assume the absence of transaction costs, illiquidity, model risks etc. In any case, it is crucial to understand the limitations and weaknesses of mathematical models when decisions are based on them in market practice.

Appendix QF & MF

8.A From discrete to continuous time

In this chapter, we have tackled continuous time models by using concepts analogous to the notions in discrete time. Here, we take another point of view and try to approximate the Black-Scholes model (and other continuous time models) using discrete market models introduced before.

Our standard models in discrete time were of the form

$$A_i := S_i^1 = A_0 \prod_{j=1}^{i} Y_j, \quad i = 1, \dots, N$$

where $Y_1, Y_2, ..., Y_N$ are independent, identically distributed (i.i.d.) random variables, and we also assume that they are non-negative. Probability theory *prefers sums to products*. So we write

$$\begin{aligned} A_i &= A_0 \exp\left(\log\left(\prod_{j=1}^{i} Y_j \right) \right) = A_0 \exp\left(\sum_{j=1}^{i} \log(Y_j) \right) \\ &= A_0 \exp\left(\sum_{j=1}^{i} [\log(Y_j) - E(\log(Y_j))] + iE(\log(Y_1)) \right). \end{aligned}$$

(Note that all moments of $\log(Y_1)$ exist in finite probability spaces.) Now, write $X_j := \log(Y_j) - E(\log(Y_j))$ for all j and

$$S_0 = 0, S_1 = X_1, S_2 = X_1 + X_2, \dots, S_i = X_1 + \dots + X_i, \dots,$$

so that

$$A_i = A_0 \exp\left(S_i + iE(\log(Y_1))\right).$$

Process of sums

Let us more generally look at a sequence of sums

$$S_0 = 0, S_1 = X_1, S_2 = X_1 + X_2, \dots, S_i = X_1 + \dots + X_i, \dots$$

where $X_1, X_2, \dots$ are i.i.d. with $E(X_i) = 0, \mathrm{Var}(X_i) = \sigma^2$, for simplicity take $\sigma^2 = 1$.

One of the most fundamental results from probability theory is the *central limit theorem* (CLT), which states that $\frac{1}{\sqrt{n}}S_n$ is approximately $N(0,1)$-distributed for large n. More precisely,

$$P\left(\frac{1}{\sqrt{n}}S_n \le t\right) \overset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}\,dx \text{ for all } t,$$

written as

$$\frac{1}{\sqrt{n}}S_n \overset{n\to\infty}{\longrightarrow} Y \sim N(0,1) \text{ in distribution.}$$

Now take $t \in [0,T]$, for simplicity $T=1$.
For

$$t = 0 = \frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n} = 1$$

write

$$S_t^{(n)} = S_j, \quad t = \frac{j}{n}.$$

In between, we can either use linear interpolation or use horizontal lines to get a process

$$S_t^{(n)}, \quad t \in [0,1].$$

The horizontal line process is simply

$$S_t^{(n)} = S_{[nt]}, \quad t \in [0,1]$$

where $[nt]$ is the integer part of nt

$$[nt] = j \text{ for } \frac{j}{n} \le t < \frac{j+1}{n}$$

For both cases we have

$$\frac{1}{\sqrt{n}}S_t^{(n)} \overset{n\to\infty}{\longrightarrow} N(0,t)$$

e.g. for the horizontal line case

$$\frac{1}{\sqrt{n}}S_t^{(n)} = \frac{1}{\sqrt{n}}S_{[nt]} = \underbrace{\frac{1}{\sqrt{[nt]}}S_{[nt]}}_{\to N(0,1)} \cdot \underbrace{\frac{\sqrt{[nt]}}{\sqrt{n}}}_{\to\sqrt{t}} \to N(0,t)$$

So for each t

$$\frac{1}{\sqrt{n}}S_t^{(n)} \to W_t \text{ in distribution,}$$

where W denotes standard Brownian motion. In fact it can be shown that also as a stochastic process

$$\left(\frac{1}{\sqrt{n}}S_t^{(n)}\right)_{t\in[0,T]} \to (W_t)_{t\in[0,T]}$$

in a strong sense. This result is known as *Donsker's theorem*, see the lecture on stochastic processes. So we can take normalized by $\frac{1}{\sqrt{n}}$ sum processes to approximate the standard Brownian motion.

Approximating the Black-Scholes model by a binomial tree model

$$
\begin{array}{ccccccc}
 & & & & S_0u^2 & & \\
 & & & \overset{p}{\nearrow} & & & \cdots \\
 & & S_0u & & & & \\
 & \overset{p}{\nearrow} & & \overset{1-p}{\searrow} & & & \cdots \\
S_0 & & & & S_0ud & & \\
 & \overset{1-p}{\searrow} & & \overset{p}{\nearrow} & & & \cdots \\
 & & S_0d & & & & \\
 & & & \overset{1-p}{\searrow} & & & \cdots \\
 & & & & S_0d^2 & &
\end{array}
$$

$$S_0 \prod_{j=1}^{i} Y_j, \quad Y_j = \begin{cases} u & \text{prob. } p \\ d & \text{prob. } 1-p \end{cases}$$

We want to use such a model to approximate a Black-Scholes model on $[0, T]$. For n, which denotes the accuracy of the approximation, we look at the time points

$$0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{[nT]}{n}$$

and a stock price

$$A^{(n)}_{\frac{k}{n}} = A_0 \prod_{i=1}^{k} Y_i^{(n)}$$

where $Y_1^{(n)}, Y_2^{(n)}, \cdots$ are independent with

$$P(Y_i^{(n)} = u_n) = 1 - P(Y_i^{(n)} = d_n) = p_n.$$

We want to approximate

$$A_t = \exp(\sigma W_t + at), \quad t \in [0, T].$$

Let $a_n = E\left(\log(Y_i^{(n)})\right)$, $\sigma_n^2 = \operatorname{Var}\left(\log(Y_i^{(n)})\right)$. Then for $k = 0, 1, \cdots, n$

$$
\begin{aligned}
A^{(n)}_{\frac{k}{n}} &= A_0 \exp\left(\sum_{i=1}^{k} \log\left(Y_i^{(n)}\right)\right) \\
&= A_0 \exp\left(\left[\underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{k} \overbrace{\left(\frac{\log\left(Y_i^{(n)}\right) - a_n}{\sigma_n}\right)}^{X_i \text{ with } EX_i=0, \operatorname{Var}X_i=1}}_{\approx W_{\frac{k}{n}} \text{ in distribution with general CLT}}\right] \cdot \sigma_n \sqrt{n} + \frac{k}{n} a_n \cdot n\right) \quad \approx A_0 e^{\sigma W_{\frac{k}{n}} + a\frac{k}{n}},
\end{aligned}
$$

if

$$
\begin{aligned}
a_n n &= a \, (or \approx a), \\
\sigma_n \sqrt{n} &= \sigma \, (or \approx \sigma)
\end{aligned}
$$

Now insert for a_n, σ_n to obtain the equations

$$\begin{aligned} na_n &= n(p_n \log(u_n) + (1 - p_n)\log(d_n)) = a\, [\approx a] \\ n\sigma_n^2 &= n(p_n \log(u_n)^2 + (1 - p_n)\log(d_n)^2 - a_n^2) = \sigma^2 \left[\approx \sigma^2\right] \end{aligned}$$

This is easily solved for p_n, u_n, d_n; we may put the additional constraint $d_n = \frac{1}{u_n}$. An approximate solution is given by

$$u_n = e^{\frac{\sigma}{\sqrt{n}}}, \quad d_n = \frac{1}{u_n}, \quad p_n = \frac{1}{2}\left(1 + \frac{a}{\sigma\sqrt{n}}\right).$$

Appendix MF

8.B Change of measure for Brownian motion

In our approach to the Black-Scholes model, the question remained open whether an equivalent martingale measure exists. For such kind of questions, theorems of *Girsanov* type are used. Here, we present an elementary version, which is enough for our needs.

