Portfolio Theory

P.J.C. Spreij

this version: May 25, 2023

Contents

Preface

These lecture notes have originally been written for and during the course *Portfolio Theory* at the Universiteit van Amsterdam in Fall 2007.

The aim of the course is to introduce the fundamental concepts that underly the problem of portfolio optimization. Needed for this is also an exposition of the fundamental notions of financial markets, such as absence of arbitrage and completeness. Other concepts that will be developed are preference relations and utility. All this will first be done for a one-period market and later on extended to markets with a larger horizon. In the latter case we show how to use Dynamic Programming for optimization problems. We confine ourselves to models in discrete time, although portfolio optimization in continuous time is a topic that equally well deserves a place in this set of lecture notes. This will be a topic of future consideration.

The lecture notes are for a large deal based on the book *Stochastic Finance, An Introduction in Discrete Time* by Alexander Schied and Hans Föllmer. But also other sources like *Introduction to Mathematical Finance* by Stanley R. Pliska have been consulted, as well as *Stochastic systems: estimation, identification and adaptive control* by P.R. Kumar and Pravin Varaiya.

Finally, these lectures notes are constantly in a state of revision. Throughout the years, many errors and omissions in earlier versions have been corrected, thanks to careful reading by Attila Herczegh, Demeter Kiss and Kamil Kosiński, who were among the first students that took this course. Later on, Johan du Plessis, Hailong Bao, Nicos Starreveld, Laurens Sluijterman and Hymke ten Have provided me with more very useful feedback that resulted in the elimination of other inaccuracies. But it is almost inevitable that some remained, or new ones appear. Readers are kindly invited to report remaining mistakes. Suggestions for improvements are equally welcome.

Amsterdam, Spring 2021 Peter Spreij

1 Valuation in a one-period model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We will assume that all random variables that we encounter below are defined on this space and real valued. We assume that there are only two relevant time instants, $t = 0$ and $t = 1$. First we describe a market consisting of $d + 1$ assets. The assets are numbered from 0 to d. The zero-th asset is a non-risky asset. Given an interest rate $r > -1$ and a price π_0 of the asset at time $t = 0$, its price at $t = 1$ is given by $S_0 = \pi_0(1 + r) > 0$. We will make the convention that $\pi_0 = 1$. The other d assets are risky. This means that their prices π_i at $t = 0$ are known deterministic nonnegative numbers, but their prices S_i at $t = 1$ are not exactly known at $t = 0$ and are modeled as nonnegative random variables. Next to prices, we also have quantities, which are for the i-th asset given by real constants ξ_i. Note that negative ξ_i are allowed, this has to be interpreted as borrowing, or short selling products. Also non-integer quantities are allowed, one can buy or sell $\log 12$ units of an asset. We introduce the following vectors.

The price vector at $t = 0$ of these assets will be denoted by $\bar{\pi} = (\pi_0, \pi)$, where π denotes the vector of the d risky assets. At $t = 1$ we have with similar notation $\bar{S} = (S_0, S)$. Likewise we have for the quantities $\bar{\xi} = (\xi_0, \xi)$. The vector $\bar{\xi}$ will often be referred to as a *portfolio*. The *value* of the portfolio at $t = 0$ is then $W_0 = \bar{\xi} \cdot \bar{\pi}$ and $t = 1$ it is $W_1 = \bar{\xi} \cdot \bar{S}$. The dot here denotes the ordinary inner product.

1.1 Arbitrage

In a realistic market there will not exist arbitrage opportunities, making a sure profit by investing in a portfolio. We give a formal definition of this.

Definition 1.1 A portfolio $\bar{\xi}$ is an arbitrage opportunity if $W_0 \leq 0$, $W_1 \geq 0$ a.s. and $\mathbb{P}(W_1 > 0) > 0$. A market is called arbitrage free, if arbitrage opportunities don't exist.

Remark 1.2 Consider an arbitrage free market, with $\pi_i = 0$ for some asset i. Take the portfolio consisting of 1 unit of asset i only. Then $W_1 = S_i$. Since this portfolio is not an arbitrage opportunity, we must have that $S_i = 0$ a.s. Hence this asset is always worthless, and therefore we exclude zero initial prices. All π_i will be assumed strictly positive.

It will turn out useful to characterize arbitrage opportunities in terms of the risky assets only.

Lemma 1.3 *The existence of an arbitrage opportunity is equivalent to the existence of a vector ξ having the properties $\xi \cdot S \geq (1+r)\xi \cdot \pi$ a.s. and $\mathbb{P}(\xi \cdot S > (1+r)\xi \cdot \pi) > 0$.*

Proof Let $\bar{\xi} = (\xi_0, \xi)$ be an arbitrage opportunity. Then

$$\xi \cdot S - (1+r)\xi \cdot \pi = W_1 - (1+r)W_0 \geq W_1.$$

Since also $W_1 \geq 0$ a.s. and $\mathbb{P}(W_1 > 0) > 0$, the characterization follows.

Conversely, assume that the characterization holds true for some vector ξ. Choose $\xi_0 = -\xi \cdot \pi$. Then $W_0 = 0$ and $W_1 = \xi \cdot S - (1+r)\xi \cdot \pi$. It follows that $\bar{\xi}$ is an arbitrage opportunity. □

We now proceed with another characterization of arbitrage opportunities. We need the vector Y of *discounted net gains*. Its elements Y_i $(i = 1, \ldots, d)$ are given by

$$Y_i = \frac{S_i}{1+r} - \pi_i.$$

Corollary 1.4 *Existence of an arbitrage opportunity is equivalent to the existence of a vector ξ such that $\xi \cdot Y \geq 0$ a.s and $\mathbb{P}(\xi \cdot Y > 0) > 0$. An arbitrage free market is characterized by the implication $\xi \cdot Y \geq 0$ a.s. $\Rightarrow \xi \cdot Y = 0$ a.s.*

Proof This is an immediate consequence of Lemma 1.3. □

Definition 1.5 A probability measure $\mathbb{P}^*$ on $\mathcal{F}$ is called a risk-neutral measure, or a martingale measure, if

$$\pi_i = \mathbb{E}^* \frac{S_i}{1+r}, \; i = 0, \ldots, d.$$

It follows that $\mathbb{P}^*$ on $\mathcal{F}$ is a risk-neutral measure, iff $\mathbb{E}^* Y = 0$. Notice that $\mathbb{E}^* Y$ is well defined for any $\mathbb{P}^*$, since each Y_i is lower bounded by $-\pi_i$. By $\mathcal{P}$ we denote the set of all risk-neutral measures that are equivalent to $\mathbb{P}$ (they define the same null sets). Elements of $\mathcal{P}$ are called *equivalent martingale measures*.

Theorem 1.6 below is a version of the (first) *Fundamental Theorem of Asset Pricing* (FTAP).

Theorem 1.6 *A market is free of arbitrage iff the set $\mathcal{P}$ is nonempty. In this case there exists a $\mathbb{P}^* \in \mathcal{P}$ such that the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}$ is bounded.*

Proof Let $\mathcal{P}$ be non-empty and take $\mathbb{P}^* \in \mathcal{P}$. Let ξ be such that $\xi \cdot Y \geq 0$ $\mathbb{P}$-a.s. Then the same is true under $\mathbb{P}^*$. Since $\mathbb{P}^* \in \mathcal{P}$, we have $\mathbb{E}^* \xi \cdot Y = 0$ and hence $\xi \cdot Y = 0$ $\mathbb{P}^*$-a.s., but then also under $\mathbb{P}$. The result follows from Corollary 1.4.

Conversely, assume that the market is arbitrage free. We have to show the existence of a $\mathbb{P}^* \sim \mathbb{P}$ such that $\mathbb{E}^* Y = 0$. We first assume that $\mathbb{E}|Y| < \infty$. Put

$$\mathcal{Q} = \{\mathbb{Q} : \mathbb{Q} \sim \mathbb{P} \text{ and } \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \text{ bounded}\},$$

and

$$\mathcal{C} = \{\mathbb{E}_{\mathbb{Q}} Y : \mathbb{Q} \in \mathcal{Q}\}.$$

Notice that $\mathcal{C}$ is well defined, since $\mathbb{E}_{\mathbb{Q}}|Y| < \infty$ for all $\mathbb{Q} \in \mathcal{Q}$. One easily shows that $\mathcal{Q}$ is a convex set, and hence $\mathcal{C}$ is a convex subset of $\mathbb{R}^d$. For any $\mathbb{P}^*$ one has $\mathbb{E}^* Y = 0$, and we therefore show that $0 \in \mathcal{C}$.

Assume the contrary, $0 \notin \mathcal{C}$. By virtue of the *Separating Hyperplane Theorem*, Theorem A.1, there exists a vector $\xi \in \mathbb{R}^d$ such that $\xi \cdot x \geq 0$ for all $x \in \mathcal{C}$ and $\xi \cdot x_0 > 0$ for some $x_0 \in \mathcal{C}$. But elements of $\mathcal{C}$ are expectations, so we have $\mathbb{E}_{\mathbb{Q}}\xi \cdot Y \geq 0$ for all $\mathbb{Q} \in \mathcal{Q}$ and $\mathbb{E}_{\mathbb{Q}_0}\xi \cdot Y > 0$ for some $\mathbb{Q}_0 \in \mathcal{Q}$. Since the latter expectation is strictly positive, we must have $\mathbb{Q}_0(\xi \cdot Y > 0) > 0$, and by equivalence we then also have $\mathbb{P}(\xi \cdot Y > 0) > 0$. If we can show that $\xi \cdot Y \geq 0$ a.s., then we have shown existence of an arbitrage opportunity in view of Lemma 1.3, a contradiction. Consider thereto the set $A = \{\xi \cdot Y < 0\}$. The following arguments are aimed at showing $\mathbb{P}(A) = 0$.

Define for $n \geq 2$ the functions ϕ_n by

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

Note that $\frac{1}{n} \leq \phi_n \leq 1 - \frac{1}{n}$ as $n \geq 2$, from which we obtain that $\mathbb{E}\phi_n \geq 1/n > 0$. We can therefore define the probability measures $\mathbb{Q}_n$ by $\frac{\mathrm{d}\mathbb{Q}_n}{\mathrm{d}\mathbb{P}} = c_n\phi_n$, where $c_n = 1/\mathbb{E}\phi_n$. The $\mathbb{Q}_n$ are in $\mathcal{Q}$ as now $\frac{\mathrm{d}\mathbb{Q}_n}{\mathrm{d}\mathbb{P}} \leq n - 1$. We have the following string of equalities (we also use the Dominated Convergence Theorem)

$$\begin{aligned}
\mathbb{E}(\xi \cdot Y\mathbf{1}_{\{\xi \cdot Y<0\}}) &= \mathbb{E}(\xi \cdot Y\mathbf{1}_A) \\
&= \mathbb{E}(\xi \cdot Y \lim \phi_n) \\
&= \lim \mathbb{E}(\xi \cdot Y\phi_n) \\
&= \lim \frac{1}{c_n}\mathbb{E}_{\mathbb{Q}_n}(\xi \cdot Y) \geq 0.
\end{aligned}$$

So the nonpositive random variable $\xi \cdot Y\mathbf{1}_{\{\xi \cdot Y<0\}}$ has a nonnegative expectation. This implies $\mathbb{P}(\xi \cdot Y \geq 0) = 1$, which we wanted to prove.

The case $\mathbb{E}|Y| = \infty$ is left as Exercise 1.3. Note that the assertion on the existence of a bounded Radon-Nikodym derivative follows by the definition of the set $\mathcal{C}$. □

The following example shows that in an arbitrage free market, risk neutral measures are in general not unique.

Example 1.7 Suppose that $\Omega = \{\omega_1, \ldots, \omega_n\}$ with $n \geq 2$. Assume that $p_i = \mathbb{P}(\{\omega_i\}) > 0$ for all i, there is only one risky asset S_1 and $s_i = S_1(\omega_i) > 0$ for all i. Assume $s_1 < \cdots < s_n$. Consider an arbitrage opportunity with $\xi > 0$. Lemma 1.3 implies that $s_i \geq (1 + r)\pi$ for all i and hence $s_1 \geq (1 + r)\pi$. For an arbitrage opportunity with $\xi < 0$, the same lemma implies that $s_n \leq (1+r)\pi$. If no arbitrage exists we must have $s_1 < (1 + r)\pi < s_n$. Theorem 1.6 says that in the latter case for every $\mathbb{P}^*$, represented by a probability vector $(p_1^*, \ldots, p_n^*)$, the p_i^* solve $\sum_{i=1}^n p_i^* s_i = (1 + r)\pi_1$, next to $\sum_{i=1}^n p_i^* = 1$, and thus that a solution exists. Moreover, the risk-neutral measure is unique iff $n = 2$. This is related to the content of Section 1.3, where we discuss *completeness*.

Definition 1.8 The set of attainable pay-offs is $\mathcal{W} = \{\bar{\xi} \cdot \bar{S} : \bar{\xi} \in \mathbb{R}^{d+1}\}$. So, the elements of $\mathcal{W}$ are the random variables that can be seen as values of portfolios.

Lemma 1.9 *Assume that the market is arbitrage free. Suppose that $W \in \mathcal{W}$ can be represented both as $\bar{\xi} \cdot \bar{S}$ and as $\bar{\zeta} \cdot \bar{S}$. Then the values of the two portfolios at $t = 0$ are equal, $\bar{\xi} \cdot \bar{\pi} = \bar{\zeta} \cdot \bar{\pi}$.*

Proof Any $\mathbb{P}^* \in \mathcal{P}$ (which is non-empty) satisfies $\mathbb{E}^*\bar{S} = (1 + r)\bar{\pi}$. Hence $\mathbb{E}^*W = \mathbb{E}^*(\bar{\xi} \cdot \bar{S}) = \mathbb{E}^*(\bar{\zeta} \cdot \bar{S})$ yields $(1 + r)\bar{\xi} \cdot \bar{\pi} = (1 + r)\bar{\zeta} \cdot \bar{\pi}$. □

The above lemma yields *the principle, or law, of one price.* The price at $t = 0$ of an attainable pay-off W is equal to and defined by $\frac{\mathbb{E}^*W}{1+r}$, for any $\mathbb{P}^* \in \mathcal{P}$. This can be rephrased by saying that two portfolios which generate the same pay-off at $t = 1$ must have the same price at $t = 0$, which is then given by the expectation under any of the risk-neutral measures. Check that if the initial price vector $\bar{\pi}$ is different from $\frac{\mathbb{E}^*\bar{S}}{1+r}$, an arbitrage opportunity can explicitly be constructed.

1.2 Contingent claims and derivatives

There are many financial products other then portfolios (whose pay-off is attainable by definition). Some of these depend on the underlying risky assets, like call options. An example is the European call option whose pay-off is $C := (S_1 - K)^+$ (the constant K is called the strike price). We see that this pay-off is a function of S_1. We will also consider pay-offs that are (in principle) not functions of $\bar{S}$.

Definition 1.10 A contingent claim C is by definition a nonnegative random variable (so C is $\mathcal{F}$-measurable). Such a C is called a derivative if C is $\sigma(\bar{S}) = \sigma(S)$-measurable.

In this definition we require nonnegativity to ensure existence of an expectation. Alternatives to this are conceivable, like C lower bounded, or $\mathbb{E}|C| < \infty$. Another reason for nonnegativity is that we now treat C similar to the given assets, which are always assumed to have a nonnegative payoff.

Since we have seen how to price portfolios in arbitrage free markets (at $t = 0$), the natural question is how to price contingent claims. By analogy, an obvious candidate is $\frac{\mathbb{E}^*C}{1+r}$. It turns out that this is true, but also that in general this price is not unique and that, unlike for attainable pay-offs, it depends on the specific choice of the risk neutral measure.

Since absence of arbitrage is the key to finding a pricing rule for a contingent claim, we will look at the *extended market.* Next to the assets we already have, we consider the extra security $S_{d+1} = C$ and its price π_{d+1} at $t = 0$. No arbitrage considerations in the extended market will give the possible values of π_{d+1}.

Definition 1.11 A (nonnegative) real number π^C is called an arbitrage free price of the contingent claim C if the extended market with $\pi_{d+1} = \pi^C$ is free of arbitrage. We denote by $\Pi(C)$ the set of all prices π^C.

The next theorem confirms the conjecture made at the beginning of this section and gives a precise formulation.

Theorem 1.12 *Let C be a contingent claim. Suppose that the original market is arbitrage free (so $\mathcal{P} \neq \emptyset$). Then also $\Pi(C)$ is non-empty and in fact, one has*

$$\Pi(C) = \{\mathbb{E}^* \frac{C}{1+r} : \mathbb{P}^* \in \mathcal{P} \text{ such that } \mathbb{E}^* C < \infty\}. \tag{1.1}$$

Proof We first show that the set on the right hand side of (1.1) is not empty. Call this set E. Introduce a probability measure $\tilde{\mathbb{P}}$ by $\mathrm{d}\tilde{\mathbb{P}} = \frac{c}{1+C}\mathrm{d}\mathbb{P}$, where c is the normalizing constant. One sees that $\tilde{\mathbb{P}} \sim \mathbb{P}$ and that $\tilde{\mathbb{E}}C < \infty$. From the equivalence it follows that the market is also free of arbitrage under $\tilde{\mathbb{P}}$. Then Theorem 1.6 yields the existence of a risk-neutral measure $\mathbb{P}^*$ equivalent to $\tilde{\mathbb{P}}$ (and then also to $\mathbb{P}$) with $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\tilde{\mathbb{P}}}$ bounded, by B say. Then we have $\mathbb{E}^*C = \tilde{\mathbb{E}}(\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\tilde{\mathbb{P}}}C) \leq B\tilde{\mathbb{E}}C < \infty$. Hence the number $\frac{\mathbb{E}^*C}{1+r}$ belongs to the set E, which is thus not empty.

We now show that $\Pi(C) \subset E$. If $\Pi(C) = \emptyset$, there is nothing to prove. Therefore we assume that we can pick $\pi^C \in \Pi(C)$. By definition of the arbitrage price, the extended market is free of arbitrage, so in view of Theorem 1.6 there exists a probability measure $\hat{\mathbb{P}}$ equivalent to $\mathbb{P}$ such that $\hat{\mathbb{E}}\frac{S_i}{1+r} = \pi_i$, for all $i = 0, \ldots, d+1$. In particular we have that $\hat{\mathbb{E}}\frac{C}{1+r} = \pi^C < \infty$ (take $i = d+1$). Since $\hat{\mathbb{P}}$ is then also a risk-neutral measure for the original market, we have $\Pi(C) \subset E$.

To show the reversed inclusion, we take $\mathbb{P}^* \in \mathcal{P}$ and define $\pi_{d+1} = \pi^C := \frac{\mathbb{E}^*C}{1+r}$. This definition turns $\mathbb{P}^*$ into a risk-neutral measure for the extended market as well, and so we have $E \subset \Pi(C)$. □

In the proof of Theorem 1.16 the concept of *non-redundancy* comes in handy.

Definition 1.13 A market is called *non-redundant* if the implication $\bar{\xi} \cdot \bar{S} = 0$ $\mathbb{P}$-a.s. $\Rightarrow \bar{\xi} = 0$ holds.

If the implication in Definition 1.13 doesn't hold for some portfolio $\bar{\xi}$, it has a nonzero element, ξ_i say. It follows that $S_i = -\frac{1}{\xi_i}\sum_{j \neq i} \xi_j S_j$. So, the i-th asset price is a linear combination of the other ones, a form of redundancy. In an arbitrage free market we then also have (take expectations under any risk-neutral measure) $\pi_i = -\frac{1}{\xi_i}\sum_{j \neq i} \xi_j \pi_j$, again a linear combination and with the same coefficients, and likewise Y_i is the same linear combination of the Y_j.

Proposition 1.14 *The following holds.*

(i) Any (finite) market can be reduced to a non-redundant market in the sense that there exists a non-redundant market such that for any portfolio of assets in the original market one can find a portfolio in the non-redundant market with exactly the same payoff.

(ii) In a non-redundant market the implication $\xi \cdot Y = 0$ $\mathbb{P}$-a.s. $\Rightarrow \xi = 0$ holds. Conversely, if this implication holds and the market is arbitrage free, then the market is also non-redundant.

Proof Exercise 1.5. □

Remark 1.15 We will need later that $\Pi(C)$ is an interval; this follows from the characterization of $\Pi(C)$ as the right hand side of (1.1), a convex set. This motivates to study $\inf \Pi(C)$ and $\sup \Pi(C)$. Of course these quantities are of independent interest, as they give upper and lower bounds for the possible arbitrage free prices of a contingent claim C.

Theorem 1.16 *Assume that the market is arbitrage free. Let*

$$M_0 = \{m \in [0,\infty] : \exists \xi \in \mathbb{R}^d \text{ with } m + \xi \cdot Y \le \frac{C}{1+r} \mathbb{P}\text{-a.s.}\}$$

and

$$M_1 = \{m \in [0,\infty] : \exists \xi \in \mathbb{R}^d \text{ with } m + \xi \cdot Y \ge \frac{C}{1+r} \mathbb{P}\text{-a.s.}\}.$$

Then $\inf \Pi(C) = \max M_0$ *and* $\sup \Pi(C) = \min M_1$.

Proof We only give the proof of the characterization of the supremum. The other assertion follows by similar arguments. We first notice that $M_1 \neq \emptyset$, since $\infty \in M_1$ (take $\xi = 0$). Take $m \in M_1$ and $\xi \in \mathbb{R}^d$ such that $m + \xi \cdot Y \ge \frac{C}{1+r}$ $\mathbb{P}$-a.s. For any $\mathbb{P}^* \in \mathcal{P}$ we then have $m \ge \frac{\mathbb{E}^* C}{1+r}$. Taking the supremum on the right hand side over all $\mathbb{P}^*$ and then the infimum on the left hand side over all $m \in M_1$ yields $\inf M_1 \ge \sup \Pi(C)$ in view of Theorem 1.12.

We proceed by showing the reversed inequality. Since this is trivial if $\sup \Pi(C) = \infty$, we assume that $\sup \Pi(C) < \infty$. Pick $m > \sup \Pi(C)$. If we can show that $m \ge \inf M_1$, we are done by taking the limit $m \downarrow \sup \Pi(C)$. Since $m \notin \Pi(C)$, there exists an arbitrage opportunity in the extended market with $\pi_{d+1} = m$ and $S_{d+1} = C$. From Corollary 1.4, applied to the extended market (where the vector of net gains is the original Y appended with $\frac{C}{1+r} - m$), we obtain the existence of a vector $\xi \in \mathbb{R}^d$ and a real number ξ_{d+1} with the properties that

$$\xi \cdot Y + \xi_{d+1}(\frac{C}{1+r} - m) \ge 0 \quad \mathbb{P}\text{-a.s.}, \tag{1.2}$$

and strictly positive with positive $\mathbb{P}$-probability. Since the original market is arbitrage free, we can take $\mathbb{P}^* \in \mathcal{P}$ under which $\xi \cdot Y + \xi_{d+1}(\frac{C}{1+r} - m)$ has strictly positive expectation, $\mathbb{E}^*(\xi_{d+1}(\frac{C}{1+r} - m)) > 0$. Because $m > \sup \Pi(C) \ge \mathbb{E}^* \frac{C}{1+r}$, we find that $\xi_{d+1} < 0$. Consider the portfolio $\zeta = -\frac{\xi}{\xi_{d+1}}$. From (1.2) we obtain that $m + \zeta \cdot Y \ge \frac{C}{1+r}$. So $m \in M_1$ and thus $m \ge \inf M_1$, which we wanted to show.

The last thing to do is to show that the infimum is attained. If $\inf M_1 = \infty$, this is trivial. So, we assume that $\inf M_1 < \infty$. Choose a sequence of $m^n \in M_1$ that decreases to $\inf M_1$ and pick the corresponding ξ^n. We then have

$$m^n + \xi^n \cdot Y \geq \frac{C}{1+r} \ \mathbb{P}\text{-a.s.} \tag{1.3}$$

If we would know that the sequence (ξ_n) had a finite limit ξ, we could take limits in (1.3) to get $\inf M_1 + \xi \cdot Y \geq \frac{C}{1+r}$, in which case it follows that $\inf M_1 \in M_1$ and is thus attained. In general, the existence of a limit ξ cannot be guaranteed, but in the sequel we show that we can apply the above arguments along some subsequence.

Without loss of generality, we may assume that the market is non-redundant. Otherwise, see Proposition 1.14, we could replace the $\xi^n \cdot Y$ by a linear combination of non-redundant discounted net gains. Suppose that $\liminf ||\xi^n|| = \infty$. Then the vectors $\eta^n := \frac{\xi^n}{||\xi^n||}$ all lie on the (compact) unit circle and therefore converge along a subsequence (η^{n_k}) to some vector η with $\|\eta\| = 1$. Divide the inequality (1.3) by $||\xi^{n_k}||$ and take the limit for $k \to \infty$ to obtain $\eta \cdot Y \geq 0$ $\mathbb{P}$-a.s. Absence of arbitrage entails $\eta \cdot Y = 0$ $\mathbb{P}$-a.s. and non-redundancy then yields $\eta = 0$ (see Proposition 1.14). This contradicts $||\eta|| = 1$. Hence $\liminf ||\xi^n|| < \infty$. Choose then a subsequence (again denoted by) (ξ^{n_k}) converging to a finite limit ξ. Take along the same subsequence limits in (1.3) to arrive at $\inf M_1 + \xi \cdot Y \geq \frac{C}{1+r}$ $\mathbb{P}$-a.s. This shows that $\inf M_1 \in M_1$. □

Remark 1.17 The sets M_0 and M_1 in Theorem 1.16 are of interest in their own rights. One can show that the set M_1 coincides with the set of prices (at $t = 0$) of portfolios that *superhedge* the claim C, where a portfolio $\bar{\xi}$ superhedges C if $\bar{\xi} \cdot \bar{S} \geq C$ a.s. A similar characterization can be given for M_0. See also Exercise 1.2.

Definition 1.18 A contingent claim C is called *attainable* if C a.s. belongs to the space of attainable pay-offs $\mathcal{W}$, so $C = \bar{\xi} \cdot \bar{S}$ for some portfolio $\bar{\xi}$. Such a portfolio is called *replicating* portfolio, or *hedge*.

It is obvious that in an arbitrage free market, the arbitrage-free price of an attainable claim equals that of the replicating portfolio and is thus unique, due to the law of one price.

Proposition 1.19 *Let C be a contingent claim in an arbitrage free market. Then*

(i) C is attainable iff it admits a unique arbitrage-free price.

(ii) If C is not attainable, $\Pi(C)$ is the open interval $(\inf \Pi(C), \sup \Pi(C))$.

Proof (i) If C admits a unique price, the set $\Pi(C)$ is not an open interval, so (i) follows from (ii). We now prove the latter assertion. Let C be not attainable. As observed in Remark 1.15, $\Pi(C)$ is a non-empty interval. We only have to show that it is open, which happens if both $\inf \Pi(C)$ and $\sup \Pi(C)$ are not contained in it. We only consider $\inf \Pi(C)$ and suppose that $\inf \Pi(C) \in \Pi(C)$. Theorem 1.16 then implies that there exists $\xi \in \mathbb{R}^d$ such that $\inf \Pi(C) + \xi \cdot Y \leq \frac{C}{1+r}$ $\mathbb{P}$-a.s. Since C is not attainable we have $\mathbb{P}(\inf \Pi(C) + \xi \cdot Y < \frac{C}{1+r}) > 0$. Extend the original market with the risky asset having pay-off C endowed with the price $\inf \Pi(C)$ and consider the risky portfolio $-\xi$ augmented with one unit of the extra risky asset. The total net discounted gain of this extended portfolio is equal to $-\xi \cdot Y + \frac{C}{1+r} - \inf \Pi(C)$. The above inequalities show that we have constructed an arbitrage opportunity in the extended market. Hence $\inf \Pi(C)$ is not an arbitrage free price for C and thus not an element of $\Pi(C)$. □

1.3 Complete markets

In arbitrage free *complete* markets, as we shall see, every contingent claim has a unique price. We start with a definition.