Theorem 8.B.1. *Let W be a standard Brownian motion. Let $a \in \mathbb{R}$ and $T > 0$. Define*

$$L_T = \exp\left(aW_T - \frac{a^2}{2}T\right)$$

and

$$Q(A) = E(L_T 1_A).$$

Then, Q is a probability measure equivalent to P such that $(W_t)_{t\in[0,T]}$ is a Brownian motion with volatility 1 and drift a w.r.t. Q.

Proof. Recall first that $\exp(aW_t - a^2/2t)_t$ is a stochastic exponential with respect to a continuous martingale and therefore a martingale starting in 1. As a consequence, $P(L_T > 0) = 1$ and $EL_T = 1$, so Q is equivalent to P. Now let Y be another Brownian motion with volatility 1 and drift a. We need to show that for all $0 < t_1 < \cdots < t_n = T$ and for all bounded and measurable $g : \mathbb{R}^n \to \mathbb{R}$ it holds that

$$E_Q g(W_{t_1}, \ldots, W_{t_n}) = Eg(Y_{t_1}, \ldots, Y_{t_n}).$$

To this end, we directly use the normal densities of the Brownian motion:
Let $t_0 = 0 < t_1 < \ldots < t_n = T$ and $x_0 = 0, x_1, \ldots, x_n \in \mathbb{R}$. For the densities of the n-dimensional distributions it holds that

$$\begin{aligned}
& f_{(Y_{t_1},\ldots,Y_{t_n})}(x_1, \ldots, x_n) \\
= \; & f_{(Y_{t_1}-Y_{t_0},\ldots,Y_{t_n}-Y_{t_{n-1}})}(x_1 - x_0, \ldots, x_n - x_{n-1}) \\
= \; & \prod_{i=1}^{n} \exp\left(-\frac{(x_i - x_{i-1} - a(t_i - t_{i-1}))^2}{2(t_i - t_{i-1})}\right) \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \\
= \; & \prod_{i=1}^{n} \exp\left(-\frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}\right) \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \\
& \times \prod_{i=1}^{n} \exp\left(a(x_i - x_{i-1}) - \frac{a^2}{2}(t_i - t_{i-1})\right) \\
= \; & f_{(W_{t_1},\ldots,W_{t_n})}(x_1, \ldots, x_n) \exp\left(ax_n - \frac{a^2}{2}T\right).
\end{aligned}$$

Hence, for all bounded and measurable $g : \mathbb{R}^n \to \mathbb{R}$

$$\begin{aligned}
E_Q g(W_{t_1}, \ldots, W_{t_n}) &= E L_T g(W_{t_1}, \ldots, W_{t_n}) \\
&= \int g(x_1, \ldots, x_n) \exp\left(a x_n - \frac{a^2}{2} T\right) \\
&\quad \times f_{(W_{t_1}, \ldots, W_{t_n})}(x_1, \ldots, x_n) dx_1 \ldots dx_n \\
&= \int g(x_1, \ldots, x_n) f_{(Y_{t_1}, \ldots, Y_{t_n})}(x_1, \ldots, x_n) dx_1 \ldots dx_n \\
&= E g(Y_{t_1}, \ldots, Y_{t_n}),
\end{aligned}$$

which proves the claim. □

Now, in the Black-Scholes model the discounted asset price process is given by

$$\hat{S}_t^1 = S_0^1 \exp((\tilde{\mu} - r - \sigma^2/2)t + \sigma W_t).$$

We now want to change the drift of W such that the process becomes an exponential martingale of the Brownian motion. Using the notations from the previous theorem, we obtain

$$\hat{S}_t^1 = S_0^1 \exp((\tilde{\mu} - r + \sigma a - \sigma^2/2)t + \sigma \tilde{W}_t),$$

where $\tilde{W}_t := W_t - at$, t, is a SBM under Q. Now, we set $a = \frac{-1}{\sigma}(\tilde{\mu} - r)$ to obtain the result. In this case,

$$S_t^1 = S_0^1 \exp((r - \sigma^2/2)t + \sigma \tilde{W}_t),$$

and

$$\tilde{W}_t = W_t + \frac{\tilde{\mu} - r}{\sigma} t, \; t \in [0, T],$$

is a SBM under Q. Hence, from the discrete time results it seems reasonable to use expectations under Q for pricing, e.g. for the European call $X = (S_T^1 - K)^+$ we get for $t = 0$ the price

$$E_Q(e^{-rT}(S_T^1 - K)^+),$$

where S_T^1 is a random variable with a log-normal distribution. Calculating this expectation yields the Black-Scholes price also obtained in (8.11) for $t = 0$ (and similar for general t).

Chapter 9

Portfolio optimization

9.1 Maximizing expected utility of terminal wealth

Without any background one may think that financial mathematics is primarily concerned with maximizimg the investor's wealth. As we have seen this is not the case. Nevertheless, the question how to choose one's portfolio optimally has been studied thoroughly. In this chapter we consider maximization of expected utility of terminal wealth.

As in the previous chapters we work on a finite filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n=0,\dots,N}, P)$ with terminal date $N \in \mathbb{N}$, trivial σ-field $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and such that $\mathscr{F} = \mathscr{P}(\Omega)$ and $P(\{\omega\}) > 0$ for all ω. As usual we consider an asset price process $S = (S^0, \dots, S^d)$ that does not allow for arbitrage and such that the numeraire S^0 is positive. Later we will be more specific about the price process.

We consider an investor who wants to maximize her terminal wealth $V_N(\varphi)$ by investing her initial capital V_0 optimally. However, $V_N(\varphi)$ is random and hence any strategy's success or failure depends on whether and which asset prices rise or fall, which we cannot predict. Hence one needs to think about what really is supposed to be optimized here.

Naïvely, one may consider maximizing the expectation $E(V_N(\varphi))$. This criterion, however, does not appear to be attractive on closer inspection. It neglects that most investors are risk averse. Focusing on the expectation makes a highly speculative investment seem superior to a riskless savings account, even if the average return is only slightly higher.

As an alternative we consider a classical criterion that reflects both the desire to achieve a high average return as well as the aversion to possibly threatening losses. The idea is to replace mean terminal wealth by the mean $E(u(\hat{V}_N(\varphi)))$ of its utility. This utility is expressed in terms of a *utility function* $u : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$, which maps an amount of money to the degree of happiness it creates. For obvious reasons one assumes this function to be increasing. As a further important property one demands concavity. Intuitively, this means that an additional Euro creates less additional utility than a lost Euro kills. Put differently, a lottery paying one Euro on average is perceived as less attractive than a safe Euro. This property of weighting losses more than gains makes the criterion consider both competing goals. The choice of the utility function allows to incorporate one's personal preferences and risk aversion. Popular utility functions are the power and logarithmic utility considered below.

Definition 9.1. Any increasing, strictly concave function $u : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ is called **utility function**. A self-financing strategy φ with discounted initial wealth $\hat{V}_0$ is called **optimal** in the sense of expected utility from terminal wealth if φ maximizes $E(u(\hat{V}_N(\vartheta)))$ as a function of such strategies ϑ.

The most popular utility functions are of power type $u(x) = \frac{x^{1-p}}{1-p}$ for $x > 0$ and fixed *risk aversion parameter* $p \in (0, \infty)$ with $p \neq 1$, the logarithm $u(x) = \log(x)$ for $x > 0$, and exponential functions $u(x) = 1 - \exp(-\lambda x)$ with fixed *risk aversion parameter* $\lambda > 0$. The denominator in the power case only warrants that u is increasing for any choice of p.

Exploiting convextity, it is easy to show that optimal strategies are essentially unique. More precisely, their value process is uniquely determined. The corresponding trading strategy itself may be ambiguous if there are e.g. two assets with identical price process in the market.

Lemma 9.2. *All optimal strategies φ corresponding the same utility function u and the same initial capital have the same (discounted) wealth process $\hat{V}(\varphi)$.*

Proof. Suppose that there are optimal strategies $\varphi, \widetilde{\varphi}$ with different terminal wealth, i.e. $\hat{V}_N(\varphi) \neq \hat{V}_N(\widetilde{\varphi})$. For the average portfolio $\psi := \frac{1}{2}(\varphi + \widetilde{\varphi})$ we have

$$\begin{aligned} E(u(\hat{V}_N(\psi))) &= E\Big(u(\tfrac{1}{2}\hat{V}_N(\varphi) + \tfrac{1}{2}\hat{V}_N(\widetilde{\varphi}))\Big) \\ &> E\Big(\tfrac{1}{2}u(\hat{V}_N(\varphi)) + \tfrac{1}{2}u(\hat{V}_N(\widetilde{\varphi}))\Big) \\ &= E(u(\hat{V}_N(\varphi))), \end{aligned}$$

which contradicts the optimality of φ.

Consequently, the discounted wealth of all optimal strategies φ coincides. By Lemma 4.9 this implies that the whole value process $\hat{V}(\varphi)$ is the same for all optimal strategies. $\square$

In general, it is not easy to compute optimal portfolios because the set of trading strategies is of very high dimension. A classical approach is called *dynamic programming*, where the optimal strategy is obtained recursively, similarly as we determined the price of an American put in the Cox-Ross-Rubinstein model. Here we apply instead martingale methods to obtain optimal portfolios in concrete models. They are based on the following sufficient condition for optimality. It does not provide the solution but it helps to prove that a given candidate strategy is in fact optimal.

Theorem 9.3. *Suppose that the utility function u is differentiable (on the set where it is finite). Moreover, let φ be a self-financing strategy with discounted initial capital $\hat{V}_0$. Define a probability measure $Q \sim P$ by its density*

$$\frac{dQ}{dP} = \frac{u'(\hat{V}_N(\varphi))}{E(u'(\hat{V}_N(\varphi)))}. \tag{9.1}$$

If Q happens to be an equivalent martingale measure, then φ is the optimal strategy for u.

Proof. Suppose that ψ is another self-financing strategy with the same initial value. By concavity of u we have

$$\begin{aligned} u(\hat{V}_N(\psi)) - u(\hat{V}_N(\varphi)) &\leq u'(\hat{V}_N(\varphi))(\hat{V}_N(\psi) - \hat{V}_N(\varphi)) \\ &= E(u'(\hat{V}_N(\varphi)))\tfrac{dQ}{dP}((\psi - \varphi) \bullet \hat{S}_N). \end{aligned} \tag{9.2}$$

Since $(\psi - \varphi) \bullet \hat{S}$ is a Q-martingale by Lemma 3.16, we have $E_Q((\psi - \varphi) \bullet \hat{S}_N) = 0$ and hence $E(u(\hat{V}_N(\psi))) \leq E(u(\hat{V}_N(\varphi)))$ by (9.2). □

The previous theorem underlines once more the key role played by equivalent martingale measures in Mathematical Finance. Their existence and uniqueness characterize absence of arbitrage resp. completeness. And now they occur again in the context of portfolio optimization. In the present case of finite probability spaces, the above sufficient criterion is also necessary for most utility functions. Moreover, the above sufficient condition holds in infinite probability spaces as well. The above EMM Q can be shown to solve some dual minimization problem, i.e. it is the martingale measure which is closest to P with respect to some distance that depends on u. However, these results are beyond of the scope of the present introduction.

In arbitrary market models it is not easy to determine the optimal strategy unless one considers the logarithmic utility function. For the latter, the problem can be solved explicitly in general. Here, we focus on a time-homogeneous market model, which generalizes the two asset model of Section 4 to multiple risky assets. The money market account or bond is assumed to be of the form (4.2) as before, i.e. $S^0_n = S^0_0 \exp(rn) = S^0_0(1 + \tilde{r})^n$. Parallel to (4.3), the price processes of d stocks is supposed to be given by

$$S^i_n = S^i_0 \exp(X^i_n) = S^i_0 \prod_{m=1}^{n} (1 + \Delta\widetilde{X}^i_m) = S^i_0 \mathscr{E}(\widetilde{X}^i)_n$$

for $i = 1, \dots, d$ and $n = 0, \dots, N$. Here, X (resp. $\widetilde{X}$) denotes a $\mathbb{R}^d$-valued adapted process with $X_0 = \widetilde{X}_0 = 0$, whose increments ΔX_n (resp. $\widetilde{\Delta X}_n$), $n = 1, \dots, N$, are independent, identically distributed random variables. X and $\widetilde{X}$ are related to each other via $\widetilde{\Delta X}^i_n = e^{\Delta X^i_n} - 1$, which parallels (4.4). For the discounted price process we have

$$\hat{S}^i_n = \hat{S}^i_0 \prod_{m=1}^{n} (1 + \Delta\hat{X}^i_m) = \hat{S}^i_0 \mathscr{E}(\hat{X}^i)_n$$

with $\hat{X}^i_0 = 0$ and $\Delta\hat{X}^i_n = \frac{1+\Delta\widetilde{X}^i_n}{1+\tilde{r}} - 1$. Note that the relative price changes of *different assets* may be stochastically dependent, but not relative time changes in *different periods*.

We want to apply the above martingale criterion in order to determine optimal strategies for utility functions of power and logarithmic type. The difficult part is to guess a reasonable candidate portfolio whose optimality can then be proved by Theorem 9.3. To this end we make a parametric ansatz, hoping that the true solution is of the assumed form. The parameters are determined in second step such that the proof works.

In the above homogeneous market the following ansatz turns out to be successful. We consider portfolios investing a constant fraction of wealth in each of the risky assets. The proportions are denoted by the vector γ below. The actual number of shares varies randomly over time because both the investor's wealth and the asset prices move. The discounted wealth and the numeraire part φ^0 are determined by self financability.

Another obstacle must be overcome to apply the above criterion. Knowing the density dQ/dP generally does not suffice to decide whether Q is a martingale measure. The whole density process Z is needed for that purpose. This problem is solved by once again making an ansatz. We assume that proportionality of Z_n and $u'(\hat{V}_n(\varphi))$ holds not only at N, but also in earlier periods with a possibly changing factor.

Theorem 9.4. *Consider the utility function u with*

$$u(x) := \begin{cases} \frac{x^{1-p}}{1-p} & \text{for } x > 0 \\ -\infty & \text{for } x \leq 0 \end{cases}$$

for fixed $p \in (0, \infty)$. In the case $p = 1$ the above undefined expression is to be replaced by $u(x) = \log(x)$ for $x > 0$. Define a vector $\gamma \in \mathbb{R}^d$ of portfolio weights such that $\gamma^\top \Delta \hat{X}_1 > -1$ and

$$E\left(\frac{\Delta \hat{X}_1}{(1 + \gamma^\top \Delta \hat{X}_1)^p}\right) = 0. \tag{9.3}$$

Moreover, set

$$\hat{V}_n := \hat{V}_0 \mathscr{E}(\gamma^\top \hat{X})_n,$$

$$\varphi_n^i := \frac{\gamma^i}{\hat{S}_{n-1}^i} \hat{V}_{n-1} \quad \text{for } i = 1, \dots, d,$$

$$\varphi_n^0 := \hat{V}_{n-1} - (\varphi^1, \dots, \varphi^d)_n^\top (\hat{S}^1, \dots, \hat{S}^d)_{n-1}.$$

Then φ is the optimal strategy for u and discounted initial wealth $\hat{V}_0$. Its discounted wealth process equals $\hat{V}$.

Proof. Since

$$\begin{aligned} \hat{V}_0 + \varphi \bullet \hat{S} &= \hat{V}_0 \left(1 + \textstyle\sum_{i=1}^d (\mathscr{E}(\gamma^\top \hat{X})_- \frac{\gamma^i}{\hat{S}_-^i} \hat{S}_-^i) \bullet \hat{X}^i\right) \\ &= \hat{V}_0 \left(1 + \mathscr{E}(\gamma^\top \hat{X})_- \bullet (\gamma^\top \hat{X})\right) \\ &= \hat{V}_0 \mathscr{E}(\gamma^\top \hat{X}) = \hat{V}, \end{aligned}$$

we have that $\hat{V}$ is the discounted wealth process of the self-financing strategy φ corresponding to $(\varphi^1, \dots, \varphi^d)$ and discounted initial wealth $\hat{V}_0$. Moreover, the numeraire component φ^0 is given by the above expression.