Definition 1.20 A market is called *complete* if every contingent claim is attainable.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote for $p \geq 1$ by $L^p(\Omega, \mathcal{F}, \mathbb{P})$ the usual Banach space with the L^p-norm $||\cdot||_p$ defined by $||X||_p = (\mathbb{E}|X|^p)^{1/p}$. For $p = \infty$ we have $||X||_\infty = \inf\{c \geq 0 : \mathbb{P}(|X| > c) = 0\}$. Also for $0 < p < 1$ we consider $L^p(\Omega, \mathcal{F}, \mathbb{P})$, the set of X with $\mathbb{E}|X|^p < \infty$. By $L^0(\Omega, \mathcal{F}, \mathbb{P})$ we simply denote the linear space of a.s. finitely valued random variables. It can be considered as a (complete) *metric* space by defining the metric ρ by $\rho(X, Y) = \mathbb{E}\frac{|X-Y|}{1+|X-Y|}$. It then holds that $X_n \xrightarrow{\rho} X$ iff $X_n \xrightarrow{\mathbb{P}} X$. We will often use the same notation $L^p(\Omega, \mathcal{F}, \mathbb{P})$ when its elements are finite-dimensional random vectors $X = (X_1, \ldots, X_k)$ with $|X(\omega)|$ denoting any norm of the vector $X(\omega)$, usually $|X(\omega)| = (\sum_{j=1}^k X_j(\omega)^2)^{1/2}$.

In every market $\mathcal{W}$ is a subset of the set of $\mathcal{F}$-measurable random variables. If $\mathcal{C}$ is the set of attainable claims, we have the inclusions $\mathcal{C} \subset \mathcal{W} \subset L^0(\Omega, \mathcal{F}, \mathbb{P})$. In a complete market we have $L^0(\Omega, \mathcal{F}, \mathbb{P}) = \mathcal{C} + (-\mathcal{C}) \subset \mathcal{W}$ and so $L^0(\Omega, \mathcal{F}, \mathbb{P}) = \mathcal{W}$. It follows that in this case $L^0(\Omega, \mathcal{F}, \mathbb{P})$ is finite dimensional, with dimension less than or equal to $d + 1$ and also that $\mathcal{F}$ and $\sigma(S)$ are identical modulo null sets.

Recall the definition of an atom. A set $A \in \mathcal{F}$ is called an atom of $(\Omega, \mathcal{F}, \mathbb{P})$ if $\mathbb{P}(A) > 0$ and every measurable subset B of A either has probability zero or $\mathbb{P}(A)$.

Proposition 1.21 *Let $\mathcal{N}$ be the set of integers m for which there exists a measurable partition of Ω into m sets with positive probability. For any $p \in [0, \infty]$ one has $\dim L^p(\Omega, \mathcal{F}, \mathbb{P}) = \sup \mathcal{N}$. Moreover, $n := \dim L^p(\Omega, \mathcal{F}, \mathbb{P}) < \infty$ iff there exists a partition of Ω into n atoms.*

Proof Let $m \in \mathcal{N}$, then there exists a measurable partition of m sets with positive probability, $A_1, \ldots, A_m$ say. Then the set of corresponding indicators $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_m}$ is linearly independent in any $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Therefore $n \geq m$

and as m is arbitrary, also $n \geq \sup \mathcal{N}$. If $\mathcal{N}$ is unbounded, we trivially have $n = \sup \mathcal{N}$. Assume therefore that $n_0 := \sup \mathcal{N} < \infty$. Since $n_0 \in \mathcal{N}$ there exists a partition $A_1, \ldots, A_{n_0}$ of *atoms*. Indeed, if one of the A_i were not an atom, we split it into more sets of positive probability, but then n_0 would not be maximal. Hence we have that $\mathcal{F} = \sigma(A_1, \ldots, A_{n_0})$ modulo null sets and *every* $\mathcal{F}$-measurable function is almost surely constant on the A_i and hence a linear combination of the linearly independent $\mathbf{1}_{A_i}$. It follows that $n = n_0$. □

The following theorem is known as the second Fundamental Theorem of Asset Pricing.

Theorem 1.22 *An arbitrage free market is complete iff there exists exactly one risk-neutral measure. In this case* $\dim L^0(\Omega, \mathcal{F}, \mathbb{P}) \leq d+1$ *and* Ω *can be decomposed into at most* $d+1$ *atoms.*

Proof Let the market be complete. Consider the claim $C = \mathbf{1}_A(1+r)$ for some $A \in \mathcal{F}$. Then C is attainable and thus admits the unique fair price $\mathbb{P}^*(A)$, by virtue of Proposition 1.19, valid for any $\mathbb{P}^* \in \mathcal{P}$. But since A is arbitrary, this shows that $\mathbb{P}^*$ is unique.

Conversely, assume that $\mathcal{P} = \{\mathbb{P}^*\}$. Let C be a contingent claim. Theorem 1.12 says that the set $\Pi(C)$ of its arbitrage free prices has elements $\frac{\mathbb{E}^* C}{1+r} < \infty$. But since $\mathcal{P}$ is a singleton, also $\Pi(C)$ is a singleton and then Proposition 1.19 says that C is attainable. As C is arbitrary, any claim is attainable and thus the market is complete.

For a complete market we already observed that $L^0(\Omega, \mathcal{F}, \mathbb{P}) = \mathcal{W}$, and so $\dim L^0(\Omega, \mathcal{F}, \mathbb{P}) \leq d+1$, since $\dim \mathcal{W} \leq d+1$. Proposition 1.21 then implies that $(\Omega, \mathcal{F}, \mathbb{P})$ has at most $d+1$ atoms. □

It follows from Theorem 1.22 that for a complete market with finitely many securities we can as well set up a model with a finite space Ω. Let us take an example with n atoms, $\Omega = \{\omega_1, \ldots, \omega_n\}$. Let there be one risky asset with initial price π and S as its price at $t = 1$. Let $s_1 = S(\omega_1) < \cdots < s_n = S(\omega_n)$. We have seen before (Example 1.7) that any risk-neutral measure $\mathbb{P}^*$, represented by a vector $(p_1, \cdots, p_n)$ on the simplex, must satisfy $\sum_i p_i s_i = (1+r)\pi \in (s_1, s_n)$. It is obvious that $\mathbb{P}^*$ is unique iff $n = 2$, which is then the only case where the market is complete, in view of Theorem 1.22. Moreover, in this case one easily computes $p_1 = \frac{s_2 - (1+r)\pi}{s_2 - s_1}$ and $p_2 = \frac{(1+r)\pi - s_1}{s_2 - s_1}$.

It is also straightforward to compute the replicating portfolio for a given contingent claim C. Let $c_i = C(\omega_i), i = 1, 2$. If the claim is attainable one should find ξ_0 and ξ_1 such that $c_i = \xi_0(1+r) + \xi_1 s_i$, $i = 1, 2$. This linear system of equations is solved by

$$\xi_0 = \frac{c_1 s_2 - c_2 s_1}{(s_2 - s_1)(1+r)}$$

$$\xi_1 = \frac{c_2 - c_1}{s_2 - s_1}.$$

One also easily computes an explicit expression for $\pi^C = \xi_0 + \frac{\xi_1 \mathbb{E}^* S}{1+r}$. Furthermore, here we have $\dim L^p(\Omega, \mathcal{F}, \mathbb{P}) = d+1 = 2$.

1.4 Exercises

1.1 Consider a financial market as in Example 1.7 and assume $n = 3$.

(a) Characterize explicitly the set $\mathcal{P}$ as a set of vectors (p_1^*, p_2^*, p_3^*).

(b) Consider a call option $C = (S_1 - K)^+$ for some $K > 0$. The set $\Pi(C)$ will turn out to be an interval. What are the upper and lower limits?

1.2 Let C be a contingent claim in an arbitrage free market with discounted net gains vector Y. Consider the set M_1 of Theorem 1.16. Put $\pi^* := \inf M_1$. Show that π^* is the lowest price of all portfolios $\bar{\xi}$ that are such that $\bar{\xi} \cdot \bar{S} \geq C$ a.s.

1.3 Finish the proof of Theorem 1.6, by considering the case $\mathbb{E}|Y| = \infty$. *Hint:* Consider the auxiliary probability measure $\tilde{\mathbb{P}}$ defined by $\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}} = \frac{c}{1+|Y|}$, with c the normalizing constant, and reason as in the proof of Theorem 1.12

1.4 Write $\pi(W) = \bar{\pi} \cdot \bar{\xi}$ for $W = \bar{\xi} \cdot \bar{S}$. If $\pi(W) \neq 0$ we define the *return* of W by $R(W) := \frac{W - \pi(W)}{\pi(W)}$.

(a) Let $W = S_0$, so $\xi_0 = 1$ and the other ξ_i are zero. Compute $R(W)$.

(b) Assume an arbitrage free market. Show that $\mathbb{E}^*(W) = (1+r)\pi(W)$, for any $W \in \mathcal{W}$. Compute, for every $W \in \mathcal{W}$ with $\pi(W) \neq 0$, the expected return $\mathbb{E}^* R(W)$.

1.5 Here you prove Proposition 1.14.

(a) To prove the first statement you may proceed along the following pattern. Let $N = \{\bar{\xi} \in \mathbb{R}^{d+1} : \bar{\xi} \cdot \bar{S} = 0\}$ and assume that $\dim N > 0$. Let $N^\perp$ be the orthogonal complement of N in $\mathbb{R}^{d+1}$ with $\dim N^\perp = k + 1 < d + 1$. Show that there exists a matrix $B \in \mathbb{R}^{(d+1)\times(k+1)}$ of full column rank and a $(k+1)$-dimensional random vector $\bar{S}'$ such that $\bar{S} = B\bar{S}'$.

(b) Show that any portfolio of assets in the original market has a counterpart, a portfolio of assets in $\bar{S}'$ with the same payoff.

(c) Show that the $(k+1)$-dimensional market described by $\bar{S}'$ is non-redundant. Does it have a riskless asset?

1.6 Consider a market defined on a finite probability space (as in Example 1.7) with 1 riskless and d (instead of one) risky assets. Find a relation between n and d if the market is non-redundant and complete. Show that this condition is not sufficient and provide a sufficient condition for non-redundancy and completeness.

1.7 Consider next to the given market the alternative market with the original non-risky asset and d risky assets whose values at $t = 1$ are represented by the vector of discounted net gains Y. Assume that at $t = 0$ the non-risky asset has price $\pi_0' = \pi_0 = 1$ and the alternative risky assets have price vector π' equal to zero.

(a) Show that any portfolio in the original market has a counterpart in the alternative market with same pay-off at $t = 1$.

(b) Show that the alternative market is arbitrage free iff the original market is arbitrage free.

1.8 Assume that a markt admits an arbitrage opportunity. Show there exists a $\bar{\xi} \in \mathbb{R}^{d+1}$ such that $W_0 = 0$ a.s., $W_1 \geq 0$ a.s., and $\mathbb{P}(W_1 > 0) > 0$. [Conversely, such a $\bar{\xi}$ is an arbitrage opportunity in the sense of Definition 1.1.]

1.9 Suppose that the initial price vector $\bar{\pi}$ is different from $\mathbb{E}^* \frac{\bar{S}}{1+r}$. Construct an arbitrage opportunity.

1.10 Consider the sets M_0 and M_1 of Theorem 1.16. Show by a direct argument that $\inf M_1 \geq \sup M_0$.

1.11 Complete the proof of Theorem 1.16, i.e. show that $\inf \Pi(C) = \max M_0$.

1.12 Complete the proof of Proposition 1.19, i.e. show that $\sup \Pi(C) \notin \Pi(C)$.

1.13 Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assume Ω has three elements, $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Suppose there is, next to the riskless S_0, one risky asset S_1 that is not constant. Construct a non-attainable claim C and show that the market extended with C is complete.

1.14 Consider a market, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consisting of d risky assets and one riskless asset. Assume that it is arbitrage free, complete and non-redundant. Show that $\dim L^p(\Omega, \mathcal{F}, \mathbb{P}) = d+1$, for any $p \in [0, \infty]$, and that Ω can be decomposed into exactly $d+1$ atoms. Conversely, if $\dim L^p(\Omega, \mathcal{F}, \mathbb{P}) = d+1$, show that the market is arbitrage free, complete and non-redundant. [This is a new exercise, not sure it is correct.]

1.15 Consider the sets M_0 and M_1 of Theorem 1.16. Show that $M_0 \cap M_1 \neq \emptyset$ iff C is attainable and find for that case the intersection.

2 Preferences

In a market commodities are traded and an agent acting in this market will have certain preferences of some commodities over others. Think that he likes apples more than pears. Preferences will be made explicit by introducing *preference relations*. Traded commodities include risky assets as well, or contingent claims. Future pay-offs of these products are uncertain, random. We have seen in the previous section, that in a complete market such claims have a unique (arbitrage free) price, computed as an expectation under the unique risk-neutral measure. In incomplete markets, there is usually an interval of possible arbitrage free prices. To select one of these, preference relations or their numerical counterparts, *utility functions*, are instrumental. Utility functions also describe the attitude of an economic agent towards *risk* (incurred by the uncertain pay-offs). In this case, the resulting price of a contingent claim depends on how risk is quantified and for different utility functions, usually, different prices will be the result. Utility functions can also be used in a complete market to choose between two portfolios that have the same price. The treatment of utility functions is deferred to Section 4, here we study preference relations.

2.1 Preference relations

Let $\mathcal{X}$ be a non-empty set representing commodities or securities, or in general possible choices an economic agent can make. Recall that a binary relation R on $\mathcal{X}$ can be represented as a subset of $\mathcal{X} \times \mathcal{X}$ and that xRy means $(x, y) \in R$. The binary relations that are the topic of this section are denoted $\succ$, $\succeq$, $\prec$ and $\preceq$.

Definition 2.1 A *(strict) preference relation* or *preference order* on $\mathcal{X}$ is a binary relation $\succ$ with the properties

(i) Asymmetry: If $x \succ y$, then $y \not\succ x$.

(ii) Negative transitivity: If $x \succ y$ and $z \in \mathcal{X}$, then $x \succ z$ or $z \succ y$.

Definition 2.2 A *weak preference relation* on $\mathcal{X}$ is a binary relation $\succeq$ with the properties

(i) Completeness: For all $x, y \in \mathcal{X}$ one has $x \succeq y$ or $y \succeq x$.

(ii) Transitivity: If $x \succeq y$ and $y \succeq z$, then $x \succeq z$.

Proposition 2.3 *Strict and weak preference relations are connected in the following way. If $\succ$ is a strict preference relation on $\mathcal{X}$, then $x \succeq y$ defined by $y \not\succ x$ yields a weak preference relation on $\mathcal{X}$. Conversely, if $\succeq$ is a weak preference relation, then $x \succ y$ defined by $y \not\succeq x$ yields a strict preference relation.*

Proof Exercise 2.1. □

When dealing with strict and weak preference relations $\succ$ and $\succeq$, we will always assume that they are related as in Proposition 2.3. Given a weak preference relation $\succeq$, an *indifference relation* $\sim$ is defined by $x \sim y$ iff $x \succeq y$ and $y \succeq x$. One easily verifies that an indifference relation is an equivalence relation. Note that we also have $x \succ y$ iff $x \succeq y$ and $x \nsim y$ (see also Exercise 2.6).

Sometimes it is notationally convenient to use reverse preference relations. So, instead of $x \succ y$ we also write $y \prec x$ and likewise we use $y \preceq x$ for $x \succeq y$. Let us also introduce various types of *preference intervals*. The first are $((x, \rightarrow)) := \{y \in \mathcal{X} : y \succ x\}$ and $((\leftarrow, x)) := \{y \in \mathcal{X} : x \succ y\}$. We will also use the notation $((x, y))$ for $((\leftarrow, y)) \cap ((x, \rightarrow))$. Likewise we define $[[x, \rightarrow)) = \{y \in \mathcal{X} : y \succeq x\}$, $((\leftarrow, x]] = \{y \in \mathcal{X} : x \succeq y\}$ etc. In terms of preference intervals, negative transitivity of $\succ$ can be casted as $((\leftarrow, x)) \cup ((y, \rightarrow)) = \mathcal{X}$ if $x \succ y$.

2.2 Numerical representations

In a rather general setting abstract preference orders can be replaced with equivalent *numerical representations* for which the usual order on $\mathbb{R}$ can be used.

Definition 2.4 A function $U : \mathcal{X} \to \mathbb{R}$ is called a *numerical representation* of a preference relation $\succ$, if $x \succ y$ is equivalent to $U(x) > U(y)$.

An alternative definition of a numerical representation is obtained by putting

$$x \succeq y \text{ iff } U(x) \geq U(y) \tag{2.1}$$

instead of the equivalence in Definition 2.4. Of course, any strictly increasing transformation of a numerical representation U yields another numerical representation, so numerical representations are inherently non-unique. Numerical representations and the underlying preference relations having certain additional properties can be casted in terms of the so-called *utility functions*. We'll come back to this in Section 4.

Definition 2.5 Given a preference relation $\succ$ on $\mathcal{X}$, a subset $\mathcal{Z}$ of $\mathcal{X}$ is called *order dense* (in $\mathcal{X}$) if for all $x, y \in \mathcal{X}$ with $x \succ y$, there exists a $z \in \mathcal{Z}$ such that $x \succeq z \succeq y$.

Theorem 2.6 *A given preference relation $\succ$ on $\mathcal{X}$ admits a numerical representation iff $\mathcal{X}$ contains a countable order dense subset.*

Proof Let $\mathcal{Z}$ be a countable order dense subset of $\mathcal{X}$. Choose a probability measure μ on $\mathcal{Z}$ with $\mu(z) = \mu(\{z\}) > 0$ for all $z \in \mathcal{Z}$. Then we put

$$U(x) := \sum_{z : x \succ z} \mu(z) - \sum_{z : z \succ x} \mu(z). \tag{2.2}$$

Notice that existence of such a probability distribution is guaranteed by countability of $\mathcal{Z}$, that also makes $U(x)$ well defined in terms of the given summations. By construction we have $x \succ y$ iff $U(x) > U(y)$. To see this, we first compute for $x \succeq y$ the difference

$$U(x) - U(y) = \sum_{z : x \succ z \succeq y} \mu(z) + \sum_{z : x \succeq z \succ y} \mu(z).$$

If $x \succ y$, then there is $z_0 \in \mathcal{Z}$ such that $x \succeq z_0 \succeq y$. By negative transitivity we also have $z_0 \succ y$ or $x \succ z_0$. Hence we have $x \succ z_0 \succeq y$ or $x \succeq z_0 \succ y$, and we see that at least one of the two sums in the display is strictly positive, which yields $U(x) > U(y)$. If $x \succeq y$, the right hand side of the displayed formula is still well defined, has nonnegative terms (possibly zero) and hence $U(x) \geq U(y)$. It then follows by contraposition that $U(x) > U(y)$ implies $x \succ y$. We conclude that U as in (2.2) is a numerical representation of $\succ$.

Conversely, we assume that a numerical representation is given. We also assume that $\mathcal{X}$ is uncountable, otherwise there is nothing to prove. Let $\mathcal{J} := \{[a, b] : a, b \in \mathbb{Q}, a < b, U^{-1}([a, b]) \neq \emptyset\}$. Then, for every $I \in \mathcal{J}$, there exists $z_I \in \mathcal{X}$ with $U(z_I) \in I$. Put $A := \{z_I : I \in \mathcal{J}\}$ and observe that A is countable. The set A is almost the set $\mathcal{Z}$ we are after. A naive approach could be as follows. Suppose $y \succ x$, then $U(y) > U(x)$ and there are rational a and b such that $U(x) < a < b < U(y)$. The problem arises that it is not guaranteed that $U^{-1}([a, b])$ is non-void.

To remedy this, we will enlarge the set A with certain elements of A^c and consider thereto first the set $C := \{(x, y) \in A^c \times A^c : y \succ x \text{ and } \forall z \in A : x \succeq z \text{ or } z \succeq y\}$. Hence, in terms of preference intervals, $A \subset ((\leftarrow, x]] \cup [[y, \rightarrow))$. Let $(x, y) \in C$, but suppose that there exists $z \in \mathcal{X} \setminus A$ such that $y \succ z \succ x$. Then we can also find rational a and b such that $U(x) < a < U(z) < b < U(y)$ and therefore $I := [a, b] \in \mathcal{J}$. By definition of A, we can then find $z_I \in A$ that then also has the property $U(x) < a \leq U(z_I) \leq b < U(y)$ and hence $y \succ z_I \succ x$. This contradicts $(x, y) \in C$. We conclude that if $(x, y) \in C$, then for all $z \in \mathcal{X}$ it holds that $x \succeq z$ or $z \succeq y$. So, in terms of preference intervals, $\mathcal{X} = ((\leftarrow, x]] \cup [[y, \rightarrow))$ for $(x, y) \in C$.

This implies the following observation. If $(x, y) \in C$ and $(x', y') \in C$, such that $U(x) \neq U(x')$ or $U(y) \neq U(y')$, then $(U(x), U(y)) \cap (U(x'), U(y')) = \emptyset$. We argue as follows. The situation $x \sim x'$ and $y \sim y'$ is ruled out by assumption. Therefore assume w.l.o.g. that $x \nsim x'$. Since $(x, y) \in C$, we must have $x \succeq x'$ or $x' \succeq y$, which implies that either $U(x) \geq U(x')$ or $U(x') \geq U(y)$. In the latter case we are done. Let then the former inequality hold. Since also $(x', y') \in C$, we have $x' \succeq x$ or $x \succeq y'$. The first of these possibilities can not happen, since we ruled out $x \sim x'$, and therefore the second one holds, and we obtain $U(x) \geq U(y')$, from which the conclusion follows as well.

Knowing that the intervals of the type $(U(x), U(y))$ with $(x, y) \in C$ are disjoint, we conclude that there are only countably many of them and it follows that the collection of these intervals can be written as a collection of intervals $(U(x), U(y))$, where x and y run through a countable subset of $\mathcal{X}$, B say.

We put $\mathcal{Z} = A \cup B$, a countable set as well, and we will see that it is order dense. Take $x, y \in \mathcal{X} \setminus \mathcal{Z}$ with $y \succ x$. If there is $z \in A$ such that $y \succ z \succ x$, we are done. If such a z doesn't exist, then $(x, y) \in C$, in which case we have for instance $U(x) = U(z)$ for some $z \in B$. But then $y \succ z \succeq x$. □

Not every preference order admits a numerical representation. It turns out that the *lexicographical order* on $[0, 1] \times [0, 1]$ provides us with a counterexample. This order is defined by $x \succ y$ iff $x_1 > y_1$ or $x_1 = y_1$ and $x_2 > y_2$.

Example 2.7 Let $\mathcal{X} = [0,1] \times [0,1]$ endowed with the lexicographical order $\succ$. Suppose that $\succ$ admits a numerical representation U. Since $(\alpha, 1) \succ (\alpha, 0)$, we have $d(\alpha) := U(\alpha, 1) - U(\alpha, 0) > 0$ for all $\alpha \in [0,1]$, and hence $[0,1] = \cup_n A_n$, where $A_n = \{\alpha \in [0,1] : d(\alpha) > \frac{1}{n}\}$. Since $[0,1]$ is uncountable, there must be a set A_m with infinitely many elements. In this set we can choose for any positive integer N real numbers $\alpha_0 < \cdots < \alpha_N$. Note that $U(\alpha_{i+1}, 0) > U(\alpha_i, 1)$, and so we get $U(\alpha_{i+1}, 0) - U(\alpha_i, 0) > d(\alpha_i) > \frac{1}{m}$. Hence we get

$$\begin{aligned} U(1,1) - U(0,0) &= U(1,1) - U(\alpha_N, 0) \\ &\quad + \sum_{i=0}^{N-1} (U(\alpha_{i+1}, 0) - U(\alpha_i, 0)) \\ &\quad + U(\alpha_0, 0) - U(0,0) \\ &> \frac{N}{m}. \end{aligned}$$

Letting $N \to \infty$ yields $U(1,1) - U(0,0) = \infty$, which is excluded.

Definition 2.8 Let $\mathcal{X}$ be a topological space. A preference relation $\succ$ is called *continuous* if for every $x \in \mathcal{X}$ the sets $((x, \to))$ and $((\leftarrow, x))$ are open.

Remark 2.9 Suppose that $\succ$ admits a numerical representation U. Because of the identity $((x, \to)) = U^{-1}(U(x), \infty)$, it follows that $\succ$ is continuous if U is a continuous function. But there are also examples of preference orders that are not continuous. Consider again the lexicographical order on $\mathcal{X} = [0,1] \times [0,1]$. Then $\{(y_1, y_2) \in \mathcal{X} : (y_1, y_2) \succ (\frac{1}{2}, \frac{1}{2})\}$ is not open in the ordinary topology (draw a picture).

Proposition 2.10 *Let $\mathcal{X}$ be a Hausdorff space. On $\mathcal{X} \times \mathcal{X}$ we use the product topology. Then the following are equivalent.*

(i) $\succ$ is continuous.

(ii) The set $\{(x, y) : y \succ x\}$ is open.

(iii) The set $\{(x, y) : y \succeq x\}$ is closed.

Proof First we show (i) $\Rightarrow$ (ii): Let $(x_0, y_0) \in M := \{(x, y) : y \succ x\}$. We show that there are open subsets U and V of $\mathcal{X}$ such that $(x_0, y_0) \in U \times V \subset M$. Suppose first that $((x_0, y_0)) \neq \emptyset$. Pick a z from this preference interval, then $y_0 \succ z \succ x_0$. The sets $U := ((\leftarrow, z))$ and $V := ((z, \to))$ are open and contain x_0 and y_0 respectively. Moreover, one quickly sees that $U \times V \subset M$.

If the preference interval $((x_0, y_0))$ is empty, we choose $U = ((\leftarrow, y_0))$ and $V = ((x_0, \to))$. Take $(x, y) \in U \times V$. Then $y_0 \succ x$ and $y \succ x_0$. To show that $y \succ x$, we assume the contrary. By negative transitivity we must have $y_0 \succ y$. But then $y \in ((x_0, y_0))$, which was empty. Contradiction.

(ii) $\Rightarrow$ (iii): It follows from (ii) that also $\{(x, y) : x \succ y\}$ is open. But its complement is just $\{(x, y) : y \succeq x\}$.

(iii) $\Rightarrow$ (i): Since $\mathcal{X}$ is Hausdorff, every singleton $\{x\}$ is closed and so $\{x\} \times \mathcal{X}$ is closed in the product topology. By assumption, then also $\{x\} \times \{y : y \succeq x\} =$

$\{x\} \times \mathcal{X} \cap \{(u, y) : y \succeq u\}$ is closed. But then, the set $\{y : y \succeq x\}$ is closed in $\mathcal{X}$ since a product set is closed iff all factors are closed, and so $\{y : x \succ y\}$ is open. In a similar way on proves that $\{y : y \succ x\}$ is open. □

Proposition 2.11 *Let $\mathcal{X}$ be a connected topological space endowed with a continuous preference order $\succ$. Then every dense set $\mathcal{Z}$ of $\mathcal{X}$ is also order dense. If $\mathcal{X}$ is separable, then there exists a numerical representation of $\succ$.*

Proof First we rule out the trivial situation in which all elements of $\mathcal{X}$ are indifferent. So, we can take $x, y \in \mathcal{X}$ with $y \succ x$. Observe that $y \in ((x, \rightarrow))$ and $x \in ((\leftarrow, y))$, so both open preference intervals are non-empty. Moreover, their union is $\mathcal{X}$, because of negative transitivity. Then we must have that $((x, \rightarrow)) \cap ((\leftarrow, y)) \neq \emptyset$, because $\mathcal{X}$ is connected. The intersection is open as well, so it must contain a z from $\mathcal{Z}$, since $\mathcal{Z}$ is dense. Then $y \succ z \succ x$, and so $\mathcal{Z}$ is order dense. If $\mathcal{X}$ is separable, there exists a countable dense and thus order dense subset. Apply Theorem 2.6. □

Connectedness is essential in Proposition 2.11. Here is a counterexample. Let x_0, y_0 be irrational numbers in $\mathbb{R}$ with $x_0 < y_0$. Let $\mathcal{X} = (-\infty, x_0] \cup [y_0, \infty)$ and let $\succ$ be the usual order on $\mathbb{R}$. Obviously $\mathbb{Q} \cap \mathcal{X}$ is dense in $\mathcal{X}$, but not order dense, since $[x_0, y_0] := \{x \in \mathcal{X} : x_0 \leq x \leq y_0\}$ (by definition a subset of $\mathcal{X}$) contains no rational numbers.

The assertion of Proposition 2.11, that every continuous preference relation on a connected topological space has a numerical representation, can be sharpened.

Theorem 2.12 *Let $\mathcal{X}$ be a connected and separable topological space, endowed with a continuous preference order. Then this preference order admits a* continuous *numerical representation.*

Proof We rule out the trivial case that $x \sim y$ for all $x, y \in \mathcal{X}$. Let $\mathcal{Z}$ be a countable dense subset in $\mathcal{X}$, write $\mathcal{Z} = \{z_1, z_2, \ldots\}$. We will first construct a representation U_0 of $\succ$ restricted to $\mathcal{Z}$, and then give it a continuous extension U on $\mathcal{X}$. The construction will be recursive by 'filling the holes' and bears some similarity with the construction of the Cantor function.

We define $U_0(z_1) := \frac{1}{2}$. Consider z_2. Three possibilities arise. If $z_2 \sim z_1$, then $U_0(z_2) = U_0(z_1)$. If $z_1 \succ z_2$, then $U_0(z_2) := \frac{1}{4}$ and if $z_2 \succ z_1$ we put $U_0(z_2) = \frac{3}{4}$. For later use we define $V_n = \{k2^{-n} : k = 1, \ldots, 2^n - 1\}$. Notice that $U_0(z_1) \in V_1$ and $U_0(z_2) \in V_2$, whatever z_2, and that $V_n \subset V_{n+1}$ for all n.

We now give the general pattern. Supposing that $U_0(z_1), \ldots, U_0(z_n)$ $(n \geq 2)$ have been defined and that $U_0(z_k) \in V_k \subset V_n$, for $k \leq n$. We now define $U_0(z_{n+1})$ after realizing that only four different situations can arise.

Case 1: z_{n+1} is indifferent to some z_i with $i \leq n$. Then $U_0(z_{n+1}) := U_0(z_i) \in V_n \subset V_{n+1}$.

Case 2: $z_i \succ z_{n+1}$ for all $i \leq n$. Then $U_0(z_{n+1}) = \frac{1}{2} \min\{U_0(z_i) : i = 1, \ldots, n\} \in V_{n+1}$.

Case 3: $z_{n+1} \succ z_i$ for all $i \leq n$. Then $U_0(z_{n+1}) = \frac{1}{2}(\max\{U_0(z_i) : i = 1, \ldots, n\} + 1) \in V_{n+1}$.

Case 4: There exist $i_n, j_n \in \{1, \ldots, n\}$ such that for all other $i \in \{1, \ldots, n\}$ it holds that either $z_{i_n} \succeq z_i$ or $z_i \succeq z_{j_n}$ and $z_{j_n} \succ z_{n+1} \succ z_{i_n}$. In this case we put $U_0(z_{n+1}) = \frac{1}{2}(U_0(z_{i_n}) + U_0(z_{j_n})) \in V_{n+1}$.

Let $V = \cup_n V_n$, $u_0 = \inf U_0(\mathcal{Z})$ and $u_1 = \sup U_0(\mathcal{Z})$. Notice that $u_0 < u_1$. Obviously, $U_0(\mathcal{Z}) \subset V$, but we claim that also $V \cap (u_0, u_1) \subset U_0(\mathcal{Z})$. To see that this holds true, we argue as follows.

We proceed under the assumption that u_0 and u_1 are not attained, in fact then $u_0 = 0$ and $u_1 = 1$. Alternatively, if u_0 and/or u_1 are attained, then one proceeds further by similar arguments to those below.

So, we continue under the assumption made, u_0 and u_1 are not attained. We have $U_0(z_1) = \frac{1}{2}$. The sets $((\leftarrow, z_1))$ and $((z_1, \rightarrow))$ are both open and non-empty, otherwise we would have $u_0 = U_0(z_1)$ or $u_1 = U_0(z_1)$, contradicting our assumption that u_0 and u_1 are not attained. Hence there must be z and z' in $\mathcal{Z}$ such that $U_0(z) = \frac{1}{4}$ and $U_0(z') = \frac{3}{4}$ Hence all values in $V_2 \cap (u_0, u_1)$ are attained. We continue by induction. Suppose that all values in $V_n \cap (u_0, u_1)$ are attained. Then for all $1 \leq k < 2^n - 1$, there are $z, z' \in \mathcal{Z}$ such that $U_0(z) = k2^{-n}$ and $U_0(z') = (k+1)2^{-n}$. The set $((z, z'))$ is open and, by the fact that $\mathcal{X}$ is connected, not empty. Since $\mathcal{Z}$ is dense, there is $z'' \in \mathcal{Z} \cap ((z, z'))$. We can even choose z'' such that $U_0(z'') = (2k+1)2^{-n-1} \in V_{n+1} \setminus V_n$, by the construction of U_0.

If $2^{-n} \in V_n \cap (u_0, u_1)$ and z is such that $U_0(z) = 2^{-n}$, then $2^{-n} > u_0$ (the infimal value) and there must be a z' such that $U_0(z') < 2^{-n}$. By the construction of U_0, we can choose z' even such that $U_0(z') = 2^{-n-1}$. For z such that $U_0(z) = 1 - 2^{-n}$ there is a similar reasoning. This shows that all values in V_{n+1} are attained as well.

Having thus proved the claim, we take closures and it follows that $[u_0, u_1] = \operatorname{Cl} U_0(\mathcal{Z})$, so $U_0(\mathcal{Z})$ is dense in $[u_0, u_1]$. It is obvious that the equivalence (2.1) holds for $U = U_0$ and for all $x \in \mathcal{Z}$, by construction of U_0 on $\mathcal{Z}$, so U_0 is a numerical representation of $\succ$ on $\mathcal{Z}$.

We extend U_0 to have domain $\mathcal{X}$ by setting

$$U(x) = \sup\{U_0(z) : z \in \mathcal{Z}, z \preceq x\}. \tag{2.3}$$

First we check that $U(z) = U_0(z)$, for $z \in \mathcal{Z}$. It is obvious that $U(z) \geq U_0(z)$. Suppose that $U(z) > U_0(z)$. Then, by the definition of $U(z)$ as a supremum, there would be $z' \preceq z$ such that $U_0(z') > U_0(z)$ and then also $z' \succ z$, a contradiction.

Now we show that U is a numerical representation of $\succ$. Suppose that $x \succeq y$. Then $\{U_0(z) : z \in \mathcal{Z}, z \preceq y\} \subset \{U_0(z) : z \in \mathcal{Z}, z \preceq x\}$ and $U(x) \geq U(y)$ obviously holds. Let now $x \succ y$. Then, by $\mathcal{Z}$ dense and $\mathcal{X}$ connected, there are $z, z' \in \mathcal{Z}$ such that $x \succeq z \succ z' \succeq y$ (Exercise 2.3), and hence $U(x) \geq U(z) > U(z') \geq U(y)$.

We finally show that U is continuous, for which it is sufficient to show that the sets of the form $U^{-1}[(-\infty, u)]$ and $U^{-1}[(u, \infty)]$ are open for $u \in [u_0, u_1]$.

We focus on $U^{-1}[(-\infty, u)]$, and we only have to consider $u > u_0$. Take x from this set. By V dense in $[u_0, u_1]$, there is $v \in V$ such that $U(x) < v$ and $v < u$. But $v = U(z)$ for some $z \in \mathcal{Z}$, since all $v \in V \cap (u_0, u_1)$ are attained by U_0 and $U_{|\mathcal{Z}} = U_0$. Hence $x \in ((\leftarrow, z)) \subset U^{-1}[(-\infty, u)]$, and since $((\leftarrow, z))$ is open, also $U^{-1}[(-\infty, u)]$ is open. □

2.3 Exercises

2.1 Prove Proposition 2.3.

2.2 Assume that $\succ$ is a continuous preference relation on a connected set $\mathcal{X}$, which is endowed with a topology that is first-countable (this allows you to work below with sequences). Let $\mathcal{Z}$ be a dense subset of $\mathcal{X}$. If $U : \mathcal{X} \to \mathbb{R}$ is continuous and its restriction to $\mathcal{Z}$ is a numerical representation of $\succ$, then U is also a numerical representation of $\succ$ on all of $\mathcal{X}$. To show this, you verify the following implications.
(a) $x \succ y \Rightarrow U(x) > U(y)$
(b) $U(x) > U(y) \Rightarrow x \succ y$.

Hints: To show (a) you complete the following steps. Show that there are $z, w \in \mathcal{Z}$ such that $x \succ z \succ w \succ y$. Choose then $z_n, w_n \in \mathcal{Z}$ such that $z_n \to x$ and $w_n \to y$ and finish the proof.
For (b) you show first that $U^{-1}(U(y), \infty) \cap U^{-1}(-\infty, U(x))$ is non-void and select $z, w \in \mathcal{Z}$ such that $U(x) > U(z) > U(w) > U(y)$. Use again convergent sequences.

2.3 Let $\mathcal{X}$ be a connected topological space and $\succ$ a continuous strict preference relation on it.
(a) Let $x \succ y$. Show that $\mathcal{X} = ((y, \rightarrow))^c \cup ((y, x)) \cup ((\leftarrow, x))^c$.
(b) Show that $((y, x))$ is not empty.
(c) Let $\mathcal{Z}$ be dense in $\mathcal{X}$. Show that there are $z, z' \in \mathcal{Z}$ such that $x \succ z \succ z' \succ y$.

2.4 Show that the set $\{z \in \mathcal{Z} : U_0(z) \in (u_0, u_1)\}$ in the proof of Theorem 2.12 is dense in $((z^0, z^1))$.

2.5 Let $\mathcal{X}$ be a topological space with a continuous preference order $\succ$ and topology $\mathcal{T}$. Let $\mathcal{S}$ be the set of all preference intervals $((\leftarrow, x))$ or $((x, \rightarrow))$. Let $\mathcal{T}_0$ be the smallest topology that contains $\mathcal{S}$. Show that $\mathcal{T}_0$ is the coarsest topology for which Proposition 2.11 and Theorem 2.12 are valid.

2.6 Let $\succ$ be a strict preference order on a set $\mathcal{X}$ and $\succeq$ its associated weak preference order. Prove, using the definitions of these preference orders, the following statements.
(a) $\sim$ is an equivalence relation on $\mathcal{X}$.
(b) $x \succ y$ iff $x \succeq y$ and $x \nsim y$.
(c) $x \succeq y$ iff $x \succ y$ or $x \sim y$.

2.7 A subset K of $\mathbb{R}$ is compact in the ordinary topology iff it is closed and bounded, in which case there is a maximal number in K. Here is the counterpart for preference relations. Let $\succeq$ be a continuous weak preference relation on a set $\mathcal{X}$, which is endowed with a topology under which it becomes a Hausdorff space (a compact subset of $\mathcal{X}$ is then closed). Let K be a non-empty compact subset of $\mathcal{X}$. Show that there is a most preferred element $\bar{x}$ in K (i.e. $\bar{x} \succeq x$ for all $x \in K$) and that the set of all such $\bar{x}$ is compact. *Hint:* Consider for each $x \in K$ the preference intervals in K, $[[x, \rightarrow)) \cap K$, and show that any finite intersection of them is non-empty.

2.8 Let $\succeq$ be a continuous weak preference relation on $\mathcal{X} = (0, \infty)^m$ that has the additional property that it is monotonic: if $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ are such that $x_i \geq y_i$ for all i and $x_i \neq y_i$ for some i, then x is strictly preferred over y, $x \succ y$. Let $\mathbf{1}$ be the vector $(1, \ldots, 1)$. Define $U(x) = \sup\{\alpha > 0 : x \succeq \alpha\mathbf{1}\}$, $x \in \mathcal{X}$.

(a) Show that U is a numerical representation of $\succeq$.
(b) Consider for given $x \in \mathcal{X}$ the sets $\mathcal{U} = \{\alpha \geq 0 : \alpha\mathbf{1} \preceq x\}$ and $\mathcal{L} = \{\alpha \geq 0 : \alpha\mathbf{1} \succeq x\}$. Show that (i) $\mathcal{L} \cap \mathcal{U} \neq \emptyset$ and that (ii) actually this intersection consists of one point only.
(c) Define $U'(x) = \inf\{\alpha > 0 : x \preceq \alpha\mathbf{1}\}$, $x \in \mathcal{X}$. Show that $U = U'$.

2.9 Let $\mathcal{X} = [0, 1] \times [0, 1]$ endowed with the lexicographical order $\succ$. Show by a direct argument, not referring to Theorem 2.6, that $\mathcal{X}$ has no countable order dense subset.

2.10 Consider the function U as in the proof of Theorem 2.12, whose assumptions are in force. Define $V(x) = \inf\{U_0(z) : z \in \mathcal{Z}, z \succeq x\}$. Show that $U(x) \leq V(x)$, for all $x \in \mathcal{X}$. Is $U = V$?

2.11 Let $\mathcal{X} = \mathbb{R}^2_+$, endowed with the ordinary topology and with elements $x = (x_1, x_2)$ etc. Define $x \succeq y$ if $x_1 x_2 \geq y_1 y_2$.

(a) Show that $\succeq$ defines a weak preference relation. How does the corresponding strict preference relation look? Is it continuous?
(b) There exists an obvious numerical representation of $\succeq$. Which one? Is this representation continuous?
(c) Give a countable order dense subset.
(d) What are the points in $\mathcal{X}$ that are indifferent of $(1, 1)$? Sketch a few indifference curves.

2.12 Investigate whether the following alternatives to define the function U as in (2.3) lead to the same function.

(a) $U(x) = \sup\{U_0(z) : z \in \mathcal{Z}, z \prec x\}$.
(b) $U(x) = \inf\{U_0(z) : z \in \mathcal{Z}, z \succeq x\}$.
(c) $U(x) = \inf\{U_0(z) : z \in \mathcal{Z}, z \succ x\}$.

3 Lotteries

In this section we take the situation of the previous section as a starting point. We assume that the set $\mathcal{X}$ is a convex subset of the set of all probability measures on some measurable space $(S, \mathcal{S})$. We write $\mathcal{M}$ instead of $\mathcal{X}$. Every probability measure can be considered as a *lottery*, in the 'argot du métier'. Our aim is to consider preference relations on the space of lotteries that admit a numerical representation of a special kind.

3.1 Von Neumann-Morgenstern representations

Definition 3.1 Let $\succ$ be a preference relation on $\mathcal{M}$. A numerical representation U is called a *Von Neumann-Morgenstern* representation if there is a measurable function $u : S \to \mathbb{R}$ such that

$$U(\mu) = \int u \, \mathrm{d}\mu, \forall \mu \in \mathcal{M}. \tag{3.1}$$

Not every preference relation admits a Von Neumann-Morgenstern representation, see Exercises 3.3 and 3.7. But if it does, it is easy to check that a Von Neumann-Morgenstern representation U is an *affine* function, i.e. $U(t\mu + (1-t)\nu) = tU(\mu) + (1-t)U(\nu)$, for all $\mu, \nu \in \mathcal{M}$ and $t \in [0, 1]$. But, if a numerical representation U of $\succ$ is affine, then it implies two additional properties of $\succ$ (see Proposition 3.3), that we define now.

Definition 3.2 Let $\succ$ be a preference relation on $\mathcal{M}$. It satisfies the *independence axiom* if for all $\mu \succ \nu$ it holds that

$$t\mu + (1-t)\lambda \succ t\nu + (1-t)\lambda, \tag{3.2}$$

for all $\lambda \in \mathcal{M}$ and $t \in (0, 1]$.

The preference relation satisfies the *Archimedean axiom* (also called continuity axiom), if for all $\mu \succ \lambda \succ \nu$, there are $t, s \in (0, 1)$ such that

$$t\mu + (1-t)\nu \succ \lambda \succ s\mu + (1-s)\nu. \tag{3.3}$$

Proposition 3.3 *Assume that $\succ$ admits an affine numerical representation. Then $\succ$ satisfies the axioms of Definition 3.2.*

Proof Exercise 3.1. □

The nice thing is that Proposition 3.3 has a converse as well. This is the content of the next theorem.

Theorem 3.4 *Suppose that a preference relation $\succ$ on $\mathcal{M}$ satisfies both the independence and Archimedean axioms. Then it has an affine numerical representation, U say. Moreover, for any other affine numerical representation $\tilde{U}$, there exist $a > 0$ and $b \in \mathbb{R}$ such that $\tilde{U} = aU + b$.*

To prepare for the proof of this theorem, we present a lemma. In the proof we use a couple of times a consequence of the independence axiom, by taking $\lambda = \mu$ or $\lambda = \nu$. If $\mu \succ \nu$, then $\mu \succ t\mu + (1-t)\nu$ for all $t \in [0,1)$ and $t\mu + (1-t)\nu \succ \nu$ for all $t \in (0,1]$.

Lemma 3.5 *Suppose that a preference relation $\succ$ on $\mathcal{M}$ satisfies both the independence and Archimedean axioms. Then the following hold true.*

(i) If $\mu \succ \nu$, then for all $0 \le t < s \le 1$ it holds that $s\mu + (1-s)\nu \succ t\mu + (1-t)\nu$.

(ii) If $\mu \succ \nu$ and $\lambda \in [[\nu, \mu]]$, then there exists a unique $t \in [0,1]$ such that $\lambda \sim t\mu + (1-t)\nu$.

(iii) If $\mu \sim \nu$, then $t\mu + (1-t)\lambda \sim t\nu + (1-t)\lambda$, for all $\lambda \in \mathcal{M}$ and $t \in [0,1]$.

(iv) For any $\rho \succeq \lambda$, the preference interval $[[\lambda, \rho]]$ is convex.

Proof (i) We use (3.2), with s instead of t and $\lambda = \nu$, to get $\rho := s\mu + (1-s)\nu \succ \nu$. Using (3.2) again, with $\rho = \mu = \lambda$ and $1-u$ instead of t, we get $(1-u)\rho + u\rho \succ (1-u)\nu + u\rho$. Here the left hand side is ρ, whereas the right hand side equals $su\mu + (1-su)\nu$. Take $u = t/s \in (0,1)$.

(ii) We only have to show existence, uniqueness follows from (i). Existence for the cases $\lambda \sim \mu$ and $\lambda \sim \nu$ is trivial, so we assume $\mu \succ \lambda \succ \nu$. In view of (i), it should hold that $t = \sup A$, where $A = \{u : \lambda \succ u\mu + (1-u)\nu\}$. Note that $A \neq \emptyset$, as $0 \in A$.

Suppose that this t is not the right one, then either $\lambda \succ t\mu + (1-t)\nu$ or $\lambda \prec t\mu + (1-t)\nu$. Consider the first case, which rules out $t = 1$, so $t < 1$. We use the right hand side of (3.3) applied to the triple $\mu \succ \lambda \succ t\mu + (1-t)\nu$ to get existence of $s \in (0,1)$ such that $\lambda \succ s(t\mu + (1-t)\nu) + (1-s)\mu = (1-s+ts)\mu + (1-t)s\nu$. The definition of t implies $t \ge 1 - s + ts$, which is, since $s < 1$, equivalent to $t \ge 1$, a contradiction.

In the second of the two above cases, we apply (3.3) to the triple $t\mu + (1-t)\nu \succ \lambda \succ \nu$, to get $s \in (0,1)$ such that $st\mu + (1-st)\nu = s(t\mu + (1-t)\nu) + (1-s)\nu \succ \lambda$. This means that $st \notin A$, and hence $st \ge \sup A = t$, so $s \ge 1$, another contradiction.

(iii) We rule out the case that *all* $\rho \in \mathcal{M}$ are equivalent to μ, because then we immediately have $t\mu + (1-t)\lambda \sim \mu$, $t\nu + (1-t)\lambda \sim \mu$ and we are done. So, take $\rho \nsim \mu$ and suppose that $\rho \succ \mu$ (the other case can be treated similarly). Then also $\rho \succ \nu$ and we apply (3.2) to obtain $s\rho + (1-s)\nu \succ s\nu + (1-s)\nu = \nu$, for all $s \in (0,1)$. Then we apply (3.2) again to get

$$t(s\rho + (1-s)\nu) + (1-t)\lambda \succ t\mu + (1-t)\lambda. \tag{3.4}$$

If the assertion were not true, then we have for instance $t\mu + (1-t)\lambda \succ t\nu + (1-t)\lambda$ for some $t \in (0,1)$ (the other possibility can be treated similarly). So, assume that $t\mu + (1-t)\lambda \succ t\nu + (1-t)\lambda$, contrary to what we have to show. With Equation (3.4) it then follows that $t\mu + (1-t)\lambda \in [[t\nu + (1-t)\lambda, t(s\rho + (1-s)\nu) + (1-t)\lambda]]$. We can now apply (ii), which yields a unique $u \in (0,1)$ such that $t\mu + (1-t)\lambda \sim u(t(s\rho + (1-s)\nu) + (1-t)\lambda) + (1-u)(t\nu + (1-t)\lambda) = tsu\rho + t(1-su)\nu + (1-t)\lambda$. Equation (3.4) is true for all $s \in (0,1)$, and so we can there replace s with su, which yields $tsu\rho + t(1-su)\nu + (1-t)\lambda \succ t\mu + (1-t)\lambda$. This contradicts the last obtained indifference relation, a contradiction, caused by the assumption $t\mu + (1-t)\lambda \succ t\nu + (1-t)\lambda$. Likewise, one can eliminate $t\mu + (1-t)\lambda \prec t\nu + (1-t)\lambda$ to complete the proof.

(iv) Next we show that $[[\lambda, \rho]]$ is convex. We assume $\rho \succ \lambda$, since the case $\rho \sim \lambda$ immediately follows from part (ii).

Consider first the interior case, we take $\mu, \nu \in ((\lambda, \rho))$ and $t \in (0, 1)$. Since $\rho \succ \nu$ we use the independence axiom to get $\rho = t\rho + (1-t)\rho \succ t\rho + (1-t)\nu$. And since $\rho \succ \mu$, we use the same axiom to get $t\rho + (1-t)\nu \succ t\mu + (1-t)\nu$. Combining these relations, we obtain $\rho \succ t\mu + (1-t)\nu$. One similarly proves $t\mu + (1-t)\nu \succ \lambda$.

Next we consider a boundary case $\mu \sim \rho$, $\nu \in ((\lambda, \rho))$. Inspection of the proof of the previous case shows that this case is partly handled, and one can use part (iii) of this lemma to complete the proof (you check!). Finally we have the extreme case $\nu \sim \lambda$ and $\mu \sim \rho$. Here one can use part (iii) again (Exercise 3.6). We conclude that $[[\lambda, \rho]]$ is convex. □

Proof of Theorem 3.4 We exclude the trivial case in which all elements of $\mathcal{M}$ are indifferent. Choose $\rho, \lambda \in \mathcal{M}$ with $\rho \succ \lambda$. Let $\mu \in [[\lambda, \rho]]$. Lemma 3.5(ii) yields a unique $t = t(\mu)$ such that $\mu \sim t\rho + (1-t)\lambda$. We use this to define U on $[[\lambda, \rho]]$ by $U(\mu) := t$. So, $U(\mu)$ is the coefficient of ρ in the representation $\mu \sim t\rho + (1-t)\lambda$. Notice that $U(\lambda) = 0$ and $U(\rho) = 1$.

The first thing to show is that we have defined a numerical representation of $\succ$ on $[[\lambda, \rho]]$. Let $U(\mu) > U(\nu)$. In view of Lemma 3.5(i) we have $U(\mu)\rho + (1-U(\mu))\lambda \succ U(\nu)\rho + (1-U(\nu))\lambda$. But the probability measures on both sides are indifferent to μ and ν respectively. Hence $\mu \succ \nu$. To prove the converse implication it is sufficient to show that $U(\mu) = U(\nu)$ implies $\mu \sim \nu$. But this is obvious from the definition of U.

We now show that U is affine. Since $[[\lambda, \rho]]$ is convex (Lemma 3.5(iv)), $U(t\mu + (1-t)\nu)$ is well defined for $\mu, \nu \in [[\lambda, \rho]]$. Recall that $U(\mu)\rho + (1-U(\mu))\lambda \sim \mu$ and $U(\nu)\rho + (1-U(\nu))\lambda \sim \nu$. A double application of Lemma 3.5(iii) gives

$$t\mu + (1-t)\nu \sim t(U(\mu)\rho + (1-U(\mu))\lambda) + (1-t)(U(\nu)\rho + (1-U(\nu))\lambda).$$

Rearranging terms on the right hand side gives $(tU(\mu) + (1-t)U(\nu))\rho + (1 - tU(\mu) - (1-t)U(\nu))\lambda$, a convex combination of ρ and λ. But then, by definition of U, we have $U(t\mu + (1-t)\nu) = tU(\mu) + (1-t)U(\nu)$, as desired.

The next step is to show that U is unique up to an affine transformation. Let $\tilde{U}$ be another affine representation of $\succ$. Let $\mu \in [[\lambda, \rho]]$ and define

$$\hat{U}(\mu) = \frac{\tilde{U}(\mu) - \tilde{U}(\lambda)}{\tilde{U}(\rho) - \tilde{U}(\lambda)},$$

an affine transformation of $\tilde{U}$, having, like U, the properties $\hat{U}(\lambda) = 0$ and $\hat{U}(\rho) = 1$. Combine this with affinity of $\hat{U}$ to obtain $\hat{U}(\mu) = \hat{U}(U(\mu)\rho + (1 - U(\mu))\lambda) = U(\mu)\hat{U}(\rho) + (1 - U(\mu))\hat{U}(\lambda) = U(\mu)$. Therefore U is an affine transformation of $\tilde{U}$.

The last step is to show that U can be extended to all of $\mathcal{M}$. Consider a preference interval $[[\lambda_1, \rho_1]] \supset [[\lambda, \rho]]$. We know that $\succ$ has an affine representation U_1 on $[[\lambda_1, \rho_1]]$, which can be taken such that $U_1(\rho) = 1$, $U_1(\lambda) = 0$ (apply an affine transformation to accomplish this). So U_1 must coincide with U on $[[\lambda, \rho]]$, as there can be only one affine numerical representation that is zero at λ and one at ρ.