Set

$$\alpha := (E((1 + \gamma^\top \Delta \hat{X}_1)^{-p}))^{1/p}$$

and

$$Z_n := (\alpha^n \mathscr{E}(\gamma^\top \hat{X})_n)^{-p}$$

for $n = 0, \dots, N$. From

$$Z_n = Z_{n-1} \alpha^{-p} \left(\frac{\mathscr{E}(\gamma^\top \hat{X})_n}{\mathscr{E}(\gamma^\top \hat{X})_{n-1}}\right)^{-p} = Z_{n-1} \alpha^{-p} (1 + \gamma^\top \Delta \hat{X}_n)^{-p}$$

we conclude $Z = \mathscr{E}(M)$ with $M_n := \sum_{m=1}^n((\alpha(1+\gamma^\top\Delta\hat{X}_m))^{-p}-1)$. Since

$$E(\alpha^{-p}(1+\gamma^\top\Delta\hat{X}_n)^{-p}|\mathscr{F}_{n-1}) = \frac{E((1+\gamma^\top\Delta\hat{X}_n)^{-p}|\mathscr{F}_{n-1})}{E((1+\gamma^\top\Delta\hat{X}_n)^{-p})} = 1,$$

M and hence also Z are martingales, in particular with $E(Z_N) = 1$. Therefore Z is the density process of a probability measure $Q \sim P$ with density Z_N. According to Bayes' formula (Lemma 3.B.1) and (9.3) we have

$$\begin{aligned} E_Q(\Delta\hat{X}_n|\mathscr{F}_{n-1}) &= E(\Delta\hat{X}_n \tfrac{Z_n}{Z_{n-1}}|\mathscr{F}_{n-1}) \\ &= E(\Delta\hat{X}_n\alpha^{-p}(1+\gamma^\top\Delta\hat{X}_n)^{-p}|\mathscr{F}_{n-1}) \\ &= \alpha^{-p}E(\Delta\hat{X}_n(1+\gamma^\top\Delta\hat{X}_n)^{-p}) = 0 \end{aligned}$$

for $n = 1,\dots,N$, i.e. $\hat{X}$ is a Q-martingale. Consequently $\hat{S}^i = \hat{S}^i_0\mathscr{E}(\hat{X}^i)$ is a Q-martingale for $n = 1,\dots,d$. Theorem 9.3 yields that φ is optimal for u. □

Equation (9.3) is a system of d equations with d unknowns $\gamma^1,\dots,\gamma^d$. In most concrete models it allows for a unique solution.

Note that the logarithm corresponds to $p = 1$ in the above proof because the latter uses only the derivative of u. How do the solutions depend on the risk aversion parameter p? This is easiest to explain by considering a linear approximation. If the relative price changes $\Delta\hat{X}_n$ are small, we can approximate the denominator in (9.3) by $(1+\gamma^\top\Delta\hat{X}_1)^{-p} \approx 1 - p\gamma^\top\Delta\hat{X}_1$. Then (9.3) turns into the quadratic equation $E(\Delta\hat{X}_1) \approx pE(\Delta\hat{X}_1\Delta\hat{X}_1^\top)\gamma$. Consequently, we have

$$\gamma \approx \frac{1}{p}\left(E(\Delta\hat{X}_1\Delta\hat{X}_1^\top)\right)^{-1}E(\Delta\hat{X}_1), \tag{9.4}$$

if the matrix

$$E(\Delta\hat{X}_1\Delta\hat{X}_1^\top) = \mathrm{Cov}(\Delta\hat{X}_1) + E(\Delta\hat{X}_1)(E(\Delta\hat{X}_1))^\top \approx \mathrm{Cov}(\Delta\hat{X}_1)$$

is invertible.

(9.4) can be interpreted as follows. Any investor with power or logarithmic utility tries to keep the fraction of wealth held in any of the assets constant. The proportion invested in the numeriare S^0 on the one hand and in the risky assets $S^1,\dots,S^d$ on the other hand depends on p. However, the relative contribution of $S^1,\dots,S^d$ is approximately the same for all p if we assume price changes in single periods to be small. Put differently, the risk aversion p only determines how wealth is split between risky and riskless assets. In the case of a single risky stock, we see that the optimal investment in this stock is proportional to the excess return $E(\Delta\hat{X}_1)$ (compared to the money market account) and inversely proportional to the variance $\mathrm{Var}(\Delta\hat{X}_1)$. Since the variance can be interpreted as riskyness of the stock, (9.4) seems quite plausible.

The case of exponential utility $u(x) = 1 - \exp(-\lambda x)$ can be treated similarly. However, in this case the optimal strategy does not assign a constant *fraction of wealth* to any of the assets. Instead, a fixed *amount of money* is invested in any of the risky assets. This amount does not depend on initial wealth, which does not seem very reasonable from an economic perspective.

In the case of logarithmic utility the ansatz of Theorem 9.4 works also for arbitrary price processes whose relative increments are not necessarily independent. In this case the optimal fractions γ_n differ from period to period. Similarly to (9.3), they are given as the solution to the equation

$$E\left(\frac{\Delta\hat{X}_n}{1+\gamma_n^\top \Delta\hat{X}_n}\middle|\mathscr{F}_{n-1}\right)=0. \tag{9.5}$$

Let us have a look at the general structure of the optimal portfolio in Theorem 9.4. Surprisingly, it does not depend on the time horizon N. This contradicts the common advice to invest a larger fraction of wealth in risky assets if the time horizon is large. This advice is typically justified by arguing that random fluctuations average out in the long run, while the higher average return persists.

The optimal portfolio for logarithmic utility has another interesting property. As noted in the beginning of this chapter we cannot expect there to be a portfolio which certainly yields a higher return than any other. Surprisingly, this changes if we consider the return in the long run. To this end, note that the wealth of a bank account with fixed continuously compounded interest rate r grows expontially according to $S_N^0=S_0^0e^{rN}$. The *rate of return* r is obtained from S^0 via

$$r=\frac{1}{N}\log(S_N^0/S_0^0). \tag{9.6}$$

Now, let us consider an arbitrary portfolio φ with value process $V(\varphi)$. Inspired by (9.6),

$$\lim_{N\to\infty}\tfrac{1}{N}\log(V_N(\varphi)/V_0)$$

can be interpreted as the *long run growth rate of wealth* of portfolio φ. Somewhat surprisingly, there typically exists a portfolio which maximizes this long run growth rate of wealth with probability one. This happens to be the strategy which maximizes expected utility of terminal wealth for logarithmic utility and which—as we observed before—does not depend on the time horizon. Put differently, in the very long run this *growth optimal portfolio* outperforms any other strategy with probability one! However, it may take a rather long time to do so. The proof of this result is beyond the scope of this course.

At the end of this chapter we want to address a problem that concerns the applicability of the results in practice. In order to determine optimal portfolios, we need a reliable estimate of the common law of asset returns.

We consider the particularly simple case of a single stock which has been observed for ten years. Let us assume for simplicity that daily logarithmic returns

$$\log(S_n^1/S_{n-1}^1)=\log(1+\Delta\widetilde{X}^1)=r+\log(1+\Delta\hat{X}^1)$$

are normally distributed with mean $r+\mu/250$ and variance $\sigma^2/250$. The factor 250 stands for converting yearly into daily parameters (1 year $\approx$ 250 trading days). By (9.4) and $E(\Delta\hat{X}^1)\approx E(\log(1+\Delta\hat{X}^1))$ the parameter μ enters the optimal portfolio linearly.