But then U can be extended to all of $\mathcal{M}$. Indeed, if $[[\lambda, \rho]]^c \neq \emptyset$ and $\mu \notin [[\lambda, \rho]]$, then e.g. $\mu \prec \lambda$ and we can take $\lambda_1 = \mu$ and $[[\lambda_1, \rho_1]] = [[\mu, \rho]] \supset [[\lambda, \rho]]$ and apply the previous step. □

Remark 3.6 We note that up to here, we didn't use that $\mathcal{M}$ is a set of probability measure. The above results are valid, under the stated assumptions, for any convex set $\mathcal{M}$. On the other hand, for $\mathcal{M}$ a set of probability measures, convex combinations of the type $t\mu + (1-t)\nu$ have a nice interpretation as a compound lottery. The results that follow use essential properties of probability measures.

We return to the Von Neumann-Morgenstern representation of a preference order $\succ$ on $\mathcal{M}$. We first treat a simple case.

Example 3.7 Suppose that $\mathcal{M}$ is the set of all finite mixtures of Dirac measures δ_x, and that an affine representation U exists. Define $u(x) = U(\delta_x)$. Let $\mu = \sum t_i \delta_{x_i}$, where the $t_i \geq 0$ and $\sum t_i = 1$. Affinity of U yields $U(\mu) = \sum t_i u(x_i) = \int u \, \mathrm{d}\mu$, which is the desired representation. So, in this case, if there exists an affine representation, it is automatically of Von Neumann-Morgenstern type.

In the remainder of this section we assume that the set S is a separable metric space and that $\mathcal{S}$ is its Borel σ-algebra. Recall the definition of weak convergence of probability measures on $\mathcal{S}$: $\mu_n \to \mu$ iff $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$ for all bounded and continuous functions f on S. As a preparation for the final theorem, we have the following lemma.

Lemma 3.8 *Consider the space $\mathcal{M}$ of all probability measures on $(S, \mathcal{S})$ endowed with the weak topology. Fix $\mu, \nu \in \mathcal{M}$ and consider $A : t \to t\mu + (1-t)\nu$. Then $A : [0, 1] \to \mathcal{M}$ is continuous. If $\succ$ is a continuous preference ordering on $\mathcal{M}$, then it satisfies the Archimedean axiom.*

Proof The first assertion follows from the evident identity $\int f \, \mathrm{d}(t\mu + (1-t)\nu) = t \int f \, \mathrm{d}\mu + (1-t) \int f \, \mathrm{d}\nu$, valid for any bounded and continuous function f on S. Indeed, if $t_n \to t$, one then has for all bounded and continuous functions f on S that $\int f \, \mathrm{d}A(t_n) \to \int f \, \mathrm{d}A(t)$, which shows that $A(t)$ is the weak limit of the $A(t_n)$.

To prove the second assertion, let $\mu \succ \nu$ and choose $\lambda \in ((\nu, \mu))$. Observe that $t = 1$ is an element of $A^{-1}((\lambda, \to))$ and that this set is open in $[0, 1]$ by the just shown continuity of A. Hence there is also some $t \in (0, 1)$ belonging to it, and for this t one has $A(t) = t\mu + (1-t)\nu \succ \lambda$, as required in Definition 3.2. The existence of s in that definition follows similarly. □

Theorem 3.9 *Consider the space $\mathcal{M}$ of all probability measures on $(S, \mathcal{S})$ endowed with the weak topology, where S is assumed to be separable. Let $\succ$ be a continuous preference ordering on $\mathcal{M}$, satisfying the independence axiom. Then $\succ$ admits a Von Neumann-Morgenstern representation*

$$U(\mu) = \int u \, \mathrm{d}\mu, \tag{3.5}$$

where the function $u : S \to \mathbb{R}$ is bounded, continuous and unique up to affine transformations.

Proof Consider first the subspace $\mathcal{M}_S$ of simple distributions on S, these are the distributions as in Example 3.7. We conclude from Lemma 3.8 and Theorem 3.4 that $\succ$ restricted to $\mathcal{M}_S$ admits an affine representation, which is, by Example 3.7, automatically of Von Neumann-Morgenstern type.

The function u involved will turn out to be bounded. Suppose that this is not the case, then there is a sequence $(x_n) \subset S$ such that $(u(x_n))$ is increasing and $u(x_n) > n$ (the other possibility $u(x_n) < -n$ can be treated similarly). Put $\mu_n = (1 - \frac{1}{\sqrt{n}})\delta_{x_1} + \frac{1}{\sqrt{n}}\delta_{x_n}$. Since $u(x_2) > u(x_1)$, we have $\delta_{x_2} \succ \delta_{x_1}$, so $\delta_{x_1} \in ((\leftarrow, \delta_{x_2}))$. One easily checks that $\mu_n \to \delta_{x_1}$ weakly. Hence, for n big enough, μ_n belongs to any (nonempty) open neighborhood of δ_{x_1}, so eventually we have $\mu_n \in ((\leftarrow, \delta_{x_2}))$. But then $U(\mu_n) \leq u(x_2)$. However, by direct computation, we have $U(\mu_n) > (1 - \frac{1}{\sqrt{n}})u(x_1) + \sqrt{n}$, which yields a contradiction.

We now show that u is continuous. Suppose the contrary, then there is a sequence (x_n) converging to some $x \in S$, whereas $u(x_n)$ doesn't converge to $u(x)$. Assume e.g. that one has $\limsup u(x_n) < u(x)$. Then along a subsequence, again denoted by (x_n), one has $\lim u(x_n) =: a < u(x)$. In particular, there is $m \in \mathbb{N}$ such that $|u(x_n) - a| < \frac{1}{3}(u(x) - a)$, for $n \geq m$; equivalently $\frac{4}{3}a - \frac{1}{3}u(x) < u(x_n) < \frac{2}{3}a + \frac{1}{3}u(x)$, for $n \geq m$. Put $\mu = \frac{1}{2}(\delta_x + \delta_{x_m})$. Then also $U(\delta_x) = u(x) > \frac{2}{3}u(x) + \frac{1}{3}a > \frac{1}{2}(u(x) + u(x_m)) = U(\mu) > \frac{1}{3}u(x) + \frac{2}{3}a > U(\delta_{x_n})$, for $n \geq m$. So, $\delta_x \succ \mu \succ \delta_{x_n}$. This means that δ_{x_n} doesn't belong to the open neighborhood $((\mu, \rightarrow))$ of δ_x, contradicting the fact that $\delta_{x_n} \to \delta_x$ weakly (which follows from the ordinary convergence $x_n \to x$).

We now show that, knowing the function u, Equation (3.5) defines a numerical representation U of $\succ$. Since u is bounded and continuous, U, as defined in (3.5), is continuous w.r.t. the weak topology. It is a fact that the set of simple distributions is weak-dense in the set of all probability measures on $(S, \mathcal{S})$, see Proposition A.12. Since we know that U is a numerical representation of $\succ$ on the set of simple distributions, we can argue as in the proof of Theorem 2.12, that U is also a numerical representation on the collection of all probability measures on $(S, \mathcal{S})$. See also Exercise 2.2 for an alternative argument.

Finally, u is unique up to affine transformations. This follows from Theorem 3.4, affine numerical representations are unique up to affine transformations, and by the action of the numerical representations on Dirac measures. □

Later on we need representations of preference orders, where u is unbounded. This cannot happen under the conditions of Theorem 3.9. A way out is obtained, by replacing the weak topology with a stronger one. Let ψ be a continuous function on S with $\psi \geq 1$. Let $\mathcal{M}^\psi$ be the set of probability measures μ such that $\int \psi \, \mathrm{d}\mu < \infty$ and let C^ψ be the space of continuous functions f such that $|f|/\psi$ is bounded. If μ is a probability measure in $\mathcal{M}^\psi$, then

$$\mu^\psi(B) := \frac{\int \mathbf{1}_B \psi \, \mathrm{d}\mu}{\int \psi \, \mathrm{d}\mu}$$

defines another probability measure on $(S, \mathcal{S})$. We say that a sequence $(\mu_n) \subset \mathcal{M}^\psi$ converges in the ψ-weak topology if the corresponding sequence (μ_n^ψ) converges in the weak topology. If the ψ-weak limit is μ, then this means nothing else then $\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$, for all $f \in C^\psi$. This immediately yields the following corollary to Theorem 3.9.

Corollary 3.10 *Let $\succ$ be a preference order on $\mathcal{M}^\psi$ that is continuous w.r.t. the ψ-weak topology and that satisfies the independence axiom. Then there exists a Von Neumann-Morgenstern representation of the form (3.5), with $u \in C^\psi$. Also in this case, U and u are unique up to affine transformations.*

Proof The proof is based on the fact that the transformation $\mu \to \mu^\psi$ can be used to apply the results of Theorem 3.9. Details (Exercise 3.2) are left to the reader. □

3.2 Exercises

3.1 Prove Proposition 3.3.

3.2 Prove Corollary 3.10.

3.3 Let S be the set of positive integers, $S = \mathbb{N}$. Let $\mathcal{M}$ be the collection of probability measures μ on $\mathbb{N}$ with the property that $U(\mu) := \lim n^2 \mu(n) < \infty$. Then U is affine and induces a preference order $\succ$ on $\mathcal{M}$.

(a) Show that $\succ$ satisfies both the independence and Archimedean axioms.
(b) Show that $\succ$ does not admit a Von-Neumann-Morgenstern representation.
(c) Why can't we apply Theorem 3.9?

3.4 Finish the proof of Theorem 3.9, i.e. show that U is also a numerical representation on the collection of all probability measures on $(S, \mathcal{S})$. [If you want, you may use Exercise 2.2, and that $\mathcal{M}$ is metrizable. See also Corollary A.13. There is not a lot to do anymore.]

3.5 Suppose you are a plumber and you have a client that wants to pay you 1000 euro for installing a drainage system. If you do nothing and stay home, you don't get paid. Let μ be the sure 'lottery' that pays out 1000 euro with certainty, λ the sure 'lottery' that pays zero. Then for you $\mu \succ \lambda$. Let ν be the 'lottery' in which you will be shot. Then, most likely, $\mu \succ \lambda \succ \nu$. Is there for you a $t \in (0, 1)$ such that $t\mu + (1 - t)\nu \succ \lambda$? Same question if ν is the sure 'lottery' to get killed in a car crash on your way to the client.

3.6 Prove convexity for the remaining cases in the proof of Lemma 3.5(iv).

3.7 Let $\mathcal{M}$ be the set of probability measures on $[0,1]$ with the Borel sets, and denote by λ the Lebesgue measure. According to the Lebesgue decomposition theorem every $\mu \in \mathcal{M}$ can be decomposed as $\mu = \mu_a + \mu_s$, where where μ_s is singular with respect to λ, and μ_a is absolutely continuous. Define a function $U : \mathcal{M} \to [0,1]$ by $U(\mu) = \int_{[0,1]} x \mu_a(\,\mathrm{d}x)$.

(a) Show that U is an affine function on $\mathcal{M}$. Hence, U induces a preference order $\succ$ on $\mathcal{M}$ which satisfies both the Archimedean and the independence axioms.

It is our aim to show that the preference order $\succ$ cannot have a Von Neumann-Morgenstern representation. Assume, to reach a contradiction, that it does exist.

(b) Show $U(\delta_x) = 0$ for all $x \in [0,1]$, and deduce, see Example 3.7, that the only possible choice for u in (3.1) would be $u = 0$.

(c) Show next that $\mu \sim \lambda$ for all $\mu \in \mathcal{M}$, but also $U(\lambda) = \frac{1}{2}$ and $U(\delta_{\frac{1}{2}}) = 0$. Conclude that a Von Neumann-Morgenstern representation of $\succ$ cannot exist.

4 Utility and expected utility

In this section we consider a set $\mathcal{M}$ of probability measures on an interval S of $\mathbb{R}$, and $\mathcal{S}$ will be the Borel σ-algebra on S. We assume that $\mathcal{M}$ is convex and contains all Dirac measures on points in S, and consequently all simple measures.

First some additional remarks about the setting. We depart from the common assumption that the *fair price* of a lottery $\mu \in \mathcal{M}$ equals its *expectation* $m(\mu) := \int_S x\, \mu(\mathrm{d}x)$. We assume, unless the contrary is explicitly stated, that these expectations exist for all $\mu \in \mathcal{M}$ and are finite.

Consider concave functions $u : S \to \mathbb{R}$. These are such that for any $x, y \in S$ and $t \in [0,1]$ it holds that $u(tx + (1-t)y) \geq tu(x) + (1-t)u(y)$. Fix $y = x_0 \in \operatorname{Int} S$. One can then show u is left- and right-differentiable at x_0 with finite left- and right-derivatives $u'_-(x_0)$; and $u'_+(x_0)$ and for all $x \in S$ one has $u(x) \leq (x - x_0)u'_-(x_0) + u(x_0)$ and $u(x) \leq (x - x_0)u'_+(x_0) + u(x_0)$; make a picture to understand this and deduce that in these two inequalities one can replace the derivatives with any constant that is between the derivatives. Hence if $m(\mu)$ is finite for $\mu \in \mathcal{M}$, then $\int u\, \mathrm{d}\mu$ is well defined (but may take the value $-\infty$). We will often need the following lemma.

A function $u : S \to \mathbb{R}$ is called strictly concave if for any $x, y \in S$, $x \neq y$, and $t \in (0,1)$ it holds that $u(tx + (1-t)y) > tu(x) + (1-t)u(y)$. We will often need Jensen's inequality for (strictly) concave functions.

Lemma 4.1 (Jensen's inequality) *Assume that $u : S \to \mathbb{R}$ is concave and $m(\mu)$ and $\int u\, \mathrm{d}\mu$ are finite. Then $\int u\, \mathrm{d}\mu \leq u(m(\mu))$. If, moreover, u is strictly concave and μ is not degenerate (not a Dirac measure), then $\int u\, \mathrm{d}\mu < u(m(\mu))$.*

Proof Concavity of u implies that for any $x_0 \in S$ there exist a, b such that $u(x_0) = ax_0 + b$ and $u(x) \leq ax + b$ for all $x \in S$. If u is strictly concave, the latter inequality is strict for $x \neq x_0$. Take $x_0 = m(\mu)$ and integrate to get $\int u\, \mathrm{d}\mu \leq a\, m(\mu) + b = u(m(\mu))$ for concave u. If u is strictly concave and μ is nondegenerate, we have $\mu(\{x : u(x) < ax + b\}) > 0$ next to $\mu(\{x : u(x) \leq ax + b\}) = 1$ and hence $\int u\, \mathrm{d}\mu < u(m(\mu))$. □

4.1 Risk aversion

It frequently happens (but it depends on the circumstances) that somebody who has the choice between a lottery with an average pay-off of let's say 1000 euro and getting the same amount of money straight away, prefers the latter option. He then exhibits *risk averse* behavior, as a result of his personal preferences. This notion will be made precise in the Definition 4.2 below.

Definition 4.2 A preference order $\succ$ on $\mathcal{M}$ is called *monotone* if the implication $x > y \Rightarrow \delta_x \succ \delta_y$ holds ($x, y \in S$) ('more is better'). It is called *risk averse* if $\delta_{m(\mu)} \succ \mu$, unless μ is degenerate, $\mu = \delta_{m(\mu)}$.

Proposition 4.3 *Suppose that a preference order $\succ$ on $\mathcal{M}$ has Von Neumann-Morgenstern representation*

$$U(\mu) = \int u \, d\mu,$$

for some Borel measurable function u (the integrals are assumed to be well defined for all $\mu \in \mathcal{M}$). Then

(i) the preference order is monotone iff u is strictly increasing and

(ii) the preference order is risk averse iff u is strictly concave.

Proof (i) Notice that $U(\delta_x) = u(x)$. Then $u(x) > u(y)$ iff $U(\delta_x) > U(\delta_y)$ iff $\delta_x \succ \delta_y$.

(ii) Suppose that $\succ$ is risk averse. Take $x, y \in S$ and consider $\mu = t\delta_x + (1-t)\delta_y$ for $t \in (0, 1)$. Then $m(\mu) = tx + (1-t)y$. Then the risk averse $\succ$ yields $U(\delta_{m(\mu)}) > U(\mu)$, or $u(tx + (1-t)y) > tu(x) + (1-t)u(y)$. Hence u is strictly concave. Conversely, for strict concave u Jensen's inequality (Lemma 4.1) gives for any nondegenerate $\mu \in \mathcal{M}$ that $U(\delta_{m(\mu)}) = u(m(\mu)) > \int u \, d\mu = U(\mu)$. □

The function u in the Von Neumann-Morgenstern representation of a monotone risk averse preference relation deserves a name of its own.

Definition 4.4 A function $u : S \to \mathbb{R}$ is called a *utility function* if u is strictly increasing, strictly concave and continuous on S.

Since any concave function is continuous on the interior of the set on which it is defined, Exercise 4.9, the continuity requirement above only concerns boundary points of S. And since u is increasing, in fact it is only a condition of continuity of u at $\inf S$, if this is an element of S.

Definition 4.5 A preference order $\succ$ on $\mathcal{M}$ admits an *expected utility representation* U if there exists a utility function u such that $U(\mu) = \int u \, d\mu$, for all $\mu \in \mathcal{M}$.

In the remainder of this section, we assume that preference orders admit expected utility representations.

Continuity of a utility function u implies (Exercise 4.12) that $u(S)$ is a connected subset of $\mathbb{R}$, an interval, and hence for all $\mu \in \mathcal{M}$ there exists a number $c(\mu) \in S$ such that $u(c(\mu)) = U(\mu) \in u(S)$. Moreover, $c(\mu)$ is unique, because u is strictly increasing. Whence the indifferent relation $\delta_{c(\mu)} \sim \mu$. In words, playing a lottery μ is indifferent to obtaining the sure amount $c(\mu)$ under a given preference ordering.

Definition 4.6 The number $c(\mu)$ is called the *certainty equivalent* of the lottery μ and the difference $\rho(\mu) := m(\mu) - c(\mu)$ is called the *risk premium* of μ.

Notice that always $c(\mu) \leq m(\mu)$ for risk averse $\succ$ and that strict inequality holds for nondegenerate μ. Hence, a risk averse person with utility function u will not pay more than $c(\mu)$ to play a lottery μ. Conversely, the risk premium is the amount of money a seller of the lottery μ has to pay to a risk averse agent to convince him to exchange the sure amount $m(\mu)$ for the random pay-off of the lottery μ.

In the present context, we consider the following optimization problem. Find, if it exists, a lottery μ^* that is most preferred among all lotteries in a subset of $\mathcal{M}$, equivalently, the one with the highest value of U, where U is of expected utility type.

We specialize to a specific case. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be given and a random variable X defined on it, with values in S, that has a nondegenerate distribution μ. Let $c \in \mathbb{R}$ and consider the convex combination $X_\lambda = \lambda c + (1-\lambda)X$. Note that the distribution function of X_λ is obtained by a location-scale transformation of that of X. Write μ_λ for the distribution of X_λ ($\mu_0 = \mu$). Put

$$f(\lambda) := U(\mu_\lambda) = \int u \, \mathrm{d}\mu_\lambda = \mathbb{E}u(X_\lambda).$$

Proposition 4.7 *Assume that S is an interval, $X \geq a$ for some $a \in \mathrm{Int}S$, $\mathbb{E}X < \infty$ and $c \in \mathrm{Int}S$.*

(i) The function $f : [0,1] \to \mathbb{R}$ is strictly concave and hence its maximal value is assumed for some unique $\lambda^* \in [0,1]$.

(ii) We have $\lambda^* = 1$ if $m(\mu) = \mathbb{E}X \leq c$, and $\lambda^* > 0$ if $c \geq c(\mu)$.

(iii) If moreover u is differentiable, then we even have $\lambda^* = 1 \Leftrightarrow \mathbb{E}X \leq c$ and $\lambda^* = 0 \Leftrightarrow c \leq \frac{\mathbb{E}Xu'(X)}{\mathbb{E}u'(X)}$.

Proof (i) Since $f(\lambda) = \mathbb{E}u(X_\lambda)$, strict concavity of f follows from exploiting first strict concavity of u and then taking expectations.

(ii) Jensen's inequality yields

$$f(\lambda) \leq u\big(\mathbb{E}X_\lambda\big) = u\big(\mathbb{E}X + \lambda(c - \mathbb{E}X)\big),$$

with equality iff $\lambda = 1$. Since u is increasing, the right hand side is non-decreasing in λ if $c \geq \mathbb{E}X$. Under this condition, $\lambda^* = 1$.

Concavity of u yields $u(X_\lambda) \geq (1-\lambda)u(X) + \lambda u(c)$, hence

$$f(\lambda) \geq (1-\lambda)u(c(\mu)) + \lambda u(c),$$

with equality iff $\lambda = 0, 1$. The right hand side is non-decreasing in λ under the condition $c \geq c(\mu)$, in which case $\lambda^* > 0$.

(iii) Assume that u is differentiable (and see Exercise 4.13 for a heuristic argument). Because f is concave, $\lambda^* = 0$ can only happen if f is decreasing in a neighborhood of zero, so when the right derivative $f'_+(0) \leq 0$. Let us compute this derivative. We have

$$\frac{u(X_\lambda) - u(X)}{\lambda} = \frac{u(X_\lambda) - u(X)}{X_\lambda - X}(c - X). \tag{4.1}$$

The nonnegative difference quotient on the right of (4.1), is bounded from above by the derivative of u in the left endpoint of the involved interval. Observe that for all $\lambda \in [0, 1]$ one has $X_\lambda = \lambda c + (1 - \lambda)X \geq \lambda c + (1 - \lambda)a \geq \min\{a, c\}$. Hence the absolute value of (4.1) is bounded by $u'_+(c \wedge a)|c - X|$, which has finite expectation, since $u'_+(c \wedge a) < \infty$ because $c, a \in \text{Int}S$ and $\mathbb{E}|X| < \infty$. Taking expectations in (4.1) and letting $\lambda \downarrow 0$, we get by the Dominated Convergence Theorem that the limit is $f'_+(0) = \mathbb{E}u'(X)(c - X)$. Hence $f'_+(0) \leq 0$ iff $c \leq \frac{\mathbb{E}Xu'(X)}{\mathbb{E}u'(X)}$.

In much the same way, $\lambda^* = 1$ iff f is non-decreasing in a neighborhood of $\lambda = 1$, $f'_-(1) \geq 0$, Exercise 4.7. Working with a difference quotient like (4.1) for $\lambda \uparrow 1$ and using that $X_1 = c$, we get $f'_-(1) = u'(c)(c - \mathbb{E}X)$. The last assertion now also follows. □

Example 4.8 Consider a risky asset S_1 with price π_1, and a riskless asset with interest rate r ($S_0 = 1 + r$). Suppose that an agent has a C^1 utility function u and a capital (initial wealth) w. Suppose that he builds a portfolio by investing a fraction λ of his capital in the riskless asset and the rest in the risky asset. The value of the portfolio ("at time $t = 1$") is then $\lambda w(1 + r) + (1 - \lambda)wS_1/\pi_1$, and the discounted net gain is

$$(1 - \lambda)\frac{w}{\pi_1}(\frac{S_1}{1 + r} - \pi_1).$$

Proposition 4.7 (take $X = \frac{w}{\pi_1}S_1$) shows that $\lambda^* = 1$ (all capital invested in the riskless asset) iff $\frac{\mathbb{E}S_1}{1+r} \leq \pi_1$. Hence such an agent is only willing to invest in the risky asset, when the price is below the expected discounted value. Note that this holds for any risk averse investor, regardless the special form of the utility function u. Compare this with what happens under the risk-neutral measure of Section 1.1.

4.2 Arrow-Pratt coefficient

Suppose that one considers a probability measure μ that has finite variance and that is concentrated on a small interval around its mean $m = m(\mu)$. Let u be a C^2 utility function on a neighborhood of this interval and let U be the associated expected utility representation. Look at the following heuristic.

A first order Taylor expansion of u around m gives

$$u(x) \approx u(m) + (x - m)u'(m).$$

With $x = c(\mu)$ one obtains $u(c(\mu)) \approx u(m) + (c(\mu) - m)u'(m)$.

A second order Taylor expansion of u around m gives

$$u(x) \approx u(m) + (x-m)u'(m) + \frac{1}{2}(x-m)^2 u''(m).$$

Integrating w.r.t. μ yields $u(c(\mu)) = U(\mu) = \int u \, \mathrm{d}\mu \approx u(m) + \frac{1}{2}\mathbb{V}\mathrm{ar}\,(\mu)u''(m)$.

Hence, combining the two approximations, for the risk premium $\rho(\mu) = m - c(\mu)$ we have the approximation

$$\rho(\mu) \approx -\frac{1}{2}\frac{u''(m)}{u'(m)}\mathbb{V}\mathrm{ar}\,(\mu). \tag{4.2}$$

We shall see that, in spite of the rough heuristics, the right hand side of (4.2) contains a useful quantity.

Definition 4.9 For u, a twice differentiable utility function on some (open) interval S, the quantity

$$\alpha(x) := -\frac{u''(x)}{u'(x)}$$

is called the *Arrow-Pratt coefficient of absolute risk aversion of u at the level x*.

Note that by u being strictly concave and strictly increasing, $\alpha(x) \geq 0$ for every x. It moreover follows from Equation (4.2), that for probability measures μ that are concentrated around the mean m, the risk premium $\rho(\mu)$ approximately factors as a product of the Arrow-Pratt coefficient at the level m (a measure of the location of μ) and half the variance, the latter being an intrinsic quantity of μ only and location invariant.

Arrow-Pratt coefficients have the attractive feature that they are invariant under affine transformations. Since in Von Neumann-Morgenstern representations of preference orders, the function u is unique up to affine transformations, this means that the Arrow-Pratt coefficient in such a situation is an intrinsic feature of the preference order, not of its numerical representation (of course modulo the fact that we have to assume that u is C^2, and that u is not constant, which would lead anyway to an uninteresting preference order).

We now give some examples of widely used utility functions.

Example 4.10 Let u be such that the Arrow-Pratt function $\alpha(\cdot)$ is a (positive) constant, also denoted by α. Then, by solving a second order linear differential equation, one finds, for some constants $a \in \mathbb{R}$ and $b > 0$,

$$u_{a,b}(x) = a - be^{-\alpha x},$$

which is an affine transformation of $u(x) = 1 - \exp(-\alpha x)$. Note that u is defined on all of $\mathbb{R}$. The functions $u_{a,b}$ are called CARA functions (from Constant Absolute Risk Aversion).

Example 4.11 Here we introduce the HARA (from Hyperbolic Absolute Risk Aversion) utility functions. For these functions we have that $\alpha(x) = \frac{c}{x}$, for $c, x > 0$. For convenience we write $c = 1 - \gamma$, and hence $\gamma < 1$. Solving the corresponding differential equation for u yields

$$u_{a,b}(x) = \frac{a}{\gamma}x^\gamma + b,$$

for $\gamma \neq 0$ and $u_{a,b}(x) = a \log x + b$ for $\gamma = 0$. Note that $\gamma \geq 1$ is excluded by requiring that u is strictly concave and that for all $\gamma < 1$ it holds that $u'_{a,b}(x) = ax^{\gamma-1}$. The functions $u_{a,b}$ are affine transformations of $u_{1,0}$.

Remark 4.12 HARA utility functions with $\gamma > 0$ are examples of utility functions $u : [0, \infty) \to \mathbb{R}$ satisfying the *Inada conditions*, i.e. $u \in C^1(0, \infty)$, with $\lim_{x\to 0} u'(x) = \infty$ and $\lim_{x\to\infty} u'(x) = 0$.