Suppose that the volatility is known to be 25%, i.e. $\sigma=0.25$. If the average logarithmic return in the previous 10 years amounts to 5% above the riskless interest rate, the standard estimate for μ is 0.05. To be more precise, standard statistical theory yields a 95%-confidence interval $[-0.10, 0.20]$ for the unkown drift parameter μ. This precision does

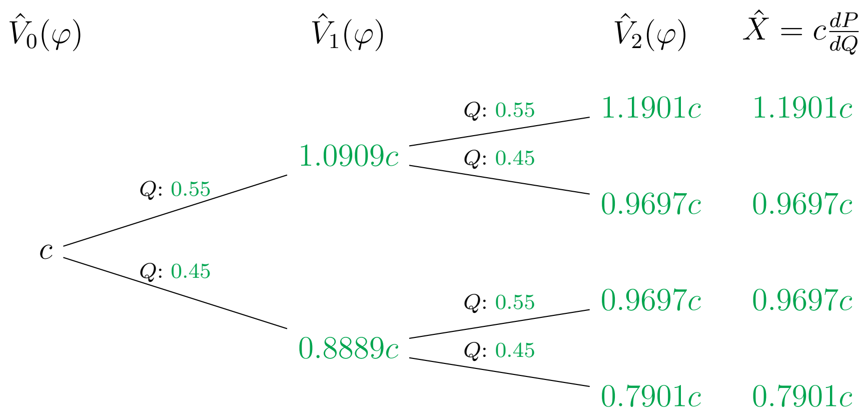

Figure 9.1.1: The discounted fair price $\hat{V}(\varphi)$ of the claim with discounted payoff $\hat{X} = cu'(\frac{dQ}{dP}) = c\frac{dP}{dQ}$ for the market in Figures 4.1, 4.6 and utility function $u(x) = \log(x)$

not increase if we dispose of high frequency or even tick-by-tick data. Only if longer periods have been observed, we obtain more reliable estimates. On the other hand, it is less obvious whether the assumption of constant parameters can be trusted if the time horizon spans several decades.

Consequently, with ten years of data it is hard even to make a reliable statement whether the stock's excess rate of return compared to the riskless asset is positive at all. Since φ is approximately linear in μ, the same is true for the fraction of wealth invested in stock. And this holds even if we make the simpifying assumption of constant parameters in a Gaussian framework.

Nevertheless the above statements provide interesting insights, even if real-world parameters are hard to come by. Indeed, we have seen that it makes sense to hold assets in constant fractions of wealth that—at least according to the above criteria—do not depend on the time horizon. In particular, it seems preferable to distribute wealth over all the available assets rather than to invest only in the stock with highest rate of return. Finally, for the long-term investor thinking in centuries rather than years (foundations, the church, ...) it may make sense to invest in the growth optimal portfolio.

9.2 Utility-based pricing and hedging

We consider the situation is Sections 5.3 und 7.1, where a client asks the bank for a derivative with discounted payoff $\hat{X} = X/S_N^0$. If the bank is offered the premium π, it can choose between two alternatives: either to enter the trade or to decline it. In this section we discuss the utility indifference principle, which provides a threshold premium dividing the favourable from the unfavourable deals. Moreover, it indicates how to sensibly invest the premium.

We assume that the bank's objective is to maximize its expected utility based on some utility function $u : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$. Its initial discounted wealth is denoted as $\hat{V}_0$. If no

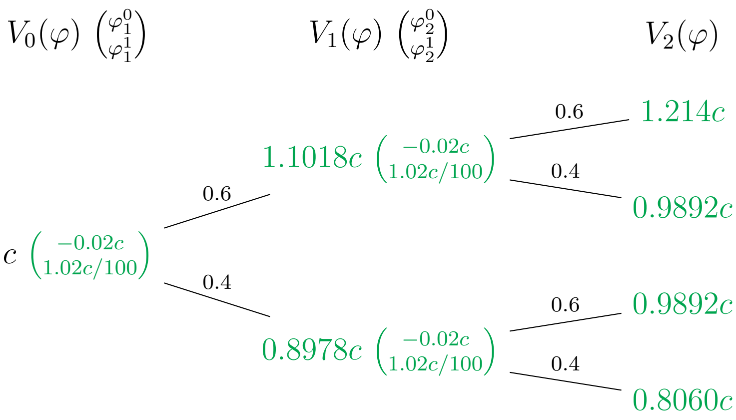

Figure 9.1.2: The optimal portfolio φ for expected logarithmic utility and its wealth process $V(\varphi)$

option trade is involved, the maximal expected utility amounts to

$$U_0 := \sup_{\varphi} E(u(\hat{V}_N(\varphi))), \tag{9.7}$$

where φ runs through all self-financing trading strategies with discounted initial value $\hat{V}_0$. If, on the other hand, the bank decides to enter the OTC option trade at a discounted premium $\hat{\pi} = \pi/S_0^0$, it can obtain the maximal expected utility

$$U_X(\pi) := \sup_{\varphi} E(u(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi))) \tag{9.8}$$

because the option's premium and obligation change the discounted terminal wealth by $\hat{\pi} - \hat{X}$. The bank will enter the trade only if it is not unfavourable, i.e. if $U_X(\pi) \geq U_0$. This happens if and only if $\pi \geq \pi^\star$ for some threshold premium $\pi^\star$. This **utility indifference price** is characterised by the condition

$$U_X(\pi^\star) = U_0.$$

$\pi^\star$ is unique because the utility function is supposed to be strictly increasing. We denote the optimal portfolios in (9.7) und (9.8) as $\varphi^\star$ resp. $\varphi^{X,\pi}$. Their difference $\varphi := \varphi^{X,\hat{\pi}} - \varphi^\star$ (or, more precisely, the corresponding self-financing trading strategy with initial value $\hat{\pi}$) is called **utility-based hedging strategy**. It represents the correction of the bank's optimal portfolio which is due to the option trade.

How can we compute $\pi^\star$, $\varphi^\star$, $\varphi^{X,\hat{\pi}}$? The plain utility maximisation problem (9.7) is discussed in the previous section. The modified problem (9.8) can be treated along the same lines. In contrast to (9.1), the density of an equivalent martingale measure Q must be of the form

$$\frac{dQ}{dP} = \frac{u'(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi))}{E(u'(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi)))} \tag{9.9}$$

in order for φ to solve the optimisation problem (9.8). The utility indifference price can be determined in a second step. However, closed-form expression are available only in rare cases.

Remark. In general it is not obvious whether the optimal values in (9.7) and (9.8) are attained. We simply assume the existence of optimal strategies in the following. For the most common utility functions, absence of arbitrage implies that it suffices in (9.7, 9.8) to consider portfolio φ from some compact set. By continuity of expected utility this implies that the maximal value is in fact obtained by some φ. In the sequel we also assume without proof that the optimisers in (9.7, 9.8) are linked to some EMM's via Equations (9.1, 9.9).

More specific statements can be made in the case of exponential utility. To this end, we henceforth suppose that $u(x) = 1 - \exp(-\lambda x)$ for some $\lambda > 0$. Observe that the discounted initial wealth $\hat{V}_0$ affects the shifted utility of terminal wealth $u(\hat{V}_N(\varphi)) - 1)$ only by a constant factor $e^{-\lambda \hat{V}_0}$. Consequently, it essentially does not affect optimality of strategies, indifference prices etc. For ease of notation, we therefore focus without loss of generality on the case $\hat{V}_0 = 0$.

In order to solve (9.7, 9.8) and for computing the indifference price, the notion of entropy turns out useful.

Definition 9.1. For probability measures $P \sim Q$

$$H(Q,P) := E_Q(\log \tfrac{dQ}{dP}) = E_P(\tfrac{dQ}{dP} \log \tfrac{dQ}{dP})$$

is called **relative entropy** of Q with respect to P.

We generally have

$$H(Q,P) \geq H(P,P) = 0,$$

because Jensen's inequality for $f(x) = -\log x$ yields

$$\begin{aligned} H(Q,P) &= E_Q(\log \tfrac{dQ}{dP}) \\ &= E_Q(-\log \tfrac{dP}{dQ}) \\ &\geq -\log E_Q(\tfrac{dP}{dQ}) \\ &= -\log 1 \\ &= 0 \quad = \quad H(P,P). \end{aligned}$$

The entropy plays a key role in the context of optimization for exponential utility.