There are close connections between utility functions, risk premia and Arrow-Pratt coefficients for different preference orders.

Proposition 4.13 *Suppose $u_1, u_2 : S \to \mathbb{R}$ are two C^2 utility functions, with corresponding risk premia $\rho_1(\cdot)$, $\rho_2(\cdot)$, certainty equivalents $c_1(\cdot)$, $c_2(\cdot)$ and Arrow-Pratt coefficients $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$. The following are equivalent.*

(i) $\alpha_1(x) \geq \alpha_2(x), \forall x \in S$.

(ii) There exist a strictly increasing concave function F, defined on the range of u_2, such that $u_1 = F \circ u_2$.

(iii) $\rho_1(\mu) \geq \rho_2(\mu), \forall \mu \in \mathcal{M}$.

Proof (i) $\Rightarrow$ (ii): The obvious choice of F is $F(x) = u_1(u_2^{-1}(x))$. Clearly, F is well defined, since u_2 is strictly increasing, and since u_2^{-1} and u_1 are strictly increasing, so is F. To show that F is strictly increasing and concave, we compute its first and second derivative and use that (i) is assumed. Notice that it sufficient to show that $F''(u_2(x)) \leq 0$, for all $x \in S$. We start with $u_1(x) = F(u_2(x))$ and get

$$\begin{aligned}
u_1'(x) &= F'(u_2(x))u_2'(x) \\
u_1''(x) &= F''(u_2(x))u_2'(x)^2 + F'(u_2(x))u_2''(x).
\end{aligned}$$

The first of these two equations shows that $F'(y) > 0$ for all y in the range of u_2, which implies that F is strictly increasing. Solving the second of these two equations for $F''(u_2(x))$ and using the first one yields

$$\begin{aligned}
F''(u_2(x)) &= \frac{u_1''(x) - \frac{u_1'(x)}{u_2'(x)}u_2''(x)}{u_2'(x)^2} \\
&= \frac{u_1'(x)}{u_2'(x)^2}\left(\frac{u_1''(x)}{u_1'(x)} - \frac{u_2''(x)}{u_2'(x)}\right) \\
&= \frac{u_1'(x)}{u_2'(x)^2}(\alpha_2(x) - \alpha_1(x)),
\end{aligned} \tag{4.3}$$

by definition of the Arrow-Pratt coefficients. By assumption (i) and the fact that u_1 is increasing, we have $F''(u_2(x)) \leq 0$.

(ii) ⇒ (iii): By Jensen's inequality, applied to the concave function F, it holds that

$$
\begin{aligned}
u_1(c_1(\mu)) &= \int u_1 \,\mathrm{d}\mu = \int F \circ u_2 \,\mathrm{d}\mu \\
&\leq F(\int u_2 \,\mathrm{d}\mu) = F(u_2(c_2(\mu))) = u_1(c_2(\mu)).
\end{aligned} \tag{4.4}
$$

Since u_1 is increasing, we must have $c_1(\mu) \leq c_2(\mu)$, from which the result follows, since $\rho_1(\mu) = m(\mu) - c_1(\mu)$ and $\rho_2(\mu) = m(\mu) - c_2(\mu)$.

(iii) ⇒ (i): Suppose that (i) doesn't hold. Then for some x one has $\alpha_1(x) < \alpha_2(x)$, and by continuity of α_1 and α_2, this equality extends to an open neighborhood O of x. By (4.3), which is also valid without assumptions (i) or (ii), we then have $F''(u_2(x)) > 0$ on O. Take now a nondegenerate probability measure μ such that $\mu(O) = 1$. Then strict convexity of F on $u_2(O)$ leads to a strict equality in the opposite direction as compared to (4.4), $u_1(c_1(\mu)) > u_1(c_2(\mu))$, from which it follows that $c_1(\mu) > c_2(\mu)$, contradicting assumption (iii). □

4.3 Exercises

4.1 Show that for a utility function $u \in C^1(\mathbb{R})$ it holds that $m(\mu) > c(\mu) > \frac{\mathbb{E}Xu'(X)}{\mathbb{E}u'(X)}$, where X has nondegenerate distribution μ and all expectations are assumed to be finite.

4.2 Let $u(x) = 1-\exp(-x)$, a CARA function. Consider an investor with utility function u who wants to invest an initial capital. There is one riskless asset, having value 1 and interest rate $r = 0$, and one risky assets with random pay-off S_1 having a normal $N(m, \sigma^2)$ distribution with $\sigma^2 > 0$. Suppose he invests a fraction λ in the riskless asset and the remainder in the risky asset. The pay-off of this portfolio is thus $\lambda + (1-\lambda)S_1$. The aim is to maximize his expected utility.

(a) Show that $\mathbb{E}\exp(uS_1) = \exp(um + \frac{1}{2}u^2\sigma^2)$ $(u \in \mathbb{R})$.

(b) Compute for each λ the certainty equivalent of the portfolio.

(c) Let λ^* be the optimal value of λ. Give, by direct computations, conditions on the parameters such that each of the cases $\lambda^* = 0$, $\lambda^* = 1$ or $\lambda^* \in (0,1)$ occurs.

(d) Compare the results of (c) with the assertions of Proposition 4.7.

4.3 In Exercise 4.2 the optimization problems turns out to be of the form: maximize $\mathbb{E}Z - c\,\mathbb{V}\mathrm{ar}\,Z$. This seems reasonable, if one thinks of Z as a random revenue. One wants to maximize the expected revenue and to keep the 'risk' in terms of variance low. In general such a maximization problems leads to odd results. Consider the following example. In two lotteries the random pay-off Z satisfies $\mathbb{P}(Z = h) = p_i$ and $\mathbb{P}(Z = \ell) = 1 - p_i$, $i = 1, 2$ and $h > \ell$. Find an example of values of $p_1 > p_2$ and $c > 0$ such that the second lottery is preferred to the first one.

4.4 Consider a twice differentiable utility function $u : S \to \mathbb{R}$. Define for fixed x such that $tx \in S$ the function $t \mapsto v_x(t) = u(tx)$. A way to establish the *relative* risk around x can obtained by inspection of $v_x(t)$ in a neighborhood of $t = 1$. A measure of relative risk at x is defined by $r(x) = -v_x''(1)/v_x'(1)$.

(a) Show that $r(x) = x\alpha(x)$ ($\alpha(x)$ is the Arrow-Pratt risk coefficient).

(b) Characterize the CRRA (Constant Relative Risk Aversion) utility functions, the functions u for which $r(x)$ is constant, not depending on x.

4.5 A utility function $u : \mathbb{R} \to \mathbb{R}$ is said to exhibit *decreasing risk aversion* if the function $x \mapsto \alpha(x)$ (the Arrow-Pratt coefficient) is decreasing. Show that this property is equivalent to saying that for every $x_1 < x_2$ there exists a concave function g such that $g(u(x_2 + z)) = u(x_1 + z)$ for all z (for which the given expressions make sense).

4.6 Let u be a (continuous) strictly increasing function, $u : S \to \mathbb{R}$. Consider the fair game represented by a random variable X with values $x \pm \varepsilon$ ($\varepsilon > 0$) in S that are attained with equal probabilities $\frac{1}{2}$. Given x, ε, the *probability premium* $\pi = \pi(x, \varepsilon)$ is by definition such that the lottery with the same outcomes but with probability $\mathbb{P}(\xi = x - \varepsilon) = \frac{1}{2} - \pi$ has expected utility $u(x)$. Show that an individual who uses u for a Von-Neumann-Morgenstern representation is risk averse (in which case u is a utility function) iff $\pi(x, \varepsilon) > 0$ for all x, ε. Sketch the graph of u and that of $\tilde{u}$ (the latter graph is the line segment joining the points $(x - \varepsilon, u(x - \varepsilon))$ and $(x + \varepsilon, u(x + \varepsilon))$), construct a point $\tilde{x}$ such that $\tilde{u}(\tilde{x}) = u(x)$ and indicate that $\tilde{x} > x$ iff $\pi(x, \varepsilon) > 0$.

4.7 Show the statement concerning $\lambda^* = 1$ in Proposition 4.7(iii).

4.8 The optimal value λ^* in Proposition 4.7 should be nondecreasing and continuous as a function of c, and strictly increasing on the interval $\big(\frac{\mathbb{E}Xu'(X)}{\mathbb{E}u'(X)}, m(\mu)\big)$. Verify whether this is true.

4.9 Let $f : S \to \mathbb{R}$ be concave, where S is an interval. Show that f is continuous on the interior of S and give an example where f is not continuous in a boundary point of S (which is assumed to belong to S).

4.10 Let u be an exponential utility function, $u(x) = -\exp(-\alpha x)$, $x \in \mathbb{R}$, $\alpha > 0$. Find the maximizing λ for the problem in Proposition 4.7 in each of the cases (a) X assumes two values only, (b) X has an exponential distribution, (c) X has a log-normal distribution. [I have not checked whether explicit solutions exist.]

4.11 Let $g : I \to \mathbb{R}$ be a concave, strictly increasing function on an interval I. Show that the inverse function g^{-1} defined on $g[I]$ is strictly increasing and convex.

4.12 Show that for an interval S and a utility function u there exists for all $\mu \in \mathcal{M}$ a unique number $c(\mu) \in S$ such that $u(c(\mu)) = U(\mu) \in u(S)$.

4.13 Verify heuristically assertion (iii) of Proposition 4.7 by computing the derivative $f'(\lambda)$, assuming that swapping differentiation and expectation is allowed (which is guaranteed to be true if the expectation is a finite sum).

5 Stochastic dominance

Results in the previous sections were depending on the preference orders, or the utility functions, at hand. In the present section, we will look at preferences that are independent of a particular choice of a utility function belonging to a certain class. The *standing assumptions* are that we deal with the set $\mathcal{M}$ of *all* probability measures on $(\mathbb{R}, \mathcal{B})$ that admit a *finite expectation*. As a consequence, for any utility function $u : \mathbb{R} \to \mathbb{R}$, the integrals $\int u \, \mathrm{d}\mu$ are well defined, but may take on the value $-\infty$. This holds, since every concave function has an affine function as a majorant. Indeed, since for some $a, b > 0$, one has $u(x) \leq ax + b$ for all x, it holds that $u(x)^+ \leq a|x| + b$ and hence $\int u^+ \, \mathrm{d}\mu < \infty$.

5.1 Uniform order

Definition 5.1 Let $\mu, \nu \in \mathcal{M}$. One says that μ is *uniformly preferred* over ν, denoted by $\mu \succeq_{\text{uni}} \nu$, if

$$\int u \, \mathrm{d}\mu \geq \int u \, \mathrm{d}\nu, \text{ for all utility functions } u : \mathbb{R} \to \mathbb{R}.$$

Remark 5.2 The uniform preference of the above definition is also called *second order stochastic dominance*. Notice that it is not a weak preference order (see Definition 2.2), since it is not complete. In Section 5.2 we will discuss first order stochastic dominance.

The next theorem gives a number of characterizations of uniform preference, there are many more. The functions f below are defined on all of $\mathbb{R}$.

Theorem 5.3 *There is equivalence between the following statements.*

(i) $\mu \succeq_{\text{uni}} \nu$.

(ii) For all increasing concave functions $f : \mathbb{R} \to \mathbb{R}$, *one has* $\int f \, \mathrm{d}\mu \geq \int f \, \mathrm{d}\nu$.

(iii) For all $c \in \mathbb{R}$, *it holds that* $\int (c - x)^+ \mu(\mathrm{d}x) \leq \int (c - x)^+ \nu(\mathrm{d}x)$.

(iv) If F_μ *and* F_ν *are the distribution functions of* μ *and* ν *respectively, then* $\int_{-\infty}^{c} F_\mu(x) \, \mathrm{d}x \leq \int_{-\infty}^{c} F_\nu(x) \, \mathrm{d}x$, *for all* $c \in \mathbb{R}$.

Proof (i) $\Leftrightarrow$ (ii): Obviously (ii) $\Rightarrow$ (i). For the converse implication we need a utility function that has finite integral under μ and ν. This can be accomplished as follows. Take a given utility function u and an arbitrary $x_0 \in \mathbb{R}$. Modify u on $(-\infty, x_0]$ by replacing u with $x \mapsto u'_+(x_0)(2(x - x_0) - \exp(x - x_0) + 1) + u(x_0)$ (see Figure 1 for an illustration, the dashed line is the modified utility function). Check that the modified function is still a utility function! Moreover, the modified utility function (denoted u again) has finite integral w.r.t. any probability measure with finite expectation. If f is increasing and concave, then $u_\alpha(x) := \alpha f(x) + (1 - \alpha)u(x)$ defines a strictly increasing, strictly concave continuous function, so a utility function, for every $\alpha \in [0, 1)$. Note that the

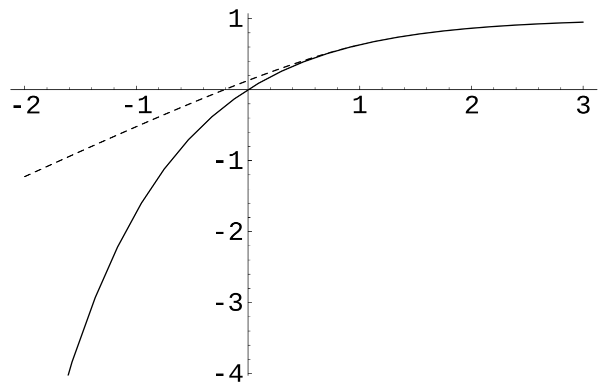

Figure 1: $u(x) = 1 - e^{-x}$ and $x_0 = 1$

integral $\int u_\alpha \, \mathrm{d}\mu$ is now always well defined, possibly taking the value $-\infty$. The assertion follows from

$$\int f \, \mathrm{d}\mu = \lim_{\alpha \uparrow 1} \int u_\alpha \, \mathrm{d}\mu \geq \lim_{\alpha \uparrow 1} \int u_\alpha \, \mathrm{d}\nu = \int f \, \mathrm{d}\nu.$$

(ii) $\Leftrightarrow$ (iii): Clearly (ii) $\Rightarrow$ (iii). The converse implication basically follows from the fact that every nonnegative convex decreasing function, with limit zero at infinity, is a pointwise limit of positive linear combinations of functions $x \mapsto (c-x)^+$ and that $-f$ is decreasing and convex. More formally, we have that $h = -f$ admits right derivatives $h'_+(x)$ at every point x. The function h'_+ is increasing, right continuous and on any interval $(a, b]$, *up to scaling*, it is a distribution function of a probability measure. Stated otherwise, there is a measure γ on $(\mathbb{R}, \mathcal{B})$ such that $\gamma(a, b] = h'_+(b) - h'_+(a)$, for all $a < b$. Since there exists only countably many discontinuity points of h'_+, we have for $x < b$

$$\begin{aligned} h(x) &= h(b) - \int_{(x,b]} h'_+(y) \, \mathrm{d}y \\ &= h(b) - \int_{(x,b]} (h'_+(y) - h'_+(b)) \, \mathrm{d}y - h'_+(b)(b-x). \end{aligned} \tag{5.1}$$

We first rewrite the integral in (5.1). Let $B = \{(u, y) : x < y < u \leq b\}$. We have

$$\begin{aligned} \int_{(x,b]} (h'_+(y) - h'_+(b)) \, \mathrm{d}y &= -\int_{(x,b]} \gamma(y, b] \, \mathrm{d}y \\ &= -\int_{(x,b]} \int \mathbf{1}_{(y,b]} \, \mathrm{d}\gamma \, \mathrm{d}y \\ &= -\int \int \mathbf{1}_B(u, y) \gamma(\mathrm{d}u) \, \mathrm{d}y \\ &= -\int \int \mathbf{1}_B(u, y) \, \mathrm{d}y \, \gamma(\mathrm{d}u) \text{ (by Fubini)} \end{aligned}$$

$$= -\int \mathbf{1}_{(x,b]}(u)(u-x)\gamma(\mathrm{d}u)$$
$$= -\int \mathbf{1}_{(-\infty,b]}(u-x)^+\gamma(\mathrm{d}u).$$

Hence, going back to (5.1), we can rewrite $h(x)$ as

$$h(x) = h(b) - h'_+(b)(b-x) + \int \mathbf{1}_{(-\infty,b]}(u-x)^+\gamma(\mathrm{d}u).$$

Let μ be a probability measure on $(\mathbb{R}, \mathcal{B})$. Integration of the last expression w.r.t. μ and using Fubini's theorem again, yields

$$\int_{(-\infty,b]} h\,\mathrm{d}\mu = h(b)\mu(-\infty,b] - h'_+(b)\int (b-x)^+\mu(\mathrm{d}x)$$
$$+ \int_{(-\infty,b]}\int_{(-\infty,b]} (u-x)^+\mu(\mathrm{d}x)\gamma(\mathrm{d}u)$$

$$= h(b)\mu(-\infty,b] - h'_+(b)\int (b-x)^+\mu(\mathrm{d}x)$$
$$+ \int_{(-\infty,b]}\int (u-x)^+\mu(\,\mathrm{d}x)\gamma(\mathrm{d}u).$$

Using condition (iii) and the fact that $h'_+ \leq 0$, we have an upper bound for the last displayed expression by replacing μ with ν. It follows that

$$\int_{(-\infty,b]} h\,\mathrm{d}\mu \leq \int_{(-\infty,b]} h\,\mathrm{d}\nu + h(b)(\mu(-\infty,b] - \nu(-\infty,b]).$$

Since h is lower bounded by an affine function, we have that $\int_{(b,\infty)} h\,\mathrm{d}\mu$ and $\int_{(b,\infty)} h\,\mathrm{d}\nu$ are both finite. Hence we obtain

$$\int h\,\mathrm{d}\mu \leq \int h\,\mathrm{d}\nu + \int_{(b,\infty)} h\,\mathrm{d}\mu - \int_{(b,\infty)} h\,\mathrm{d}\nu + h(b)(\mu(-\infty,b] - \nu(-\infty,b])$$
$$= \int h\,\mathrm{d}\nu - \int_{(b,\infty)} (h(b)-h(x))\mu(\mathrm{d}x) + \int_{(b,\infty)} (h(b)-h(x))\nu(\mathrm{d}x).$$

We finally show that the last two integrals vanish for $b \to \infty$. Since they are similar, we treat only the first of the two. Fix b_0 and let $b > b_0$. It holds that $0 \leq h(b) - h(x) \leq -h'_+(b_0)(x-b_0)$ for $x > b$. Hence

$$\int_{(b,\infty)} (h(b)-h(x))\mu(\mathrm{d}x) \leq -h'_+(b_0)\int (x-b_0)\mathbf{1}_{(b,\infty)}(x)\,\mathrm{d}\mu,$$

which tends to zero by the Dominated convergence theorem, since $\int |x|\mu(\mathrm{d}x)$ is finite. Hence we obtain $\int h\,\mathrm{d}\mu \leq \int h\,\mathrm{d}\nu$, which is equivalent to (ii).

(iii) ⇔ (iv): This is just a matter of rewriting, using Fubini's theorem. One has

$$\int_{-\infty}^{c} F_\mu(y)\,\mathrm{d}y = \int_{-\infty}^{c}\int \mathbf{1}_{(-\infty,y]}(x)\mu(\mathrm{d}x)\,\mathrm{d}y$$

$$= \int_{(-\infty,c]}\int_x^c \mathrm{d}y\,\mu(\mathrm{d}x)$$

$$= \int_{(-\infty,c]}(c-x)\mu(\mathrm{d}x)$$

$$= \int (c-x)^+\mu(\mathrm{d}x). \tag{5.2}$$

The integral with F_ν can be rewritten in similar terms and the equivalence of (iii) and (iv) becomes obvious. □

Remark 5.4 It follows from Theorem 5.3(ii), that $\mu \succeq_{\mathrm{uni}} \nu$ implies $m(\mu) \geq m(\nu)$. The integrals w.r.t. the measure μ in assertion (iii) of the same theorem in fact determine μ. Indeed, by the computations leading to (5.2), we see that knowing integrals of $(c-x)^+$ for all c is equivalent to knowing the integrals of F_μ up to c. Taking right derivatives w.r.t. c gives $F_\mu(c)$ and knowing this for all c determines μ. This fact can be used to show that $\succeq_{\mathrm{uni}}$ defines a partial order, Exercise 5.1.

If μ is the distribution of a random variable $X \in L^1(\Omega,\mathcal{F},\mathbb{P})$ and ν that of $Y \in L^1(\Omega,\mathcal{F},\mathbb{P})$, such that $\mu \succeq_{\mathrm{uni}} \nu \succeq_{\mathrm{uni}} \mu$, then $\mu = \nu$, so X and Y have the same *distribution* (under $\mathbb{P}$). Yet, X and Y are in general very different as random variables. It may happen that $\mathbb{P}(X = Y) = 0$.

When two lotteries with the same mean are compared, we can develop the assertions of Theorem 5.3 a little further.

Proposition 5.5 *For all probability measures $\mu,\nu \in \mathcal{M}$ the following are equivalent.*

(i) $\mu \succeq_{\mathrm{uni}} \nu$ *and* $m(\mu) = m(\nu)$.

(ii) $\int f\,\mathrm{d}\mu \geq \int f\,\mathrm{d}\nu$, *for all concave functions* f.

(iii) $m(\mu) \geq m(\nu)$ *and* $\int (x-c)^+\mu(\mathrm{d}x) \leq \int (x-c)^+\nu(\mathrm{d}x)$, *for all* $c \in \mathbb{R}$.

Proof (i) ⇒ (ii): First we show that the assertion holds true for decreasing concave functions. Such a function is $x \mapsto -(c-x)^-$, for arbitrary $c \in \mathbb{R}$. Since $-(c-x)^- = c - x - (c-x)^+$, the assertion for such a function follows from Theorem 5.3 and the assumptions that $m(\mu) = m(\nu)$ and $\mu \succeq_{\mathrm{uni}} \nu$, because $x \mapsto -(c-x)^+$ is concave and increasing. The proof for arbitrary decreasing concave functions is then similar to the proof of (iii) ⇒ (ii) of Theorem 5.3. The second assertion of Theorem 5.3 also tells us that (ii) is true for increasing concave functions, and hence (ii) holds for monotone concave functions.

If f is concave, but not monotone, then there exists a $x_0 \in \mathbb{R}$, such that $f(x) \leq f(x_0)$, for all $x \in \mathbb{R}$. Let

$$f_1(x) = \begin{cases} f(x) \text{ if } x \leq x_0 \\ f(x_0) \text{ if } x > x_0 \end{cases}$$

and

$$f_2(x) = \begin{cases} f(x_0) \text{ if } x \leq x_0 \\ f(x) \text{ if } x > x_0. \end{cases}$$

Then f_1 is concave and increasing and f_2 is concave and decreasing. Knowing that the assertions hold true for f_1 and f_2, we obtain the same result for f, because $f(x) = f_1(x) + f_2(x) - f(x_0)$ and integration of the constant is the same for each probability measure.

(ii) $\Rightarrow$ (iii): Take first $f(x) \equiv x$ to get the first assertion, and then $f(x) \equiv -(x-c)^+$, which is concave, to get the second one from (ii).

(iii) $\Rightarrow$ (i): Rewrite the inequality between the integrals in (iii) as

$$\int_{(c,\infty)} x\,\mu(\mathrm{d}x) - c + c\,\mu(-\infty, c] \leq \int_{(c,\infty)} x\,\nu(\mathrm{d}x) - c + c\,\nu(-\infty, c].$$

Let $c \to -\infty$ and use that both measures have a finite first moment to conclude that $c\,\mu(-\infty, c]$ and $c\,\nu(-\infty, c]$ tend to zero as well as $\int_{(c,\infty)} x\,\mu(\mathrm{d}x) \to \int_{\mathbb{R}} x\,\mu(\mathrm{d}x)$ and $\int_{(c,\infty)} x\,\nu(\mathrm{d}x) \to \int_{\mathbb{R}} x\,\nu(\mathrm{d}x)$. One then arrives at $\int x\,\mu(\mathrm{d}x) \leq \int x\,\nu(\mathrm{d}x)$, or $m(\mu) \leq m(\nu)$. Together with the assumption, this gives $m(\mu) = m(\nu)$.

To prove $\mu \succeq_{\text{uni}} \nu$ we use the identity $y^+ = y + (-y)^+$ $(y \in \mathbb{R})$ to get

$$\int (c-x)^+ \mu(\mathrm{d}x) = c - m(\mu) + \int (x-c)^+ \mu(\mathrm{d}x).$$

A similar equality holds for ν. Using the assumption and the just proved identity $m(\mu) = m(\nu)$, we arrive at $\int (c-x)^+ \mu(\mathrm{d}x) \leq \int (c-x)^+ \nu(\mathrm{d}x)$, condition (iii) in Theorem 5.3 to get $\mu \succeq_{\text{uni}} \nu$. □

Remark 5.6 Assume that $\mu_1 \succeq_{\text{uni}} \mu_2$ and $m(\mu_1) = m(\mu_2)$. Then it follows from Proposition 5.5 that $\mathbb{V}\text{ar}\,\mu_1 \leq \mathbb{V}\text{ar}\,\mu_2$. For normal distributions there is a converse relationship, see Exercise 5.2.

5.2 Monotone order

We turn to another concept of stochastic dominance, also called *first order stochastic dominance*. There are more of these concepts conceivable.

Definition 5.7 Let μ, ν be two probability measures on $(\mathbb{R}, \mathcal{B})$. One says that μ *stochastically dominates* ν, if for all bounded increasing continuous functions $f : \mathbb{R} \to \mathbb{R}$ it holds that

$$\int f\,\mathrm{d}\mu \geq \int f\,\mathrm{d}\nu. \tag{5.3}$$

In this case one writes $\mu \succeq_{\text{mon}} \nu$.

It is almost trivial to check that $\succeq_{\text{mon}}$ defines a partial order on the space of probability distributions on $(\mathbb{R}, \mathcal{B})$. Below we give an easy characterization of $\mu \succeq_{\text{mon}} \nu$.

Proposition 5.8 *Let μ, ν be two probability measures on $(\mathbb{R}, \mathcal{B})$ and let F_μ and F_ν be their distribution functions. The following are equivalent.*

(i) It holds that $\mu \succeq_{\text{mon}} \nu$.

(ii) For all $x \in \mathbb{R}$ one has $F_\mu(x) \leq F_\nu(x)$.

Proof (i) $\Rightarrow$ (ii): We'd like to apply the definition of stochastic dominance to the function $u \mapsto \mathbf{1}_{(x,\infty)}(u)$, which is bounded and increasing. The result would then follow. However this function is not continuous. Therefore one first uses the functions $u \mapsto (\min\{n(u-x), 1\})^+$ and let $n \to \infty$.

(ii) $\Rightarrow$ (i): Let f be continuous, bounded and increasing. We can obtain f (which is measurable) as the pointwise limit of an increasing sequence of simple functions f_n, that are increasing themselves. To see this, we assume for simplicity that $0 \leq f \leq 1$ and we follow the usual approximation scheme, known from measure theory.

Let $n \in \mathbb{N}$ and define $E_{ni} = \{(i-1)2^{-n} < f \leq i2^{-n}\}$ for $i = 1, \ldots, 2^n$. Put

$$f_n = 2^{-n} \sum_{i=1}^{2^n} (i-1) \mathbf{1}_{E_{ni}}.$$

Then we know that $f_n \leq f$ and $f_n \uparrow f$. Using that, for each n, the E_{ni} with $i = 0, \ldots, 2^n$ are disjoint, $\bigcup_{i \geq j+1} E_{ni} = \{f > j2^{-n}\}$ and $\{f > 1\} = \emptyset$, we rewrite

$$f_n = 2^{-n} \sum_{i=1}^{2^n} (\sum_{j=1}^{i-1} 1) \mathbf{1}_{E_{ni}} = 2^{-n} \sum_{j=1}^{2^n-1} \sum_{i=j+1}^{2^n} \mathbf{1}_{E_{ni}} = 2^{-n} \sum_{j=1}^{2^n} \mathbf{1}_{\{f > j2^{-n}\}}.$$

Since f is continuous, the sets $\{f > j2^{-n}\}$ are open and since f is increasing, there are real numbers a_{nj} such that $\{f > j2^{-n}\} = (a_{nj}, \infty)$. Hence,

$$\int f_n \, \mathrm{d}\mu = 2^{-n} \sum_{j=1}^{2^n} \mu((a_{nj}, \infty)) = 2^{-n} \sum_{j=1}^{2^n} (1 - F_\mu(a_{nj})).$$

It follows from the assumption that $\int f_n \, \mathrm{d}\mu \geq \int f_n \, \mathrm{d}\nu$. The assertion follows by application of the Monotone Convergence Theorem. $\square$

Remark 5.9 It follows from Theorem 5.3 and Proposition 5.8 that $\mu \succeq_{\text{mon}} \nu$ implies $\mu \succeq_{\text{uni}} \nu$.

5.3 Exercises

5.1 Show that $\succeq_{\text{uni}}$ defines a partial order on the set of probability measures with finite mean (see also Remark 5.4).

5.2 Consider two normal distributions, $\mu_1 = N(m_1, \sigma_1^2)$ and $\mu_2 = N(m_2, \sigma_2^2)$.

(a) Compute $\int_{\mathbb{R}} \exp(-ax)\mu_i(\mathrm{d}x)$ and show that $\mu_1 \succeq_{\text{uni}} \mu_2$ implies both $m_1 \geq m_2$ and $\sigma_1^2 \leq \sigma_2^2$.

(b) Assume that $m_1 = m_2$. Show (use Theorem 5.3(iv)) that $\sigma_1^2 \leq \sigma_2^2$ implies $\mu_1 \succeq_{\text{uni}} \mu_2$.

(c) Let u be a utility function and assume $m_1 \geq m_2$. Put $\tilde{u}(x) = u(x + m_2)$. Verify that $\mathbb{E}u(N(m_1, \sigma_1^2)) \geq \mathbb{E}\tilde{u}(N(0, \sigma_1^2))$ (the notation should be obvious).

(d) Let $m_1 \geq m_2$ and $\sigma_1^2 \leq \sigma_2^2$. Show that $\mu_1 \succeq_{\text{uni}} \mu_2$.

5.3 Let $\mu, \nu \in \mathcal{M}$ and f an increasing function such that $\int |f| \, \mathrm{d}\mu$ and $\int |f| \, \mathrm{d}\nu$ are both finite. Show that $\mu \succeq_{\text{mon}} \nu$ implies $\int f \, \mathrm{d}\mu \geq \int f \, \mathrm{d}\nu$ and thus $\mu \succeq_{\text{uni}} \nu$.

5.4 Let $\mu \succeq_{\text{mon}} \nu$ and $m(\mu) = m(\nu)$. Show that $\mu = \nu$. *Hint*: compute $0 \leq \int_a^b (F_\nu(x) - F_\mu(x)) \, \mathrm{d}x$ for any $a < b$. Use integration by parts and let $a \to -\infty, b \to \infty$.

5.5 A random variable X has a log-normal distribution with parameters α and σ, if $X = \exp(\alpha + \sigma Z)$, where $\sigma \geq 0$ and Z has a standard normal distribution.

(a) Compute $\mathbb{E}X^p$ for $p > 0$. In particular, one has $\mathbb{E}X = \exp(\alpha + \frac{1}{2}\sigma^2)$.

(b) Let μ_i be log-normal distributions ($i = 1, 2$) with parameters α_i, σ_i. Show that $\mu_1 \succeq_{\text{uni}} \mu_2$ implies $m(\mu_1) \geq m(\mu_2)$ and $\sigma_1 \leq \sigma_2$.

(c) Conversely, if $m(\mu_1) \geq m(\mu_2)$ and $\sigma_1 \leq \sigma_2$, then $\mu_1 \succeq_{\text{uni}} \mu_2$. To prove this, proceed as follows. Let $X_1 = \exp(\alpha_1 + \sigma Z_1)$ and $X_2 = \exp(\alpha_2 + \sigma Z_2)$ (in obvious notation). Let further $X_3 = \exp(\alpha_2 - \alpha_1 + \sqrt{\sigma_2^2 - \sigma_1^2} Z_3$, where Z_3 is standard normal, independent of Z_1. Verify that $X_1 X_3$ has the same distribution as X_2 and that $\mathbb{E}X_3 = \frac{m(\mu_2)}{m(\mu_1)}$. Use then Jensen's inequality for conditional expectations to show that $\mathbb{E}u(X_2) \leq \mathbb{E}u(X_1)$.

5.6 Definition 5.1 requires utility functions u to be defined on all of $\mathbb{R}$, and thus rules out for instance $u(x) = \sqrt{x}$. To include such a utility function, or rather a utility function defined on an interval $[a, \infty)$, one can extend the definition of u to all of $\mathbb{R}$ by putting $u(x) = -\infty$ for $x < a$. This extended u is not strictly increasing anymore, nor strictly concave, nor continuous everywhere. But it is still concave. If Definition 5.1 is extended to include such extended utility functions, are the integrals there still well defined (possibly with values $-\infty$), and is Theorem 5.3 still valid?

6 Portfolio optimization

In this section we return to the setting of Section 1 and combine it with the expected utility setting of Section 4. We consider an investor, whose preferences are determined by a utility function $\tilde{u}$, and who wants to invest a capital w (w from wealth). On the market there are d risky assets having a price (at $t = 0$) given by the vector π and whose future, at time $t = 1$, random pay-off is described by the random vector S, defined on a underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Both vectors are assumed to have strictly positive entries. Next to the risky assets, there is a riskless asset, with price $\pi_0 = 1$ and future value $S_0 = 1 + r > 0$. Let $\bar{S} = (S_0, S)$. A portfolio is given by $\bar{\xi} \in \mathbb{R}^{d+1}$ and we also write $\bar{\xi} = (\xi_0, \xi) \in \mathbb{R} \times \mathbb{R}^d$. The future value of the portfolio is then $\bar{\xi} \cdot \bar{S}$, and it has expected utility

$$\mathbb{E}\tilde{u}(\bar{\xi} \cdot \bar{S}). \tag{6.1}$$

In order for the investor to purchase the portfolio $\bar{\xi}$, the price of it should at most be equal to the initial capital. Thus we have the *budget constraint*

$$\bar{\xi} \cdot \bar{\pi} \leq w. \tag{6.2}$$

We will study the problem of maximizing (6.1) over portfolios $\bar{\xi}$, satisfying the constraint (6.2).

6.1 Optimization and absence of arbitrage

We start this section by casting the above problem in a different, but equivalent form. Our first observation is that it can never be optimal to use only a fraction of the initial capital w. Indeed, suppose one has a portfolio $\bar{\xi}$ with $\bar{\xi} \cdot \bar{\pi} < w$. Change the investment ξ_0 into $\xi_0' = \xi_0 + w - \bar{\xi} \cdot \bar{\pi}$. Then we have $(\xi_0', \xi) \cdot \bar{S} = \bar{\xi} \cdot \bar{S} + (w - \bar{\xi} \cdot \bar{\pi})(1 + r) > \bar{\xi} \cdot \bar{S}$. But then, since $\tilde{u}$ is strictly increasing, also $\mathbb{E}\tilde{u}((\xi_0', \xi) \cdot \bar{S}) > \mathbb{E}\tilde{u}(\bar{\xi} \cdot \bar{S})$. Therefore, *we assume from now on that equality holds in (6.2)*, and so we work with

$$\bar{\xi} \cdot \bar{\pi} = w. \tag{6.3}$$

Recall that we denoted by Y the d-dimensional random vector of discounted net gains,

$$Y = \frac{S}{1 + r} - \pi.$$

Hence we have, assuming (6.3), $\bar{\xi} \cdot \bar{S} = (1 + r)(\xi \cdot Y + w)$. Define a new utility function u by $u(x) = \tilde{u}((1 + r)(w + x))$. Then we have $\tilde{u}(\bar{\xi} \cdot \bar{S}) = u(\xi \cdot Y)$. Of course, this expression only makes sense if $\xi \cdot Y \in D$, where D is the domain of u. Given a risky portfolio ξ, by (6.3) one can always choose a non-risky investment ξ_0 such that the total portfolio has initial price w. This makes the constraint (6.3) redundant, if one considers only ξ as the free variable. All these arguments motivate to study the following *unconstrained* optimization problem, equivalent to the original one.

Problem 6.1 Let $u : D \to \mathbb{R}$ be a utility function. Maximize

$$\mathbb{E}u(\xi \cdot Y)$$

over all risky portfolios ξ that satisfy $\xi \cdot Y \in D$.

We will study this problem under each of the two cases in the assumption below.

Assumption 6.2 Let $u : D \to \mathbb{R}$ be a utility function and Y the vector of discounted net gains. Assume either of the following.

(i) $D = \mathbb{R}$ and u is bounded from above, or

(ii) $D = [a, \infty)$ for some $a < 0$, and we optimize over the set of ξ such that $\xi \cdot Y \geq a$ a.s. In this case, we also assume that for those ξ the expected utility $\mathbb{E}u(\xi \cdot Y)$ is finite.

For both of theses case we write $\Xi = \{\xi \in \mathbb{R}^d : \xi \cdot Y \in D \text{ a.s.}\}$.

Theorem 6.5 below shows that the maximization problem 6.1 only makes sense in an arbitrage-free market, just as pricing of portfolios and derivatives. In the proof of it we use the following two lemmas. Recall the definition of a upper semicontinuous function h (often abbreviated as u.s.c. function), it is such that $\limsup h(x_n) \leq h(x)$, whenever $x_n \to x$. A characterization of a function h to be u.s.c. is that all sets $\{h \geq c\}$ ($c \in \mathbb{R}$) are closed (This is Exercise 6.1).

Lemma 6.3 *Let $h : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ be a concave and upper semicontinuous function with $h(0) > -\infty$. Then h attains its supremum, if for all $\xi \neq 0$*

$$\lim_{\alpha \uparrow \infty} h(\alpha\xi) = -\infty. \tag{6.4}$$

Proof Let $c < \sup h$. We will see that the non-empty (!) level set $\{h \geq c\}$ is compact. By the fact that h is u.s.c., this set is closed. So, by the Heine-Borel theorem we only have to show that it is bounded. Suppose that it is unbounded, then there exists a sequence (x_n) such that $|x_n| \to \infty$ and $h(x_n) \geq c$, for all n. We may assume that the normalized vectors $x_n/|x_n|$ converge to some limit ξ. Let $\alpha > 0$ and consider $h(\alpha\xi)$. One has for all n large enough $\alpha/|x_n| \in (0, 1)$, which will be used below, when concavity comes into play.

$$\begin{aligned}
h(\alpha\xi) &= h(\lim \alpha \frac{x_n}{|x_n|}) \geq \limsup h(\alpha \frac{x_n}{|x_n|}) \\
&= \limsup h(\frac{\alpha}{|x_n|} x_n + (1 - \frac{\alpha}{|x_n|}) \cdot 0) \\
&\geq \limsup(\frac{\alpha}{|x_n|} h(x_n) + (1 - \frac{\alpha}{|x_n|}) h(0)) \\
&\geq \limsup(\frac{\alpha}{|x_n|} c + (1 - \frac{\alpha}{|x_n|}) h(0)) \\
&= h(0) > -\infty,
\end{aligned}$$

which contradicts the assumption (6.4) on h.

Consider the identity $\{h = \sup h\} = \bigcap_{c<\sup h}\{h \geq c\}$. Knowing that $\{h \geq c\}$ is compact for all $c < \sup h$, we have here an (infinite) intersection of *nested* non-empty compact sets. By a property of compactness, this intersection is non-empty. □

Lemma 6.4 *Let $u : D \to \mathbb{R}$ be a utility function, where $D = [a, \infty)$, $a < 0$. Let $0 \leq b < -a$. Let $X \geq 0$ be a random variable. Then for all $\alpha \in (0, 1]$ the implication*

$$\mathbb{E}u(\alpha X - b) < \infty \Rightarrow \mathbb{E}u(X) < \infty$$

holds.

Proof From concavity of u and $X \geq 0$ one obtains for $X > 0$ the following relations between slopes (a picture makes this clear)

$$\frac{u(X) - u(0)}{X} \leq \frac{u(\alpha X) - u(0)}{\alpha X} \leq \frac{u(\alpha X - b) - u(-b)}{\alpha X}.$$

Hence $u(X) - u(0) \leq (u(\alpha X - b) - u(-b))/\alpha$ (also valid if $X = 0$), from which the result follows. □

Theorem 6.5 *Let $u : D \to \mathbb{R}$ be a utility function and Y the vector of discounted net gains. Let Assumption 6.2 be satisfied. A maximizer in Problem 6.1 exists if and only if the market is free of arbitrage. In this case the maximizer is unique if the market is non-redundant (see Definition 1.13).*

Proof First we consider uniqueness under non-redundancy. Proposition 1.14(ii) tells us that in a non-redundant market the a.s. equality $\xi \cdot Y = \xi' \cdot Y$ implies that $\xi = \xi'$. Hence the function $\xi \mapsto u(\xi \cdot Y)$ is a.s. strictly concave, and then also $\xi \mapsto h(\xi) := \mathbb{E}u(\xi \cdot Y)$ is strictly concave (use Jensen's inequality for the latter statement). Suppose that two different maximizers ξ^* and ξ' exist. Then by strict concavity one would have $h(\frac{1}{2}(\xi^* + \xi') = \mathbb{E}u(\frac{1}{2}(\xi^* + \xi') \cdot Y) > \frac{1}{2}(\mathbb{E}u(\xi^* \cdot Y) + \mathbb{E}u(\xi' \cdot Y)) = \mathbb{E}u(\xi^* \cdot Y) = h(\xi^*)$, meaning that $h(\xi^*)$ is not maximal. One concludes that there can be at most one maximizer.

We turn to existence. Suppose that the market admits an arbitrage opportunity. Let ξ be any risky portfolio. By Corollary 1.4, there exists a portfolio ξ' such that $\xi' \cdot Y \geq 0$ a.s. and $\mathbb{P}(\xi' \cdot Y > 0) > 0$. In this case one has $h(\xi + \xi') = \mathbb{E}u((\xi + \xi') \cdot Y) > \mathbb{E}u(\xi \cdot Y) = h(\xi)$ and therefore a maximizing portfolio cannot exist. Hence absence of arbitrage is a necessary condition.

To show sufficiency, we assume that the market is free of arbitrage. Without loss of generality we can even assume that the market is non-redundant. Indeed, if the non-redundancy condition doesn't hold, one proceeds as follows, assuming that a maximizer in the non-redundant case exists. Let $N = \{\xi \in \mathbb{R}^d : \xi \cdot Y = 0 \text{ a.s.}\}$. Then N is a closed subspace of $\mathbb{R}^d$ and hence every vector ξ in $\mathbb{R}^d$ can be orthogonally decomposed as $\xi = \xi_0 + \xi^\perp$ with $\xi_0 \in N$ and $\xi^\perp \in N^\perp$. It follows that $h(\xi) = h(\xi^\perp)$, so effectively, h is a function on $\Xi' := \Xi \cap N^\perp$, which has, by assumption, a maximizer $\xi' \in \Xi'$. Then $\xi^* = \xi' + \xi_0$, for some $\xi_0 \in N$, is a maximizer in Ξ. Henceforth in the proof, non-redundancy is assumed.

Consider first the case where $D = \mathbb{R}$ and u has an upper bound. We will invoke Lemma 6.3 applied to the function $h(\xi) = \mathbb{E}u(\xi \cdot Y)$. We first show that h is u.s.c. Since u has an upper bound, we can apply the lim sup version of Fatou's lemma and obtain for every sequence (ξ^n) with limit ξ

$$\limsup h(\xi^n) = \limsup \mathbb{E}u(\xi^n \cdot Y) \leq \mathbb{E} \limsup u(\xi^n \cdot Y) = h(\xi),$$

by continuity of u.

Next we check condition (6.4). In view of Corollary 1.4, absence of arbitrage is equivalent to saying that for every $\xi \in \mathbb{R}^d \setminus \{0\}$ one has $\mathbb{P}(\xi \cdot Y < 0) > 0$. Indeed, $\mathbb{P}(\xi \cdot Y \geq 0) = 1$ implies $\mathbb{P}(\xi \cdot Y = 0) = 1$, which in turn implies $\xi = 0$ by non-redundancy. Hence if $\xi \neq 0$, then $\mathbb{P}(\xi \cdot Y < 0) > 0$.

Since u is concave and increasing one has $\{\xi \cdot Y < 0\} = \{\lim_{\alpha \uparrow \infty} u(\alpha\, \xi \cdot Y) = -\infty\}$. From the fact that the latter set has positive probability, it follows by the Monotone convergence theorem or the reverse Fatou lemma (use also that u has an upper bound) that for all $\xi \neq 0$

$$\lim_{\alpha \to \infty} h(\alpha\, \xi) = \lim_{\alpha \to \infty} \mathbb{E}u(\alpha\, \xi \cdot Y) = -\infty.$$

We have shown that for the present case, absence of arbitrage leads to condition (6.4), which is sufficient for existence of a maximum of $\mathbb{E}u(\xi \cdot Y)$.

We turn to the case, where all $\xi \cdot Y$ involved have a lower bound $a < 0$. We show that $\Xi = \{\xi \in \mathbb{R}^d : \xi \cdot Y \geq a \text{ a.s.}\}$ is compact. We follow a familiar way of reasoning (see also the proof of Theorem 1.16), working towards a contradiction. Supposing that the set Ξ is unbounded, we can take a sequence (ξ^n) in this set such that $|\xi^n| \to \infty$ and $\xi^n/|\xi^n| \to \eta$, for some vector η with $|\eta| = 1$. Then

$$\eta \cdot Y = \lim \frac{\xi^n \cdot Y}{|\xi^n|} \geq \lim \frac{a}{|\xi^n|} = 0 \text{ a.s.}$$

By absence of arbitrage and non-redundancy we conclude that $\eta = 0$, a contradiction. We conclude to optimize over a compact set.

To show that an optimizer exists, it is now sufficient to show that h is continuous on Ξ. This follows by an application of the Dominated convergence theorem, $\lim h(\xi_n) = \lim \mathbb{E}u(\xi_n \cdot Y) = \mathbb{E} \lim u(\xi_n \cdot Y) = h(\xi)$ if $\xi_n \to \xi$. For a valid application of this theorem one has to find a random variable X such that $\sup_{\xi \in \Xi} u(\xi \cdot Y) \leq u(X)$ a.s. and $\mathbb{E}u(X) < \infty$, an integrable *upper* bound is sufficient since u is lower bounded by $u(a)$.

Define $\eta \in \mathbb{R}^d_+$ by its elements $\eta_i = 0 \vee m_i$, where $m_i = \max\{\xi_i : \xi \in \Xi\}$. The m_i are finite by compactness of Ξ. By positivity of S, we have $\eta \cdot S \geq \xi \cdot S$ for $\xi \in \Xi$ and hence

$$\xi \cdot Y \leq \frac{\eta \cdot S}{1+r} - \xi \cdot \pi \leq \frac{\eta \cdot S}{1+r} + M =: X,$$

where $M = \max\{-\xi \cdot \pi : \xi \in \Xi\} \vee 0$ (also a finite number). We also have $\eta \cdot Y = \eta \cdot (\frac{S}{1+r} - \pi) \geq -\eta \cdot \pi$ and, because $\eta \cdot \pi \geq 0$, there is $\alpha \in (0, 1]$ such that $\alpha\eta \cdot \pi < -a$, which implies that $\alpha\eta \cdot Y > a$. Then $\alpha\eta \in \Xi$ and, by Assumption 6.2,

$\mathbb{E}u(\alpha\eta\cdot Y)<\infty$. One has $a\leq\xi\cdot Y\leq X$ and $\alpha\eta\cdot Y=\alpha X-\alpha(\eta\cdot\pi+M)$. With $b=\alpha(\eta\cdot\pi+M)$ we wish to apply Lemma 6.4 to get $\mathbb{E}u(X)<\infty$, as desired. One then has to verify that $0\leq\alpha(\eta\cdot\pi+M)<-a$, which can be accomplished by taking α small enough, meanwhile maintaining $\mathbb{E}u(\alpha\eta\cdot Y)<\infty$. This finishes the proof. □

Knowing that under Assumption 6.2, the maximization problem has a solution, we now turn to a characterization of it under additional assumptions.

Theorem 6.6 *Let $u:D\to\mathbb{R}$ be a continuously differentiable utility function. Let Assumption 6.2 hold and assume, additionally, $\mathbb{E}|u(\xi\cdot Y)|<\infty$ for all $\xi\in\Xi$. Let the Problem 6.1 maximizing ξ^* be an interior point of Ξ. Then $Yu'(\xi^*\cdot Y)\in\mathcal{L}^1(\Omega,\mathcal{F},\mathbb{P})$ and*

$$\mathbb{E}Yu'(\xi^*\cdot Y)=0. \tag{6.5}$$

Proof If differentiation and expectation commute, one has (writing a gradient as a column vector)

$$\nabla_\xi\mathbb{E}u(\xi\cdot Y)=\mathbb{E}u'(\xi\cdot Y)Y,$$

and the result follows by taking $\xi=\xi^*$. Since it is not clear that the commutation is valid, we directly show that the right hand side is zero at $\xi=\xi^*$. Take $\eta\in\mathbb{R}^d$ and $\varepsilon\in(0,1]$. Put $\xi_\varepsilon=\xi^*+\varepsilon\eta$, then $\xi_\varepsilon\in\Xi$ for all ε sufficiently small, $\varepsilon<\varepsilon_0$ say. For those ε we put $f(\varepsilon):=u(\xi_\varepsilon\cdot Y)$ and

$$\Delta_\varepsilon:=\frac{f(\varepsilon)-f(0)}{\varepsilon}=\frac{u(\xi_\varepsilon\cdot Y)-u(\xi^*\cdot Y)}{\varepsilon}=\eta\cdot Y\frac{u(\xi_\varepsilon\cdot Y)-u(\xi^*\cdot Y)}{\varepsilon\eta\cdot Y}.$$

Note that $\mathbb{E}\Delta_\varepsilon\leq 0$, because $\mathbb{E}u(\xi^*\cdot Y)$ is maximal. Concavity of u gives that f is concave too. Hence Δ_ε is increasing for $\varepsilon\downarrow 0$, with limit $\eta\cdot Yu'(\xi^*\cdot Y)$. The assumption that $u(\xi\cdot Y)\in\mathcal{L}^1(\Omega,\mathcal{F},\mathbb{P})$ for all $\xi\in\Xi$ implies that $\Delta_{\varepsilon_0}\in\mathcal{L}^1(\Omega,\mathcal{F},\mathbb{P})$. Hence $\Delta_\varepsilon-\Delta_{\varepsilon_0}$ is nonnegative and increasing for $\varepsilon\downarrow 0$, which enables us to apply the Monotone convergence theorem to get

$$0\geq\mathbb{E}\Delta_\varepsilon\uparrow\mathbb{E}[\eta\cdot Yu'(\xi^*\cdot Y)],$$

where the expectation on the right hand side is a finite number. We conclude that $\eta\cdot\mathbb{E}Yu'(\xi^*\cdot Y)\leq 0$ for all $\eta\in\mathbb{R}^d$. So we can replace η with $-\eta$ in the last inequality and we conclude that the linear map $\eta\mapsto\eta\cdot\mathbb{E}Yu'(\xi^*\cdot Y)$ is identically zero. But then we must have $\mathbb{E}Yu'(\xi^*\cdot Y)=0$. □

Proposition 6.7 *Let the assumptions of Theorem 6.6 hold and let the market be arbitrage-free. Let ξ^* be the maximizer of Problem 6.1. Then $\mathbb{E}u'(\xi^*\cdot Y)<\infty$ and*

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}=\frac{u'(\xi^*\cdot Y)}{\mathbb{E}u'(\xi^*\cdot Y)} \tag{6.6}$$

defines a risk-neutral measure on $(\Omega,\mathcal{F})$.

Proof First we show that $\mathbb{E}u'(\xi^* \cdot Y) < \infty$, so that $\mathbb{P}^*$ is well defined. Define

$$c := \sup\{u'(x) : x \in D \text{ and } x \in [-|\xi^*|, |\xi^*|]\}.$$

Consider first the case in which $D = \mathbb{R}$. Then, because u' is decreasing, we have $c = u'(-|\xi^*|)$. If $D = [a, \infty)$, then $c \leq \sup\{u'(x) : x \in D\} = u'(a)$. In both cases we have $c < \infty$. By the Cauchy-Schwarz inequality, we have $|\xi^* \cdot Y| \leq |\xi^*| \cdot |Y|$. Hence, if $|\xi^* \cdot Y| > |\xi^*|$, then $|Y| > 1$. From this it follows that (we split into the cases $|\xi^* \cdot Y| \leq |\xi^*|$ and $|\xi^* \cdot Y| > |\xi^*|$ and use that u' is nonnegative)

$$\begin{aligned}
0 \leq u'(\xi^* \cdot Y) &= u'(\xi^* \cdot Y)\mathbf{1}_{\{|\xi^* \cdot Y| \leq |\xi^*|\}} + u'(\xi^* \cdot Y)\mathbf{1}_{\{|\xi^* \cdot Y| > |\xi^*|\}} \\
&\leq c\,\mathbf{1}_{\{|\xi^* \cdot Y| \leq |\xi^*|\}} + u'(\xi^* \cdot Y)\mathbf{1}_{\{|Y|>1\}} \\
&\leq c + u'(\xi^* \cdot Y)\mathbf{1}_{\{|Y|>1\}} \\
&\leq c + u'(\xi^* \cdot Y)|Y|\mathbf{1}_{\{|Y|>1\}} \\
&\leq c + u'(\xi^* \cdot Y)|Y|
\end{aligned}$$

where the expression on the right hand side has finite expectation, by Theorem 6.6.

By definition, a risk-neutral measure satisfies $\mathbb{E}^*Y = 0$. This is indeed the case, since

$$\mathbb{E}^*Y = \mathbb{E}Y\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}} = 0,$$

because of Equation (6.5). □

Remark 6.8 If Y is $\mathbb{P}$-a.s. bounded, then the Radon-Nikodym derivative in (6.6) is bounded and we have constructed a risk neutral measure with bounded density as mentioned in Theorem 1.6. If Y is not bounded under $\mathbb{P}$, one may change the optimization problem by considering $\tilde{Y} = Y/(1 + |Y|)$, which is bounded, instead of Y. Indeed, along with Y, also $\tilde{Y}$ satisfies the no arbitrage condition $\mathbb{P}(\xi \cdot \tilde{Y} \geq 0) = 1 \Rightarrow \mathbb{P}(\xi \cdot \tilde{Y} = 0) = 1$ and vice versa. If $\tilde{\xi}$ is the corresponding maximizer of $\xi \mapsto \mathbb{E}u(\tilde{\xi} \cdot Y)$, then $\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}} = \frac{1}{c}\frac{u'(\tilde{\xi}\cdot\tilde{Y})}{1+|Y|}$ defines a risk-neutral measure $\tilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ for $c = \mathbb{E}\frac{u'(\tilde{\xi}\cdot\tilde{Y})}{1+|Y|}$.