Theorem 9.2. *Define the probability measure $P_X \sim P$ by its density*

$$\frac{dP_X}{dP} := \frac{e^{\lambda \hat{X}}}{E(e^{\lambda \hat{X}})}.$$

Moreover, denote by Q_0 and Q_X the EMM's that correspond via (9.1,9.9) to the utility maximisation problems (9.7) resp. (9.8).

1. *Q_0 maximizes the relative entropy $H(Q,P)$ among all EMM's Q. In particular, it does not depend on λ. The maximal expected utility in (9.7) amounts to*

$$E(u(\hat{V}_N(\varphi^\star))) = 1 - \exp(-H(Q_0,P)).$$

2. *Q_X maximizes the relative entropy*

$$H(Q,P_X) = H(Q,P) - \lambda E_Q(\hat{X}) + \log(c_X)$$

among all EMM's Q, where $c_X := E(e^{\lambda\hat{X}})$. The maximal expected utility in (9.8) amounts to

$$\begin{aligned} E(u(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi^{X,\pi}))) &= 1 - c_X \exp(-H(Q_X,P_X) - \lambda\hat{\pi}) \\ &= 1 - \exp(-H(Q_X,P) + \lambda E_{Q_X}(\hat{X}) - \lambda\hat{\pi}). \end{aligned}$$

3. *The utility-based hedging strategy $\varphi = \varphi^{X,\pi} - \varphi^\star$ and the EMM Q_X do not depend on the negotiated option premium π (except for the numeraire part φ^0) .*

4. *The utility indifference price $\pi^\star = \hat{\pi}^\star S_0^0$ is given by*

$$\begin{aligned} \hat{\pi}^\star &= \frac{1}{\lambda}\Big(\log E(e^{\lambda\hat{X}}) + H(Q_0,P) - H(Q_X,P_X)\Big) \\ &= E_{Q_X}(\hat{X}) + \frac{1}{\lambda}\left(H(Q_0,P) - H(Q_X,P)\right) && (9.10) \\ &= \frac{1}{\lambda}\log E_{Q_0}\Big(\exp(-\lambda(\varphi \bullet \hat{S}_N - \hat{X}))\Big), && (9.11) \end{aligned}$$

where φ denotes the utility-based hedging strategy.

Proof. 1. For $x > 0$ define $v(x) = x\log x$. This is a strictly convex function by $v''(x) = 1/x > 0$. For Q_0 as in 9.3 we have

$$v'(\tfrac{dQ_0}{dP}) = 1 + \log(\tfrac{dQ_0}{dP}) = c - \lambda\hat{V}_N(\varphi^\star)$$

with some constant $c \in \mathbb{R}$. Convexity of v implies $v(y) \geq v(x) + v'(x)(y-x)$ for any $x, y > 0$. This yields

$$\begin{aligned} H(Q,P) &= E(v(\tfrac{dQ}{dP})) \\ &\leq E(v(\tfrac{dQ_0}{dP})) + E(v'(\tfrac{dQ_0}{dP})(\tfrac{dQ}{dP} - \tfrac{dQ_0}{dP})) \\ &= H(Q_0,P) + E_Q(\lambda\hat{V}_N(\varphi^\star)) - E_{Q_0}(\lambda\hat{V}_N(\varphi^\star)) \\ &= H(Q_0,P) \end{aligned}$$

for any EMM Q. By strict convexity of v equality holds only for $Q = Q_0$, whence this **minimal entropy martingale measure** is unique.

From $\frac{dQ_0}{dP} = c^{-1}\exp(-\lambda\hat{V}_N(\varphi^\star))$ with $c = E(\exp(-\lambda\hat{V}_N(\varphi^\star)))$ we conclude

$$\begin{aligned} H(Q_0,P) &= E_{Q_0}(\log(\tfrac{dQ_0}{dP})) \\ &= \log(c^{-1}) - \lambda E_{Q_0}(\hat{V}_N(\varphi^\star)) \\ &= -\log c \end{aligned}$$

and hence

$$\begin{aligned} E(u(\hat{V}_N(\varphi^\star))) &= 1 - E(\exp(-\lambda \hat{V}_N(\varphi^\star))) \\ &= 1 - c \\ &= 1 - \exp(-H(Q_0, P)). \end{aligned}$$

2. Firstly we have

$$\begin{aligned} H(Q, P_X) &= E_Q(\log \tfrac{dQ}{dP_X}) \\ &= E_Q(\log \tfrac{dQ}{dP} - \log \tfrac{dP_X}{dP}) \\ &= H(Q, P) - \lambda E_Q(\hat{X}) + \log(c_X). \end{aligned}$$

Note that

$$\begin{aligned} \frac{dQ_X}{dP_X} &= \frac{dQ_X}{dP}\frac{dP}{dP_X} \\ &= \frac{\exp(-\lambda(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi^{X,\pi})))}{E(\exp(-\lambda(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi^{X,\pi}))))}\frac{E(e^{\lambda\hat{X}})}{e^{\lambda\hat{X}}} \\ &= \frac{E(e^{\lambda\hat{X}})}{E(\exp(-\lambda(-\hat{X} + \hat{V}_N(\varphi^{X,\pi}))))}\exp(-\lambda\hat{V}_N(\varphi^{X,\pi})) \\ &= cu'(\hat{V}_N(\varphi^{X,\pi})) \end{aligned}$$

for some constant $c > 0$. This density is of the form in Theorem 9.3 if we consider the plain investment problem without option, but subject to probabilities P_X rather than P. Moreover, $\varphi^{X,\pi}$ is the corresponding optimal portfolio. By Statement 1, Q_X minimises the relative entropy with respect to P_X among all EMM's Q. Statement 1 for Q_X, P_X instead of Q_0, P also yields

$$\begin{aligned} E(u(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi^{X,\pi}))) &= 1 - E(\exp(-\lambda(\hat{\pi} - \hat{X} + \hat{V}_N(\varphi^{X,\pi})))) \\ &= 1 - e^{-\lambda\hat{\pi}}E(e^{\lambda\hat{X}})E_{P_X}(\exp(-\lambda\hat{V}_N(\varphi^{X,\pi}))) \\ &= 1 + e^{-\lambda\hat{\pi}}c_X\left(E_{P_X}(u(\hat{V}_N(\varphi^{X,\pi}))) - 1\right) \\ &= 1 - e^{-\lambda\hat{\pi}}c_X \exp(-H(Q_X, P_X)) \end{aligned}$$

as claimed.

3. As observed for $\hat{V}_0$, the discounted option premium $\hat{\pi}$ enters shifted utility in (9.8) only as a multiplicative constant. Therefore, it does not affect the optimality of a strategy and the corresponding EMM Q_X in (9.9).

4. By Statements 1,3 the utility indifference price π^* satisfies

$$1 - \exp(-H(Q_0, P)) = 1 - c_X \exp(-H(Q_X, P_X))e^{-\lambda\hat{\pi}},$$

i.e.

$$e^{\lambda\hat{\pi}} = c_X \exp(-H(Q_X, P_X) + H(Q_0, P))$$

or

$$\begin{aligned}\hat{\pi} &= \frac{1}{\lambda}\left(\log(E(e^{\lambda\hat{X}})) - H(Q_X, P_X) + H(Q_0, P)\right)\\ &= \frac{1}{\lambda}\left(\lambda E_{Q_X}(\hat{X}) - H(Q_X, P) + H(Q_0, P)\right).\end{aligned}$$

On the other hand, $U_0 = U_X(\pi^\star)$ means

$$E(\exp(-\lambda\varphi^\star \bullet \hat{S}_N))) = E(\exp(-\lambda(\pi^\star - \hat{X} + \varphi^{X,\pi} \bullet \hat{S}_N)))$$

and hence

$$1 = e^{-\lambda\pi^\star} E_{Q_0}(\exp(-\lambda(-\hat{X} + \varphi \bullet \hat{S}_N)))$$

by $\varphi^{X,\pi} = \varphi^\star + \varphi$ and $\frac{dQ_0}{dP} = \frac{\exp(-\lambda\varphi^\star \bullet \hat{S}_N))}{E(\exp(-\lambda\varphi^\star \bullet \hat{S}_N)))}$. This yields the second representation. □

The preceding theorem clarifies the structure of the problem but it does not provide a concrete solution, which is generally not easy to obtain. In the following we focus on very small resp. very large risk aversion λ, i.e. we study the asymptotics for $\lambda \to 0$ and $\lambda \to \infty$. Since the utility function u, the maximal expected utilities U_0 resp. $U_X(\pi)$, the optimal portfolios $\varphi^\star, \varphi^{X,\pi}, \varphi := \varphi^{X,\pi} - \varphi^\star$, and the indifference price $\pi^\star$ depend on λ we subsequently use the notation $u_\lambda, U_0(\lambda), U_X(\pi, \lambda), \varphi^\star(\lambda), \varphi^{X,\pi}(\lambda), \varphi(\lambda), \pi^\star(\lambda)$.