6.2 Exponential utility and relative entropy

In the present section we fix the utility function to be given by $u(x) = 1 - \exp(-\alpha x)$, $x \in \mathbb{R}$ *and* $\alpha > 0$. The optimization problem 6.1 is in this case equivalent to (take $\lambda = -\alpha\xi$) the *minimization* of the function $Z : \mathbb{R}^d \to (0, \infty)$ defined by

$$Z(\lambda) = \mathbb{E}e^{\lambda \cdot Y}.$$

This optimization problem will be studied under the assumption that $Z(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$, for which we have the following equivalent formulation, all exponential moments of $|Y|$ are finite.

Lemma 6.9 *It holds that $Z(\lambda) < \infty$, for all $\lambda \in \mathbb{R}^d$ iff $\mathbb{E}\exp(\alpha|Y|) < \infty$ for all $\alpha \in \mathbb{R}$.*

Proof The condition in terms of α is certainly sufficient. To prove necessity we proceed as follows. We use that $|Y| \leq \sqrt{d}\sum_i |Y_i|$ to get for $\alpha > 0$ (which is sufficient to consider) by Hölder's inequality for d random variables

$$\mathbb{E}e^{\alpha|Y|} \leq \mathbb{E}e^{\alpha\sqrt{d}\sum_i |Y_i|} \leq \prod_i \left(\mathbb{E}e^{\alpha d\sqrt{d}|Y_i|}\right)^{1/d}.$$

Since $\exp(\alpha d\sqrt{d}|Y_i|) \leq \exp(\alpha d\sqrt{d}Y_i) + \exp(-\alpha d\sqrt{d}Y_i)$, each of the factors in the product is finite. □

In the remainder of this section we assume that the condition of Lemma 6.9 holds. Before we proceed with the optimization problem, we introduce some terminology.

Definition 6.10 *The exponential family of $\mathbb{P}$ with respect to Y* is the family of probability measures $\mathbb{P}_\lambda$ on $(\Omega, \mathcal{F})$ with $\lambda \in \mathbb{R}^d$ given by

$$\frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}} = \frac{e^{\lambda\cdot Y}}{Z(\lambda)}.$$

Expectation w.r.t. $\mathbb{P}_\lambda$ is denoted by $\mathbb{E}_\lambda$ and $m(\mathbb{P}_\lambda) := \mathbb{E}_\lambda Y$. Note that all $\mathbb{P}_\lambda$ are mutually equivalent probability measures, and equivalent to $\mathbb{P}$.

We restate Theorem 6.5 and Theorem 6.6 in the present context.

Proposition 6.11 *The function $\lambda \mapsto Z(\lambda)$ takes on its minimum iff the market is arbitrage free. If this happens, any minimizer λ^* also solves the equation*

$$m(\mathbb{P}_{\lambda^*}) = 0.$$

If the market is non-redundant, then the minimizer is unique.

Proof We apply Theorem 6.5, and so a minimizer exists iff the market is free of arbitrage. From Theorem 6.6 we obtain for this case that $m(\mathbb{P}_{\lambda^*}) = 0$. □

Below we will see a converse to this proposition, if $m(\mathbb{P}_{\lambda^*}) = 0$, then λ^* minimizes $\lambda \mapsto Z(\lambda)$.

Definition 6.12 Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$. Denote by $\mathbb{E}_\mathbb{Q}$ expectation under $\mathbb{Q}$. If $\mathbb{Q} \ll \mathbb{P}$, then the *relative entropy*, or *Kullback-Leibler information* of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ is defined as

$$H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}_\mathbb{Q} \log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \leq \infty.$$

If $\mathbb{Q}$ is not absolutely continuous w.r.t. $\mathbb{P}$, then $H(\mathbb{Q}|\mathbb{P}) := \infty$.

Since $x \mapsto x \log x$ is strictly convex and $H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}\,\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$, it follows from Jensen's inequality that always $H(\mathbb{Q}|\mathbb{P}) \geq 0$ and that $H(\mathbb{Q}|\mathbb{P}) = 0$ iff $\mathbb{Q} = \mathbb{P}$.

Proposition 6.13 *Assume there is $\lambda_0 \in \mathbb{R}^d$ such that $m(\mathbb{P}_{\lambda_0}) = 0$.*

(i) For all $\lambda \in \mathbb{R}^d$ it holds that $H(\mathbb{P}_\lambda|\mathbb{P}) = \lambda \cdot m(\mathbb{P}_\lambda) - \log Z(\lambda)$ and $H(\mathbb{P}_{\lambda_0}|\mathbb{P}) = -\log Z(\lambda_0)$.

(ii) If $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{F})$ with $\mathbb{E}_{\mathbb{Q}} Y = 0$, then $H(\mathbb{Q}|\mathbb{P}) = H(\mathbb{Q}|\mathbb{P}_{\lambda_0}) + H(\mathbb{P}_{\lambda_0}|\mathbb{P})$.

(iii) Let $\mathcal{Q}_0$ be the set of probability measures $\mathbb{Q}$ with $\mathbb{E}_{\mathbb{Q}} Y = 0$. Then the mapping $\mathbb{Q} \mapsto H(\mathbb{Q}|\mathbb{P})$ assumes on the set $\mathcal{Q}_0$ a unique minimum for $\mathbb{Q} = \mathbb{P}_{\lambda_0}$.

(iv) λ_0 is the minimizer of $\lambda \mapsto Z(\lambda)$.

Proof (i) By definition of $\mathbb{P}_\lambda$ one has $\log \frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}} = \lambda \cdot Y - \log Z(\lambda)$. The first result then follows, because $\mathbb{E}_\lambda Y = m(\mathbb{P}_\lambda)$, and the second one is a then a consequence of $m(\mathbb{P}_{\lambda_0}) = 0$.

(ii) Clearly, there is only something to prove if all entropies involved are finite. So we assume $\mathbb{Q} \ll \mathbb{P}$, and then we also have $\mathbb{Q} \ll \mathbb{P}_\lambda$. From the product rule

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}_\lambda}\frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}} = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}_\lambda}\frac{e^{\lambda \cdot Y}}{Z(\lambda)}$$

one obtains

$$\log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}_\lambda} + \lambda \cdot Y - \log Z(\lambda).$$

Take expectation under $\mathbb{Q}$ to get

$$\begin{aligned} H(\mathbb{Q}|\mathbb{P}) &= H(\mathbb{Q}|\mathbb{P}_\lambda) + \lambda \cdot \mathbb{E}_{\mathbb{Q}} Y - \log Z(\lambda) \\ &= H(\mathbb{Q}|\mathbb{P}_\lambda) - \log Z(\lambda), \end{aligned} \tag{6.7}$$

since $\mathbb{E}_{\mathbb{Q}} Y = 0$. The result now follows from (i) if we take $\lambda = \lambda_0$.

(iii) Note that $\mathbb{P}_{\lambda_0} \in \mathcal{Q}_0$. It follows from (ii) that $H(\mathbb{Q}|\mathbb{P}) \geq H(\mathbb{P}_{\lambda_0}|\mathbb{P})$ for all $\mathbb{Q} \in \mathcal{Q}_0$. Equality holds iff $H(\mathbb{Q}|\mathbb{P}_{\lambda_0}) = 0$, which happens iff $\mathbb{Q} = \mathbb{P}_{\lambda_0}$.

(iv) Take in (6.7) $\mathbb{Q} = \mathbb{P}_{\lambda_0}$ to obtain

$$H(\mathbb{P}_{\lambda_0}|\mathbb{P}_\lambda) = H(\mathbb{P}_{\lambda_0}|\mathbb{P}) + \log Z(\lambda).$$

Then minimizing $Z(\lambda)$ over λ is equivalent to minimizing $H(\mathbb{P}_{\lambda_0}|\mathbb{P}_\lambda)$. But a minimizer of the latter is λ_0. □

We close this section by connecting the preceding results for portfolio optimization to the construction of a special risk neutral measure.

Corollary 6.14 *Suppose that the market is arbitrage-free under the probability measure $\mathbb{P}$. Then there exists a unique risk-neutral measure $\mathbb{P}^*$ that minimizes the relative entropy $H(\mathbb{P}'|\mathbb{P})$ over all equivalent risk-neutral measures $\mathbb{P}' \in \mathcal{P}$. Specifically, if λ^* is the minimizer of $Z(\lambda)$, then*

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}} = \frac{e^{\lambda^* \cdot Y}}{\mathbb{E} e^{\lambda^* \cdot Y}}. \tag{6.8}$$

Proof We apply Proposition 6.13 together with Proposition 6.11 to obtain the result, upon noticing that for a risk neutral measure $\mathbb{Q}$ one has $\mathbb{E}_{\mathbb{Q}}Y = 0$. □

The assertion of Corollary 6.14 can be restated by saying that the optimal portfolio ξ^* of an optimization problem in terms of a CARA utility function can be characterized by a *relative entropy minimizing probability measure* $\mathbb{P}^*$. This measure, as presented in this corollary, is sometimes called an *exponentially tilted transformation* of $\mathbb{P}$, or an *Esscher transform* of $\mathbb{P}$.

6.3 Exercises

6.1 Show that a function $h : \mathbb{R}^d \to \mathbb{R}$, defined on some domain is upper semicontinuous iff the sets $h^{-1}(-\infty, a)$ are open for all $a \in \mathbb{R}$.

6.2 The assertion of Theorem 6.5 should also be true for a utility function $u : (a, \infty) \to \mathbb{R}$ with $a < 0$, $\lim_{x\downarrow a} u(x) = -\infty$ and u bounded from above. Investigate whether this conjecture is correct. In particular one should check whether h is upper semicontinuous and adapt the proof where needed. It could be useful to extend the definition of u by defining $u(x) = -\infty$ for $x \le a$, since otherwise certain desirable properties of $h(\xi) = \mathbb{E}u(\xi \cdot Y)$ are hard to establish.

6.3 Consider the CARA utility function $u(x) = 1 - \exp(-\alpha x)$, $x \in \mathbb{R}$, with $\alpha > 0$, the constant Arrow-Pratt coefficient.

(a) Show that the condition $\mathbb{E}|u(\xi \cdot Y)| < \infty$ for $\xi \in \Xi$ of Theorem 6.6 is equivalent to $\mathbb{E}\exp(\xi \cdot Y) < \infty$, for all $\xi \in \mathbb{R}^d$.

(b) Show that the risk-neutral measure $\mathbb{P}^*$ of Proposition 6.7 is the same for all $\alpha > 0$.

(c) Suppose that Y has a d-dimensional multivariate normal distribution with mean vector m and invertible covariance matrix Σ. Compute the optimal $\xi^* \in \mathbb{R}^d$.

6.4 Let $\mathbb{P}$ be a probability measure on some measurable space $(\Omega, \mathcal{F})$. Show that the mapping $\mathbb{Q} \mapsto H(\mathbb{Q}|\mathbb{P})$ is strictly convex on the set of all probability measures on this space such that $H(\mathbb{Q}|\mathbb{P}) < \infty$.

6.5 Consider a market with one risky asset only ($d = 1$). Let the assumptions of Theorem 6.6 hold. Let ξ^* be the optimal investment in the risky asset. Show, by inspecting the objective function $\xi \mapsto \mathbb{E}u(\xi \cdot Y)$ (u defined on $\mathbb{R}$ and differentiable) near $\xi = 0$, that $\xi^* > 0$ iff $\mathbb{E}Y > 0$. (This yields an alternative to Example 4.8).

6.6 Consider a market with $r = 0$ and one risky good S_1, that can take on arbitrarily large and arbitrarily small numbers, $\mathbb{P}(S_1 \le \varepsilon) > 0$ and $\mathbb{P}(S_1 \ge 1/\varepsilon) > 0$ for any $\varepsilon > 0$ and assume $\mathbb{E}S_1 < \pi_1$. Let $D = [-\pi_1, \infty)$ and u a sufficiently smooth utility function with domain D. Note that $Y \ge -\pi_1$, hence $u(Y)$ is well defined.

(a) Show that $\Xi = [0, 1]$.

(b) From Example 4.8 one deduces $\xi^* = 0$, a boundary point of Ξ. Show that $\mathbb{E}u'(\xi^* \cdot Y)Y < 0$, violating Condition 6.5.

6.7 Given an arbitrage free market and a utility function $\tilde{u}$ as at the beginning of this section, the transformed utility function u depends on the initial capital w. In general, an optimal portfolio will also depend on w. We study this for the case $d = 1$ and $r = 0$. We avoid redundancy of the market, Y is non-degenerate. Assume that $\tilde{u}$ is a C^2 function and that everywhere below interchanging of expectation and differentiation is allowed. Put $f(w, \xi) = \mathbb{E}\tilde{u}'(\xi Y + w)Y$.

(a) Show that $\frac{\partial f}{\partial \xi}(w, x) < 0$.

(b) Conclude that (locally) for every $w > 0$, there is a C^1 function $w \mapsto \xi^*(w)$ such that $f(w, \xi^*(w)) = 0$.

(c) Show that

$$\frac{\mathrm{d}\xi^*(w)}{\mathrm{d}w} = -\frac{\mathbb{E}\tilde{u}''(\xi^*(w)Y + w)Y}{\mathbb{E}\tilde{u}''(\xi^*(w)Y + w)Y^2}.$$

(d) Assume that $\mathbb{E}Y > 0$ and that Arrow-Pratt coefficient $\tilde{\alpha}(\cdot)$ of $\tilde{u}$ is a decreasing function. Show that $Y\tilde{\alpha}(\xi^*(w)Y + w) \leq Y\tilde{\alpha}(w)$.

(e) Conclude, under the assumptions in (d), that $\xi^*(\cdot)$ is an increasing function of w. (In Micro-economics, assets with the latter property are called *normal goods*. Assets with *decreasing demand* ξ^* are called *inferior goods*.)

6.8 Lemma 6.3 also has a converse. If h is strictly concave and upper semi-continuous, then existence of a minimizer implies (6.4). Show this and give an example that shows that the *strict* concavity of h cannot be missed here.

6.9 Consider a market with one risky good, its value at $t = 1$ is S and price π (at $t = 0$). Assume that S has under $\mathbb{P}$ a Poisson distribution with parameter $\alpha > 0$. Consider the exponential family of Definition 6.10.

(a) Show that $Z(\lambda) < \infty$ for all $\lambda \in \mathbb{R}$

(b) Show that S has a Poisson distribution with parameter αe^λ under $\mathbb{P}_\lambda$.

(c) Compute the minimizer of $\lambda \mapsto Z(\lambda)$ directly.

(d) Verify that the minimizer is in agreement with Proposition 6.13.

6.10 Assume that the market is free of arbitrage and complete. By Theorem 1.22 there exists a unique equivalent martingale measure. Let u be a utility function and consider the situation of Proposition 6.7. The proposition tells what the equivalent martingale measure is, but it depends on the utility function u, which looks paradoxal in view of the uniqueness assertion of Theorem 1.22. How to reconcile the results of Theorem 1.22 and Proposition 6.7 for a complete market?

7 Optimal contingent claims

In the previous sections we studied the problem of portfolio optimization. In a complete market, every contingent claim (recall Definition 1.10) has the same pay-off as some portfolio, but in an incomplete market this is no longer true. Therefore, in the latter case, it makes sense to study, as a new, more general problem, the maximization of the expected utility $\mathbb{E}u(X)$, where u is a utility function and X some contingent claim, belonging to some suitable convex set $\mathcal{X}$. The specification of $\mathcal{X}$ will depend on the context.

7.1 An expected utility optimization problem

Let w be the initial capital of some investor. Let $\mathbb{P}^*$ be a probability measure, equivalent to $\mathbb{P}$ with Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}} = \phi$. Given a contingent claim with *discounted pay-off* X, define a *pricing rule* by $\mathbb{E}^*X = \mathbb{E}\phi X$.

In this context, ϕ is also called a *pricing kernel*. The *budget constraint* on claims X in the present context is given by $\mathbb{E}^*X \leq w$. We introduce the *budget set*, also called the set of *admissible pay-offs*,

$$\mathcal{B} = \{X \in \mathcal{X} \cap L^1(\Omega, \mathcal{F}, \mathbb{P}^*) : \mathbb{E}^*X \leq w\}.$$

Note that $\mathcal{B}$ is a convex set. We will study the following

Problem 7.1 Maximize $\mathbb{E}u(X)$ over the set $\mathcal{B}$.

To have this problem well-defined, we need that $\mathbb{P}^* \sim \mathbb{P}$, not just $\mathbb{P}^* \ll \mathbb{P}$ or $\mathbb{P} \ll \mathbb{P}^*$, see Exercise 7.1, the standing assumption for this section. Note that this relates to Theorem 6.5, where a similar situation has been encountered and where we required the market to be arbitrage-free, which is equivalent to the existence of a Risk neutral measure. We will also assume that $\mathbb{P}(X \in D) = 1$, for all $X \in \mathcal{B}$, where D is the domain of u.

If Problem 7.1 has a solution, it is necessarily unique. This is due to the fact that u is strictly concave, see the proof of Theorem 6.5. Another fact that we encountered before is, that it can never be optimal to invest less than the initial capital w. Indeed, if a given claim $X \in \mathcal{B}$ is such that $\mathbb{E}^*X < w$, then $X' = X + w - \mathbb{E}^*X > X$ and so $\mathbb{E}u(X') > \mathbb{E}u(X)$, whereas $X' \in \mathcal{B}$, since $\mathbb{E}^*X' = w$.

For the time being, we drop the budget constraint and let $\mathcal{X}$ be the set of all random variables. Suppose that X^* is the optimal claim. Let X be any bounded random variable and consider the 'perturbed' claims, belonging to $\mathcal{B}$ as well,

$$X_\lambda = X^* + \lambda(X - \mathbb{E}^*X), \lambda \in \mathbb{R}.$$

Among the pay-offs X_λ, the optimal one is found for $\lambda = 0$. Hence, assuming differentiability where needed, we should consider the equation

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}u(X_\lambda) = 0, \text{ for } \lambda = 0.$$

Scrupulously interchanging differentiation and expectation in the above equation yields

$$\begin{aligned} 0 &= \mathbb{E}(u'(X^*)(X - \mathbb{E}^* X)) \\ &= \mathbb{E}X(u'(X^*) - \phi\,\mathbb{E}u'(X^*)). \end{aligned}$$

Let $c = \mathbb{E}u'(X^*)$, then the above identity yields

$$\mathbb{E}Xu'(X^*) = c\,\mathbb{E}(X\phi),$$

valid for all bounded X, in particular for $X = \mathbf{1}_F$, $F \in \mathcal{F}$. In that case one has

$$\mathbb{E}\mathbf{1}_F(u'(X^*) - c\,\phi) = 0.$$

Take $F = \{u'(X^*) - c\,\phi > 0\}$. Then combining $\mathbf{1}_F(u'(X^*) - c\,\phi) \geq 0$ with the last display, one obtains, assuming invertibility of u',

$$u'(X^*) = c\,\phi \text{ a.s.},$$

yielding $X^* = (u')^{-1}(c\,\phi)$ a.s. Hence we have found a candidate solution, $(u')^{-1}(c\,\phi)$. The heuristics above are justified by the following theorem.

Theorem 7.2 *Suppose that $u \in C^1(\mathbb{R})$, $\lim_{x\to-\infty} u'(x) = \infty$ and u bounded from above. Let I be the inverse of the function u', which is well defined on $(0,\infty)$. Let $c > 0$ be given and $X^* := I(c\,\phi)$. Then X^* is well defined a.s. Moreover, assume that $X^* \in L^1(\Omega, \mathcal{F}, \mathbb{P}^*)$ and let $w = \mathbb{E}^* X^*$ (which depends on c). Then X^* is the unique maximizer of Problem 7.1 for $\mathcal{X} = L^0(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof We have already discussed uniqueness and so we turn to existence. It follows from the assumptions that $u'(x) \to 0$ for $x \to \infty$. Hence every positive number is in the range of u'. Since $\mathbb{P}^* \sim \mathbb{P}$, we have that $\mathbb{P}(\phi > 0) = 1$. Hence $I(c\,\phi)$ is $\mathbb{P}$-a.s. well-defined.

Concavity of u yields for any $X \in L^1(\Omega, \mathcal{F}, \mathbb{P}^*)$ that

$$u(X) \leq u(X^*) + u'(X^*)(X - X^*) = u(X^*) + c\,\phi(X - X^*). \tag{7.1}$$

Taking expectations in this inequality yields

$$\begin{aligned} \mathbb{E}u(X) &\leq \mathbb{E}u(X^*) + c\,\mathbb{E}\phi(X - X^*) \\ &= \mathbb{E}u(X^*) + c\,\mathbb{E}^*(X - X^*) \\ &= \mathbb{E}u(X^*) + c\,(\mathbb{E}^* X - w) \\ &\leq \mathbb{E}u(X^*), \end{aligned}$$

which shows that X^* is the maximizer. □

Let W be a nonnegative random variable with values in $[0,\infty]$. Until further notice we assume that the set $\mathcal{X}$ is that of random variables X satisfying $0 \leq X \leq W$ a.s. and $\mathcal{B} = \{X \in \mathcal{X} : \mathbb{E}^* X \leq w\}$. In this case we assume that $u : [0,\infty) \to \mathbb{R}$.

Remark 7.3 One can show that for any utility function u, there exists a maximizer $X^* \in \mathcal{B}$ of $\mathbb{E}u(X)$ under the conditions stipulated above. For a constructive result the class of utility functions under consideration will be narrowed down in what follows by assuming differentiability.

Let $u \in C^1(0,\infty)$ be a utility function. We can extend the domain of u to $[0,\infty)$, by setting $u(0) = \lim_{x\to 0} u(x) \geq -\infty$. Since u' is decreasing, the limits

$$a = \lim_{x\to\infty} u'(x)$$

and

$$b = \lim_{x\to 0} u'(x)$$

exist. Moreover, we have $0 \leq a < b \leq \infty$ and $a = \inf\{u'(x) : x > 0\}$, $b = \sup\{u'(x) : x > 0\}$. On the open interval (a,b), the function u' has a well-defined continuous and decreasing inverse I. We extend I to a function $I^+ : [0,\infty] \to [0,\infty]$, by setting

$$I^+(y) = \begin{cases} +\infty & \text{if } 0 \leq y \leq a \\ I(y) & \text{if } a < y < b \\ 0 & \text{if } y \geq b. \end{cases}$$

It is obvious that I^+ is decreasing and continuous on $[0,\infty]$.

Theorem 7.4 *Consider the optimization problem under the restriction* $0 \leq X \leq W \leq \infty$. *Let* $X^* = I^+(c\,\phi) \wedge W$ *and let* $w = \mathbb{E}^* X^* < \infty$. *If* $\mathbb{E}u(X^*) < \infty$, *then* X^* *is the unique maximizer of* $\mathbb{E}u(X)$ *over* $X \in \mathcal{B}$.

Proof Consider the function $v : [0,\infty] \times \Omega \to \mathbb{R}$ (the Legendre-Fenchel transform of u, a common tool in convex analysis) defined by

$$v(y,\omega) = \sup\{u(x) - xy : 0 \leq x \leq W(\omega)\}. \tag{7.2}$$

Suppose, for the time being, that $W(\omega) < \infty$. By continuity of u, for each y and ω the supremum will be attained at some $x^* = x^*(y,\omega) \in [0, W(\omega)]$, which is unique by strict concavity of u. So we then have (also observe the similarity with (7.1) to understand the use)

$$u(x) \leq xy + u(x^*) - x^* y,$$

where the right hand side as a function of x has as a graph a line with slope y. We discern three cases.

Suppose $x^*(y,\omega) = 0$ (first case). Then for all $x \in (0, W(\omega)]$ we have $u(x) - xy < u(0)$. Consider $\Delta_h u(x) := \frac{u(x+h)-u(x)}{h}$ for $x > 0$ and $x + h > 0$. By concavity of u, $\Delta_h u$ is decreasing for every fixed h, in particular $\Delta_h u(x) \leq \Delta_h u(0) = \frac{u(h)-u(0)}{h} \leq y$. It follows that $\sup_{x>0} u'(x) \leq y$ and hence $y \geq b$. Conversely, if $y \geq b$, then $u'(x) < y$ for all $x \in (0, W(\omega))$, then $x \mapsto u(x) - xy$ is decreasing and $x^*(y,\omega) = 0$. One similarly shows that (second case) $x^*(y,\omega) =$

$W(\omega)$ iff $y \le a$. If $x^*(y,\omega) \in (0, W(\omega))$ is an interior maximizer (third case), then $y = u'(x^*(y,\omega))$ holds and $y \in (a,b)$, so $x^*(y,\omega) = I(y)$. Taking all three cases into account, one arrives at

$$x^*(y,\omega) = I^+(y) \wedge W(\omega),$$

which is a measurable function, jointly in (y,ω), since I^+ is continuous. Hence, by definition of X^*, one has $X^*(\omega) = x^*(c\phi(\omega),\omega)$ on $\{\omega : W(\omega) < \infty\}$, measurable in ω. Suppressing the dependence on ω, we thus found $X^* = x^*(c\phi)$ on $\{W < \infty\}$.

Next we switch from $W(\omega) < \infty$ to $W(\omega) = \infty$. The case where a finite maximizer $x^*(y,\omega)$ exists can be treated as before and one gets $x^*(y,\omega) = I^+(y) = I^+(y) \wedge W(\omega)$. The supremum in (7.2) is not attained for a finite argument iff $u'(x) > y$ for all $x > 0$, and then $y \le a$. By definition of I^+, we can put $x^*(y,\omega) = I^+(y) = I^+(y) \wedge W(\omega)$. On the other hand, the assumption $\mathbb{E}^* X^* < \infty$ implies $w \ge \mathbb{E}^* X^* \mathbf{1}_{\{W=\infty\}} = \mathbb{E}^* I^+(c\,\phi)\mathbf{1}_{\{W=\infty\}}$. It follows that $X^* = x^*(c\,\phi)$ is finite $\mathbb{P}^*$-a.s. on $\{W = \infty\}$.

Hence, in both situations $W < \infty$ and $W = \infty$, one obtains $X^* = x^*(c\,\phi)$, which is a $\mathbb{P}^*$-a.s. finite random variable, and then also $\mathbb{P}$-a.s. finite. But then, using the definition of x^* as the maximizer of (7.2), we get for arbitrary $X \in \mathcal{B}$, similar to (7.1),

$$u(X^*) - c\,\phi X^* \ge u(X) - c\,\phi X \text{ a.s.}$$

Take expectations to get

$$\mathbb{E}u(X^*) - c\,\mathbb{E}\phi X^* \ge \mathbb{E}u(X) - c\,\mathbb{E}\phi X,$$

equivalent to

$$\mathbb{E}u(X^*) - \mathbb{E}u(X) \ge c(\mathbb{E}^* X^* - \mathbb{E}^* X),$$

with a nonnegative right hand side by $X \in \mathcal{B}$. The uniqueness issue has already been addressed before. □

The previous theorems dealt with the existence of an optimizer for problems where the initial capital w was defined in terms of a property of the candidate optimizer, involving the constant c that was also depending on the candidate optimizer. In a practical situation, w is given before hand and so we can apply the previous theorems only if c is such that the assumptions in these theorems are met. The next corollary gives a simple sufficient condition for this.