Lemma 9.3. *We have $\varphi^\star(\lambda) = \frac{1}{\lambda}\varphi^\star(1)$ (or it can be choosen in this way in case of ambiguity). Moreover, $U_0(\lambda) = U_0(1)$ for any $\lambda > 0$.*

Proof. We have

$$\frac{dQ_0}{dP} = cu_1'(\hat{V}_N(\varphi^\star(1))) = c\exp(-\varphi^\star(1) \bullet \hat{S}_N) = \tfrac{c}{\lambda}u_\lambda'(\hat{V}_N(\tfrac{1}{\lambda}\varphi^\star(1)))$$

for some normalizing constant $c > 0$. Since Q_0 is an EMM, $\frac{1}{\lambda}\varphi^\star(1)$ is an optimal portfolio for u_λ by Th. 9.3. The second statement follows from Theorem 9.2(1). □

It is not obvious how the utility indifference price changes with λ. At least, we can show monotonicity.

Lemma 9.4. *$\pi^\star(\lambda)$ is increasing in λ.*

Proof. Theorem 9.2 yields

$$\begin{aligned}\hat{\pi}^\star(\lambda) &= E_{Q_X}(\hat{X}) + \frac{1}{\lambda}\left(H(Q_0, P) - H(Q_X, P)\right)\\ &= \inf_{Q \text{ ÄMM}}\left(E_Q(\hat{X}) + \frac{1}{\lambda}\left(H(Q_0, P) - H(Q, P)\right)\right).\end{aligned}$$

Since Q_0 is the minimal entropy martingale measure, we have $H(Q_0, P) - H(Q, P) \leq 0$ for any EMM Q. Consequently, the above expression is increasing in λ. □

We now turn to first-order approximations for small λ, on the basis of a purely heuristic reasoning. A rigorous proof is beyond of the score of this introductory text. We start from the natural assumption that the utility-based hedge converges for vanishing λ in the sense that

$$\varphi(\lambda) = \eta + O(\lambda)$$

for $\lambda \to 0$, where η denotes the desired limiting hedge and $O(\lambda)$ stands for some strategy such that $O(\lambda)/\lambda$ is bounded by some constant that does not depend on λ. For the aggregate portfolio $\varphi^{X,\pi^\star}(\lambda)$ we obtain

$$\varphi^{X,\pi^\star}(\lambda) = \varphi^\star(\lambda) + \varphi(\lambda) = \frac{1}{\lambda}\varphi^\star(1) + \eta + O(\lambda). \tag{9.12}$$

For the utility indifference price, we expect a linear approximation

$$\pi^\star(\lambda) = \pi^\star(0) + \lambda\delta + O(\lambda^2) \tag{9.13}$$

with real numbers $\pi^\star(0), \delta \in \mathbb{R}$ and some $O(\lambda^2)$ such that $O(\lambda)/\lambda^2$ is bounded by some constant that does not depend on λ. In the following we want to determine the unknown quantities $\eta, \pi^\star(0), \delta$.

To this end, we consider the maximization problem (9.8) for strategies of the form $\varphi = \frac{1}{\lambda}\varphi^\star(1)+\eta+O(\lambda)$. In view of the Taylor approximation $e^x = 1+x+\frac{x^2}{2}+O(x^3)$ we obtain

$$\begin{aligned}
&E(u_\lambda(\hat\pi - \hat X + \hat V_N(\varphi))) \\
&\quad= 1 - E\left(\exp(-\varphi^\star(1) \bullet \hat S_N - \lambda(\hat\pi - \hat X + (\eta + O(\lambda)) \bullet \hat S_N))\right) \\
&\quad= 1 - cE_{Q_0}\left(\exp(-\lambda(\hat\pi - \hat X + (\eta + O(\lambda)) \bullet \hat S_N))\right) \\
&\quad= 1 - c - \lambda c E_{Q_0}\left(\hat\pi - \hat X + (\eta + O(\lambda^2)) \bullet \hat S_N\right) \\
&\qquad + \frac{\lambda^2}{2} c E_{Q_0}\left((\hat\pi - \hat X + \eta \bullet \hat S_N)^2\right) + O(\lambda^3) \\
&\quad= 1 - c + \lambda c\left(\hat\pi - E_{Q_0}(\hat X)\right) - \frac{\lambda^2}{2} c E_{Q_0}\left((\hat\pi - \hat X + \eta \bullet \hat S_N)^2\right) + O(\lambda^3)
\end{aligned}$$

with $c = E(\exp(-\varphi^\star(1) \bullet \hat S_N)$, where the last equality holds because that $\hat S$ is a martingale relative to the EMM Q_0. If we neglect the $O(\lambda^3)$ term, we have to minimise

$$E_{Q_0}\left((\hat\pi - \hat X + \eta \bullet \hat S_N)^2\right) \tag{9.14}$$

as a function of strategy η. This can be viewed as a quadratic hedging problem as in Section 7.1. Corollary. 7.4 and the subsequent remark yield that

$$\Delta\langle \hat V, \hat S\rangle^{Q_0} = \eta^\top \Delta\langle \hat S, \hat S\rangle^{Q_0}$$

holds for the optimal strategy η, where $\hat V$ denotes the Q_0-martingale generated by the discounted payoff $\hat X$, i. e.

$$\hat V_n := E_{Q_0}(\hat X | \mathscr{F}_n)$$

and the predictable covariation refers to probability measure Q_0. In particular, we have

$$\eta_n = \frac{\Delta\langle \hat V, \hat S^1\rangle_n^{Q_0}}{\Delta\langle \hat S^1, \hat S^1\rangle_n^{Q_0}}$$

for $d = 1$. By Corollary 7.4 and the subsequent remark, the minimal value in (9.14) is given by $(\hat{\pi} - \hat{V}_0)^2 + \varepsilon^2$ for

$$\varepsilon^2 = E\left(\langle \hat{V}, \hat{V}\rangle_N^{Q_0} - \sum_{i,j=1}^{d} (\varphi^i \varphi^j) \bullet \langle \hat{S}^i, \hat{S}^j\rangle_N^{Q_0}\right).$$

We obtain

$$U_X(\pi, \lambda) = 1 - c + \lambda c(\hat{\pi} - \hat{V}_0) - \frac{\lambda^2}{2} c\left((\hat{\pi} - \hat{V}_0)^2 + \varepsilon^2\right) + O(\lambda^3)$$

if the linear expansion (9.12) of the optimal hedge holds. The maximal utility for the plain investment problem amounts to

$$U_0(\lambda) = U_0(1) = 1 - E(\exp(-\varphi^\star(1) \bullet \hat{S}_N)) = 1 - c.$$

The utility indifference price $\pi^\star(\lambda)$ solves $U_X(\pi^\star(\lambda), \lambda) = U_0(\lambda)$. If the linear approximation (9.13) holds, we obtain

$$0 = (\hat{\pi}^\star(0) - \hat{V}_0) - \frac{\lambda}{2} c\left(2\delta + (\hat{\pi}^\star(0) - \hat{V}_0)^2 + \varepsilon^2\right) + O(\lambda^2).$$

This implies

$$\hat{\pi}^\star(0) = \hat{V}_0 = E_{Q_0}(\hat{X})$$

and, in view of the linear term in λ,

$$\delta = \frac{\varepsilon^2}{2}.$$

Consequently, the discounted utility indifference price converges for $\lambda \to 0$ to the expectation of the discounted payoff relative to the minimal entropy martingale measure Q_0. On top of this zeroth order approximation, we have in first order a risk premium depending linearly on λ. It depends on how well the option can be approximated by a self-financing portfolio. Moreover, the utility-based hedge equals to the leading order the variance-optimal hedge of the claim relative to the MEMM.

As another extreme, let us consider instead the hint $\lambda \to \infty$ of large risk aversion. Here, we can provide a rigorous proof in our setup of finite underlying sample space.

Theorem 9.5. *1. We have $\pi^\star(\lambda) \to \pi_U$ for $\lambda \to \infty$, where π_U denotes the upper price in Section 5.3.*

2. $\varphi(\lambda)$ is an asymptotic superhedge in the sense that

$$\liminf_{\lambda\to\infty} \hat{V}_N(\varphi(\lambda)) \geq \hat{X}, \tag{9.15}$$

where $\hat{V}_N(\varphi(\lambda)) = \pi^\star(\lambda) + \varphi(\lambda) \bullet \hat{S}_N$ denotes the discounted terminal value of the utility-based hedge.

3. If all cheapest superhedges φ_U in Section 5.3 have the same final value $\hat{V}_N(\varphi_U)$ (e.g. since the cheapest superhedge is unique), we have

$$\hat{V}_n(\varphi(\lambda)) \to \hat{V}_n(\varphi_U)$$

fur $\lambda \to \infty$ and $n = 0, \dots, N$.

Proof. 1. The inequality $H(Q_0, P) - H(Q_X, P) \leq 0$ and (9.10) imply $\hat{\pi}^\star(\lambda) \leq E_{Q_X}(\hat{X}) \leq \hat{\pi}_U$ for $\lambda > 0$.

Wrongly suppose that $\sup_{\lambda>0} \hat{\pi}^\star(\lambda) \leq \hat{\pi}_U - 2\varepsilon$ for some $\varepsilon > 0$. For fixed λ we have $\omega_\lambda \in \Omega$ with

$$\hat{V}_N(\varphi(\lambda))(\omega_\lambda) := \hat{\pi}^\star(\lambda) + \varphi(\lambda) \bullet \hat{S}_N(\omega_\lambda) \leq \hat{X}(\omega_\lambda) - \varepsilon.$$

Indeed, otherwise $\hat{\pi}^\star(\lambda) + \varepsilon < \hat{\pi}_U$ would be the discounted initial value of a superhedge. We conclude

$$\begin{aligned} U_0(\lambda) &= U_X(\pi^\star(\lambda), \lambda) \\ &= E(u_\lambda(\hat{V}_N(\varphi(\lambda)) - \hat{X})) \\ &= 1 - E(\exp(-\lambda(\hat{\pi}^\star(\lambda) + \varphi(\lambda) \bullet \hat{S}_N - \hat{X}))) \\ &\leq 1 - P(\{\omega_\lambda\}) \exp(\lambda\varepsilon) \\ &\to -\infty \quad \text{fÃ¼r } \lambda \to \infty, \end{aligned}$$

because the finiteness of Ω implies $\min_{\omega\in\Omega} P(\{\omega\}) > 0$. This, however, contradicts $U_0(\lambda) \geq 1 - e^0 = 0$.

2. Wrongly suppose that (9.15) does not hold. Then there are $\omega \in \Omega$ and $\varepsilon > 0$, such that $\hat{V}_N(\varphi(\lambda))(\omega) \leq \hat{X}(\omega) - \varepsilon$ holds for some sequence of arbitrarily large λ. This, however leads to the contradiction in the proof of Statement 1.

3. Choose a sequence $(\lambda_k)_{k\in\mathbb{N}}$ in $\mathbb{R}_+$ such that $\lambda_k \to \infty$. We have to prove $\lim_{k\to\infty} \hat{V}_n(\varphi(\lambda_k)) = \hat{V}_n(\varphi_U)$ for $n = 0, \dots, N$. We start with $n = N$.

By (9.15) the sequence $(\hat{V}_N(\varphi(\lambda_k))(\omega))_{k\in\mathbb{N}}$ is bounded from below for any $\omega \in \Omega$. Since

$$E_{Q_0}(\hat{V}_N(\varphi(\lambda_k))) = \hat{V}_0(\varphi(\lambda_k)) = \hat{\pi}^\star(\lambda_k) \leq \hat{\pi}_U$$

and $\min_{\omega\in\Omega} P(\{\omega\}) > 0$, these sequences are bounded from above as well. Since $(\hat{V}_N(\varphi(\lambda_k)))_{k\in\mathbb{N}}$ is bounded in $\mathbb{R}^\Omega$, there is a convergent subsequence. We must show that any of these convergent subsequences converge to $\hat{V}_N(\varphi_U)$. W.l.o.g., we denote such an arbitrary subsequence again by $(\hat{V}_N(\varphi(\lambda_k)))_{k\in\mathbb{N}}$. Since $\{\hat{V}_N(\varphi) : \varphi \text{ self-financing strategy}\}$ is a subspace of $\mathbb{R}^\Omega$ and in particular bounded, we have $\hat{V}_N(\varphi(\lambda_k)) \to \hat{V}_N(\varphi)$ for some self-financing strategy φ. (9.15) yields $\hat{V}_N(\varphi) = \lim_{k\to\infty} \hat{V}_N(\varphi(\lambda_k)) \geq \hat{X}$, i.e., φ is a superhedge. On the other hand, the finiteness of Ω implies

$$\hat{V}_0(\varphi) = E_{Q_0}(\hat{V}_N(\varphi)) = \lim_{k\to\infty} E_{Q_0}(\hat{V}_N(\varphi(\lambda_k))) = \lim_{k\to\infty} \hat{\pi}^\star(\lambda) \leq \hat{\pi}_U,$$

which means that φ is a cheapest superhege and hence $\hat{V}_N(\varphi) = \hat{V}_N(\varphi_U)$.

It remains to be shown that convergence holds also for $n < N$. Pointwise convergence for $k \to \infty$ and once more finiteness of Ω yield

$$\hat{V}_n(\varphi(\lambda_k)) = E_{Q_0}(\hat{V}_N(\varphi(\lambda_k))|\mathscr{F}_n) \to E_{Q_0}(\hat{V}_N(\varphi_U)|\mathscr{F}_n) = \hat{V}_n(\varphi_U)$$

as desired. □

The above results provide a link between the different approaches to valuing and hedging OTC-derivates. For exponential utility, one obtains a continuum of prices and hedging strategies. Superreplication is located at one and of this spectrum, requiring typically – as discussed in Section 5.3 – an exceedingly high option premium. At the other end we have obtained expressions which are closely linked to variance-optimal hedging as presented in Section 7.1. Hence we observe in hindsight that this approach is of interest even if one rejects the symmetric treatment of profits and losses inherent in quadratic loss functions.

Bibliography

Nicole Bäuerle and Ulrich Rieder. *Finanzmathematik in diskreter Zeit*. Springer, 2017.

Simone Calogero. *A First Course in Options Pricing Theory*. SIAM, 2023.

Benoîte De Saporta and Mounir Zili. *Martingales and Financial Mathematics in Discrete Time*. John Wiley & Sons, 2021.

Hans Föllmer and Alexander Schied. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter, 2011.

Albrecht Irle. *Finanzmathematik*. Springer, 2003.

Jean Jacod and Philip Protter. *Probability Essentials*. Springer, Berlin, second edition, 2004.

Stanley Pliska. *Introduction to mathematical finance*. Blackwell publishers Oxford, 1997.

Index