Corollary 7.5 *Let $w > 0$ be given and assume $0 < w < \mathbb{E}^* W < \infty$ and $\mathbb{E}u(W) < \infty$. Then there exists a unique constant $c^* > 0$ such that $X^* := I^+(c^*\,\phi) \wedge W$ satisfies $\mathbb{E}^* X^* = w$. Hence X^* is the maximizer of $\mathbb{E}u(X)$ over $X \in \mathcal{B}$.*

Proof Let $\beta > 0$ and define f_β by $f_\beta(y) = I^+(y) \wedge \beta$. Then f_β is bounded, continuous and decreasing. Moreover, $\lim_{y\uparrow b} f_\beta(y) = 0$ and $f_\beta(y) = \beta$ for $y \le u'(\beta)$. Put $g(c) = \mathbb{E}^* f_W(c\,\phi) = \mathbb{E}^*[I^+(c\,\phi) \wedge W] \le \mathbb{E}^* W$. By dominated convergence, g is continuous. Furthermore, $\lim_{c\to\infty} g(c) = 0$, $\lim_{c\downarrow 0} g(c) = \mathbb{E}^*W$ and g is *strictly* decreasing (Exercise 7.5) on the interval $g^{-1}[(0, \mathbb{E}^*W)]$, which contains w by assumption. We thus obtain that there exists a unique c^* such that $w = g(c^*)$. Theorem 7.4 yields that X^* is the expected utility maximizer, as $\mathbb{E}u(X^*) < \infty$ is guaranteed by $\mathbb{E}u(W) < \infty$. □

7.2 Optimization under uniform order restrictions

In this section we study an optimization problem, that involves the uniform order $\succeq_{\text{uni}}$, recall Definition 5.1. We transplant this order to the space of random variables, all defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, by saying that $X \succeq_{\text{uni}} Y$ iff $\mu_X \succeq_{\text{uni}} \mu_Y$, where μ_X and μ_Y denote the laws (under $\mathbb{P}$) of X and Y respectively. Note that this simply means $\mathbb{E}u(X) \geq \mathbb{E}u(Y)$ for all utility functions u. The problem we are going to address is the following.

Problem 7.6 Let $\mathbb{P}^* \sim \mathbb{P}$ and $X_0 \in \mathcal{X} = L^1_+(\Omega, \mathcal{F}, \mathbb{P})$ be given. Note that $X_0 \geq 0$ a.s. under $\mathbb{P}$ and $\mathbb{P}^*$ and assume that $\mathbb{E}^*X_0 < \infty$. The objective is to minimize $\mathbb{E}^*X$ over all random variables $X \in \mathcal{X}$ satisfying $X \succeq_{\text{uni}} X_0$.

The interpretation of this problem is that one wants to find the minimal budget needed among all X that are at least as attractive as X_0 in the sense that $X \succeq_{\text{uni}} X_0$. Note that the latter requirement is stated in terms of $\mathbb{P}$, whereas we want to find a minimal expectation under $\mathbb{P}^*$.

Before we state a theorem with the solution to this problem, we need some additional properties of the $\succeq_{\text{uni}}$ order in terms of *quantile functions*. Recall the notation $f(x-) = \lim_{y\uparrow} f(y)$ for $f : \mathbb{R} \to \mathbb{R}$, assuming that the limit exists.

Definition 7.7 If F is a distribution function, then $q : (0, 1) \to \mathbb{R}$ is called a *quantile function* for F if for all $t \in (0, 1)$ it holds that $F(q(t)-) \leq t \leq F(q(t))$. If X is a random variable with distribution function F, we also say that q is a quantile function for X if q is a quantile function of F. Such a quantile function is also denoted q_F and q_X.

Recall that there are two 'extremal' quantile functions, q^- and q^+, defined by $q^-(t) = \sup\{x \in \mathbb{R} : F(x) < t\}$ and $q^+(t) = \sup\{x \in \mathbb{R} : F(x) \leq t\}$. Recall also the fundamental equivalences $q^-(t) \leq x \Leftrightarrow t \leq F(x)$ and $q^+(t) < x \Leftrightarrow t < F(x)$. Moreover, $q^+ = q^-$ a.e. w.r.t. Lebesgue measure. Since any quantile function q satisfies $q^- \leq q \leq q^+$, we also have $q = q^- = q^+$ a.e. w.r.t. Lebesgue measure, and hence integrals of these functions w.r.t. Lebesgue measure have the same value. In particular, if U has a uniform distribution on $(0, 1)$, $q(U)$ has distribution function F for any q that is a quantile function for F.

Lemma 7.8 *Let F be a distribution function of a distribution with finite mean and q an associated quantile function. Then for all $x \in \mathbb{R}$ it holds that*

$$xF(x) = \int_{-\infty}^{x} F(u)\,\mathrm{d}u + \int_0^{F(x)} q(u)\,\mathrm{d}u. \tag{7.3}$$

Moreover, for arbitrary $x \in \mathbb{R}$ and $t \in (0,1)$, one has

$$xt \leq \int_{-\infty}^{x} F(u)\,\mathrm{d}u + \int_0^t q(u)\,\mathrm{d}u. \tag{7.4}$$

Proof Both relations follow from maximizing $x \mapsto xt - \int_{-\infty}^{x} F(u)\,\mathrm{d}u$ (Exercise 7.7). □

Remark 7.9 Inequality (7.4) is also valid for $t = 0, 1$ and if the distribution doesn't have a finite mean, in which case the right hand may be infinite.

Lemma 7.10 *Let μ, ν be probability measures on $\mathbb{R}$ and let q_μ and q_ν be corresponding quantile functions. The following statements are equivalent.*

(i) $\mu \succeq_{\mathrm{uni}} \nu$.

(ii) For all $t \in (0,1)$, it holds that $\int_0^t q_\mu(s)\,\mathrm{d}s \geq \int_0^t q_\nu(s)\,\mathrm{d}s$.

(iii) For all decreasing functions $h : (0,1) \to [0,\infty)$ it holds that

$$\int_0^1 h(s) q_\mu(s)\,\mathrm{d}s \geq \int_0^1 h(s) q_\nu(s)\,\mathrm{d}s. \tag{7.5}$$

(iv) For all bounded decreasing functions $h : (0,1) \to [0,\infty)$ inequality (7.5) holds true.

Proof (i) $\Leftrightarrow$ (ii) follows from Lemma 7.8 and Theorem 5.3 (Exercise 7.8).

(ii) $\Rightarrow$ (iii): Since h is decreasing, it has at most countably many discontinuities, so the integrals in (7.5) don't change if we replace h with its right-continuous modification. Then, up to an additive positive constant, h can be seen as the 'complement of a distribution function' of a measure η on $(0,1)$, $h(t) = \eta(t,1)$. We apply Fubini's theorem as in the proof of Theorem 5.3. We have

$$\begin{aligned} \int_0^1 h(t) q_\mu(t)\,\mathrm{d}t &= \int_0^1 \int_{(t,1)} \eta(\mathrm{d}s)\, q_\mu(t)\,\mathrm{d}t \\ &= \int_{(0,1)} \int_0^s q_\mu(t)\,\mathrm{d}t\,\eta(\mathrm{d}s) \\ &\geq \int_{(0,1)} \int_0^s q_\nu(t)\,\mathrm{d}t\,\eta(\mathrm{d}s) \\ &= \int_0^1 h(s) q_\nu(s)\,\mathrm{d}s. \end{aligned}$$

(iii) $\Rightarrow$ (iv): trivial.

(iv) $\Rightarrow$ (ii): Take $h = \mathbf{1}_{(0,t]}$. □

Lemma 7.11 *Let X, Y be nonnegative random variables. Then*

$$\mathbb{E}XY \geq \int_0^1 q_X(1-t) q_Y(t)\;\mathrm{d}t,$$

where q_X and q_Y are quantile functions for X and Y respectively.

Proof First we note that by Fubini it holds that

$$\mathbb{E}XY = \mathbb{E}\int_0^\infty \int_0^\infty \mathbf{1}_{\{x<X,y<Y\}}\,\mathrm{d}x\,\mathrm{d}y = \int_0^\infty \int_0^\infty \mathbb{P}(X>x, Y>y)\,\mathrm{d}x\,\mathrm{d}y. \tag{7.6}$$

Next we have the trivial relations

$$\begin{aligned}\mathbb{P}(X>x,Y>y) &= \mathbb{P}(X>x) - \mathbb{P}(X>x, Y\le y)\\ &\geq \mathbb{P}(X>x) - \mathbb{P}(Y\le y).\end{aligned}$$

Since $\mathbb{P}(X>x,Y>y)\geq 0$, we also have

$$\mathbb{P}(X>x,Y>y) \geq \max\{\mathbb{P}(X>x) - \mathbb{P}(Y\le y), 0\}. \tag{7.7}$$

For $F_Y(y)\le 1-F_X(x)$ we have, using the special property of q^+,

$$\begin{aligned}0\le \mathbb{P}(X>x)-\mathbb{P}(Y\le y) &= \int_0^1 \mathbf{1}_{\{F_Y(y)\le t\le 1-F_X(x)\}}\,\mathrm{d}t\\ &= \int_0^1 \mathbf{1}_{\{y\le q_Y^+(t), x\le q_X^+(1-t)\}}\,\mathrm{d}t.\end{aligned}$$

In case $F_Y(y)>1-F_X(x)$, the integrand on the right hand side is zero, hence we can replace the left hand side with $(\mathbb{P}(X>x)-\mathbb{P}(Y\le y))^+$, whatever x and y, meanwhile maintaining the integral expression, so

$$(\mathbb{P}(X>x)-\mathbb{P}(Y\le y))^+ = \int_0^1 \mathbf{1}_{\{y\le q_Y^+(t), x\le q_X^+(1-t)\}}\,\mathrm{d}t.$$

Integrating the right hand side with respect to x and y yields by Fubini's theorem

$$\int_0^1\int_0^\infty\int_0^\infty \mathbf{1}_{\{y\le q_Y^+(t), x\le q_X^+(1-t)\}}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t = \int_0^1 q_X^+(1-t)q_Y^+(t)\,\mathrm{d}t,$$

which, upon invoking (7.6) and (7.7), proves the assertion, since all quantile functions are Lebesgue-a.e. the same. □

Theorem 7.12 *Let ϕ denote the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}$. Consider problem 7.6. If $X\in\mathcal{X}$ satisfies $X\succeq_{\mathrm{uni}} X_0$, then*

$$\mathbb{E}^*X \geq \int_0^1 q_\phi(1-s)q_{X_0}(s)\,\mathrm{d}s. \tag{7.8}$$

Let μ_ϕ be the law of ϕ under $\mathbb{P}$ and F_ϕ its distribution function. If ν is the measure on $(\mathbb{R},\mathcal{B})$ characterized by $\nu(-\infty,x]=\int_0^{F_\phi(x)} q_{X_0}(1-t)\,\mathrm{d}t$, then $\nu\ll\mu_\phi$ and equality in (7.8) holds for $X=X^*:=f(\phi)$ with $f=\frac{\mathrm{d}\nu}{\mathrm{d}\mu_\phi}$. Moreover, $X^*\succeq_{\mathrm{uni}} X_0$ and hence X^* is the minimizer sought for. The function f has the following explicit expression.

$$f(x) = \begin{cases} q_{X_0}(1-F_\phi(x)) & \text{if } x \text{ is a continuity point of } F_\phi\\ \dfrac{\int_{F_\phi(x-)}^{F_\phi(x)} q_{X_0}(1-t)\,\mathrm{d}t}{F_\phi(x)-F_\phi(x-)} & \text{else.}\end{cases} \tag{7.9}$$

Proof We will use the auxiliary probability space $((0,1),\mathcal{B}(0,1),\lambda)$, where λ denotes Lebesgue measure. Expectations and conditional expectations w.r.t. λ will be denoted by expressions like $\mathbb{E}_\lambda(X)$, $\mathbb{E}_\lambda[X|Y]$. Furthermore let U be the identity mapping on $(0,1)$; it has the uniform distribution on $(0,1)$. Then $\tilde{\phi} = q_\phi(U)$ has the same distribution as ϕ. As a conditional expectation given a random variable (here $\tilde{\phi}$) is measurable w.r.t. the σ-algebra generated by that random variable, Proposition B.10 yields the existence of a Borel-measurable function f such that

$$\mathbb{E}_\lambda[q_{X_0}(1-U)|\tilde{\phi}] = f(\tilde{\phi}),\ \lambda\text{-a.s.} \tag{7.10}$$

Put $X^* = f(\phi)$, then $X^* \stackrel{d}{=} f(\tilde{\phi})$. Let u be a utility function, hence concave. On has, using Jensen's inequality for conditional expectations,

$$\begin{aligned}
\mathbb{E}u(X^*) &= \mathbb{E}_\lambda(u(f(\tilde{\phi}))) \\
&= \mathbb{E}_\lambda(u(\mathbb{E}_\lambda[q_{X_0}(1-U)|\tilde{\phi}])) \\
&\geq \mathbb{E}_\lambda(\mathbb{E}_\lambda[u(q_{X_0}(1-U))|\tilde{\phi}]) \\
&= \mathbb{E}_\lambda(u(q_{X_0}(1-U)) \\
&= \mathbb{E}u(X_0),
\end{aligned}$$

which shows that $X^* \succeq_{\text{uni}} X_0$. Likewise we compute

$$\begin{aligned}
\mathbb{E}^*X^* = \mathbb{E}(X^*\phi) = \mathbb{E}(f(\phi)\phi) &= \mathbb{E}_\lambda(f(\tilde{\phi})\tilde{\phi}) \\
&= \mathbb{E}_\lambda(\mathbb{E}_\lambda[q_{X_0}(1-U)|\tilde{\phi}]\tilde{\phi}) \\
&= \mathbb{E}_\lambda(\mathbb{E}_\lambda[q_{X_0}(1-U)\tilde{\phi}|\tilde{\phi}]) \\
&= \mathbb{E}_\lambda(q_{X_0}(1-U)\tilde{\phi}) \\
&= \mathbb{E}_\lambda(q_{X_0}(1-U)q_\phi(U)) \\
&= \int_0^1 q_{X_0}(1-t)q_\phi(t)\ \mathrm{d}t.
\end{aligned}$$

For any $X \geq 0$, one has $\mathbb{E}^*X = \mathbb{E}(X\phi) \geq \int_0^1 q_X(t)q_\phi(1-t)\,\mathrm{d}t$ by virtue of Lemma 7.11. If moreover $X \succeq_{\text{uni}} X_0$, we obtain from Lemma 7.10 (applied by choosing $h(t) = q_\phi(1-t)$), that $\mathbb{E}^*X \geq \int_0^1 q_{X_0}(t)q_\phi(1-t)\,\mathrm{d}t = \mathbb{E}^*X^*$. We conclude that X^* is indeed the optimizer.

It remains to identify the function f as the Radon-Nikodym derivative $\frac{\mathrm{d}\nu}{\mathrm{d}\mu_\phi}$ and in terms of the quantile function q_{X_0} and the distribution function F_ϕ. This is the content of Exercise 7.10. □

7.3 Exercises

7.1 Let X be a bounded random variable. Suppose that $\mathbb{P}$ is not absolutely continuous w.r.t $\mathbb{P}^*$. Then there exists $F \in \mathcal{F}$ such that $\mathbb{P}^*(F) = 0$ and $\mathbb{P}(F) > 0$. Put $X_1 = X + c\,\mathbf{1}_F$. Show that X_1 'performs better than X', i.e. it gives higher expected utility under the same price. Find also an example of this phenomenon for the case where $\mathbb{P}^*$ is not absolutely continuous w.r.t. $\mathbb{P}$.

7.2 Let $u(x) = 1 - e^{-\alpha x}$, $\alpha > 0$ and assume that $H(\mathbb{P}^*|\mathbb{P}) < \infty$.

(a) Determine I and show that $\mathbb{E}^* I(c\,\phi) = -\frac{1}{\alpha}(\log\frac{c}{\alpha} + H(\mathbb{P}^*|\mathbb{P}))$.

(b) Compute X^* for problem 7.1 for a given initial capital w.

(c) Let $\mathbb{P}^*$ be the probability measure of Corollary 6.14 (and let $\lambda^* = -\alpha\xi^*$). Show that in this case $X^* = \frac{\bar{\xi}^* \cdot \bar{S}}{1+r}$, where $\bar{\xi} = (\xi_0, \xi)$ for some ξ_0 (which one?).

7.3 This exercise concerns the case where $W = \infty$ (see Theorem 7.4). Consider the CARA utility function $u(x) = -e^{-\alpha x}$.

(a) Show that

$$I^+(y) = (-\frac{1}{\alpha}\log\frac{y}{\alpha})^+$$

for $y \in [0, \infty]$.

(b) Show that the function $g : (0, \infty) \to (0, \infty]$ defined by $g(y) = \mathbb{E}^* I^+(y\,\phi)$ is decreasing and continuous on the set where it is finite and $\lim_{y\downarrow 0} g(y) = +\infty$, $\lim_{y\to\infty} g(y) = 0$.

(c) Let $\mathbb{P}^*$ be the risk-neutral measure of Proposition 6.7 and consider the optimization problem addressed in that proposition. Show that the optimal X^* is now of the form $X^* = (\xi^* \cdot Y - K)^+$ (a kind of European call option), where

$$K = \frac{1}{\alpha}\log\frac{c}{\alpha} + \frac{1}{\alpha}H(\mathbb{P}^*|\mathbb{P}).$$

7.4 In the setting of Theorem 7.4, let $W = \infty$ and let $u = u_{1,0}$ be a HARA utility function with index $\gamma \in [0, 1)$, see Example 4.11.

(a) Let $\gamma = 0$, $u(x) = \log x$. Show that for given $w > 0$ the optimal X^* is given by $X^* = w\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^*}$ and that the maximal expected utility equals $\log w + H(\mathbb{P}|\mathbb{P}^*)$ (assume that this is finite).

(b) Let $\gamma \in (0, 1)$. Compute the optimal X^* for this case.

7.5 Show that the function g in the proof of Corollary 7.5 is strictly decreasing on $g^{-1}[(0, \mathbb{E}^* W)]$.

7.6 Investigate whether the assertion of Corollary 7.5 continues to hold for the case where $W = \infty$ and $0 < w < \infty$. Impose additional assumptions (on u for instance as in Theorem 7.2), if needed.

7.7 Prove Lemma 7.8. (*Depending on the proof, it may be convenient to distinguish between $x \geq 0$ and $x < 0$. It is always a good idea to interpret integrals as areas, and to make a sketch.*)

7.8 Show the equivalence (i) $\Leftrightarrow$ (ii) of Lemma 7.10.

7.9 Let X^* be the optimal random variable of Theorem 7.12. Show that $\mathbb{E}X^* = \mathbb{E}X_0$. Are the laws of X^* and X_0 the same under $\mathbb{P}$? What is X^* if it happens that $\mathbb{P}^* = \mathbb{P}$? Is there an intuitive explanation for this?

7.10 Here you prove the remaining assertions of Theorem 7.12. Let $\nu(B) = \mathbb{E}_\lambda[\mathbf{1}_B(q_\phi(U))q_{X_0}(1-U)]$, $B \in \mathcal{B}(\mathbb{R})$ and let μ_ϕ be the distribution of ϕ.

(a) Show that $\nu \ll \mu_\phi$ and that $\nu(\mathbb{R}) = \mathbb{E}X_0$.

(b) Let f be as in (7.10). Show that (up to sets of Lebesgue measure zero) it holds that $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu_\phi}$.

(c) Identify f as given in Equation (7.9).

7.11 Give a concrete example where the X^* in Theorem 7.12 is different from X_0.

7.12 Let F be a distribution function and q any of its quantile functions. Let q^- and q^+ be the extremal quantile functions and note that $q^- \leq q^+$.

(a) Show that $\{q^- = q = q^+\}$ has Lebesgue measure one. You may use Theorem 3.10 of the MTP lecture notes.

(b) If U is a random variable with the uniform distribution on $(0,1)$, show that $q(U)$ has distribution function F.

7.13 Consider a random variable X that has a Binomial distribution with parameters $n = 2$ and $p = \frac{1}{2}$.

(a) Compute a quantile function $q : (0,1) \to \mathbb{R}$ for this case.

(b) Give an explicit expression for the integral $\int_0^x F(u)\,\mathrm{d}u$ and find an $x^* = x^*(t)$ (is it unique?) which is the maximizer of $x \mapsto xt - \int_0^x F(u)\,\mathrm{d}u$ for $t \in (0,1)$.

(c) Give an explicit expression for the integral $\int_0^t q(u)\,\mathrm{d}u$ and find a $t^* = t^*(x)$ (is it unique?) which is the maximizer of $t \mapsto xt - \int_0^t q(u)\,\mathrm{d}u$ for $x \in \mathbb{R}$.

(d) Verify that the relations (7.3) and (7.4) hold.

7.14 Consider the simpler version of Problem 7.1, where the set Ω is finite, $\Omega = \{\omega_1, \ldots, \omega_n\}$. Assume that the probabilities $p_j = \mathbb{P}(\{\omega_j\})$ are all positive, as well as the risk neutral probabilities p_j^*. The random variables X can be represented by vectors $(x_1, \ldots, x_n)$. Show by using the Langrange function $L = \mathbb{E}u(X) - \lambda\,\mathbb{E}^*X$ that the optimal x_j^* satisfy $u'(x_j^*) = c\frac{p_j^*}{p_j}$ for some constant c (which one?). Compare to Theorem 7.2.

8 Dynamic arbitrage theory

We return to the setting of Section 1 in the sense that we will work with a market of $d+1$ assets, of which one is often taken to be non-risky. The crucial difference though, is that we will work with *dynamic models*. That is, prices will be given by *stochastic processes* with a non-trivial time set. So, instead of working only with times $t = 0$, where all random quantities involved are deterministic (known) and a time $t = 1$, where prices of risky assets are understood as random variables, we will consider processes with a time index $t \in \{0, 1, \ldots, T\}$, where T is some fixed integer greater than (or equal to) one.

We denote by S_t the d-dimensional random vector representing the nonnegative prices of the risky assets at time t. The quantities S_t^0 will be the prices of the non-risky asset at times t. Usually we take the S_t^0 *non-random* and $S_0^0 = 1$. By $\bar{S}_t$ we denote the vector (S_t^0, S_t). Similar notation is used for the portfolio and we have $\bar{\xi}_t = (\xi_t^0, \xi_t)$ with the obvious interpretation. The value of a portfolio at time t will be denoted by W_t, so we have $W_t = \bar{\xi}_t \cdot \bar{S}_t$.

The reader is supposed to be familiar with the notions of *filtration*, *adapted* and *predictable* processes, *martingales* and other concepts that are standard within this context. See Section B.4 for a brief overview.

8.1 Self-financing trading strategies

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which all random variables below are defined. We assume that we are given a filtration $\mathbb{F} = \{\mathcal{F}_0, \ldots, \mathcal{F}_T\}$, where $\mathcal{F}_0$ is trivial, $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Since we fix the time horizon to be T, we assume that $\mathcal{F}_T = \mathcal{F}$. The price process $S = (S_t)_{t=0}^T$ is assumed to be adapted to the filtration $\mathbb{F}$.

Definition 8.1 A *trading strategy* $\bar{\xi} = \{\bar{\xi}_1, \ldots, \bar{\xi}_T\}$ is a $d+1$-dimensional predictable process, i.e. for every $t > 0$, the random vector ξ_t is $\mathcal{F}_{t-1}$-measurable.

The interpretation of a trading strategy is that at time $t-1$ ($t \geq 1$) an investor composes a portfolio $\bar{\xi}_t$, for which (s)he then has to pay $\bar{\xi}_t \cdot \bar{S}_{t-1}$, where $t \geq 1$. This portfolio is held until time t, when the value of the portfolio changes into $W_t = \bar{\xi}_t \cdot \bar{S}_t$. At that time (s)he can re-balance the portfolio to ξ_{t+1}, for which (s)he has to pay $\bar{\xi}_{t+1} \cdot \bar{S}_t$. This re-balancing may happen without infusion or withdrawing of money and will then only be financed by the current value. The requirement of a trading strategy to be predictable is of course reasonable, an investor is not supposed to know future price movements of the stocks (s)he invests in. By definition, a predictable process is formally only defined for $t \geq 1$, but for notational convenience, we will also use $\bar{\xi}_0 := \bar{\xi}_1$.

Definition 8.2 A trading strategy is called *self-financing*, if one has

$$\bar{\xi}_t \cdot \bar{S}_t = \bar{\xi}_{t+1} \cdot \bar{S}_t,$$

for all $t \in \{0, \ldots, T-1\}$.

For any stochastic process X we denote by ΔX the process with $\Delta X_t = X_t - X_{t-1}$, for $t \geq 1$ and $\Delta X_0 = X_0$.

Proposition 8.3 *A trading strategy is self-financing iff for all $t \in \{1, \dots, T\}$ one has*

$$\Delta W_t = \bar{\xi}_t \cdot \Delta \bar{S}_t.$$

Proof Exercise 8.1. □

We will take the process S^0 as a *numéraire*. For this we need and *assume* that S^0 is strictly positive (occasionally strictly positive a.s.). The *discounted* processes X^i $(i = 0, \dots, d)$ are defined by

$$X_t^i = \frac{S_t^i}{S_t^0}.$$

Of course $X_t^0 = 1$ for all t. Write $X_t = (X_t^1, \dots, X_t^d)$ and $\bar{X}_t = (X_t^0, X_t)$. The *(discounted) value process* V is defined by

$$V_t = \frac{W_t}{S_t^0}, \quad t = 1, \dots, T$$

or, equivalently,

$$V_t = \bar{\xi}_t \cdot \bar{X}_t.$$

Note that $V_0 = W_0$. We also need the *(discounted) gains process* G, defined by

$$G_t = \sum_{k=1}^{t} \xi_k \cdot \Delta X_k, \, t \in \{0, \dots, T\},$$

where $G_0 = 0$ by the convention that an empty sum equals zero. Note that ΔX_1 coincides with the vector of discounted net gains Y of Section 1.

We now characterize a self-financing strategy in terms of the discounted gains process.

Proposition 8.4 *Let $\bar{\xi}$ be a trading strategy. The following are equivalent.*

(i) $\bar{\xi}$ *is self-financing.*

(ii) $V_t = V_0 + G_t$, $t = 0, \dots, T$.

Proof By Definition 8.2, the strategy $\bar{\xi}$ is self-financing iff $\bar{\xi}_t \cdot \bar{X}_t = \bar{\xi}_{t+1} \cdot \bar{X}_t$ for $t = 0, \dots, T-1$, which is in turn equivalent to $\Delta V_t = \Delta G_t$, $t = 1, \dots, T$. □

We see that a strategy is self-financing iff changes in the discounted net gains process are completely due to changes in the (discounted) value process, $\Delta V_t = \xi_t \Delta X_t$.

Remark 8.5 As before, we will concentrate on the risky part ξ of the strategy $\bar{\xi}$, if $\bar{\xi}$ is self-financing. This, together with the initial investment V_0 completely determines the trading strategy. Indeed, for a self-financing strategy one has

$$\xi^0_{t+1} = \xi^0_t - (\xi_{t+1} - \xi_t) \cdot X_t \text{ for } t = 0, \ldots, T-1,$$

and also (for $t = 0$)

$$\xi^0_1 = V_0 - \xi_1 \cdot X_0.$$

Conversely, with the risky part of a strategy and V_0 known, the above two equations yield a self-financing strategy $\bar{\xi} = (\xi^0, \xi)$.

8.2 Arbitrage

As before, the intuitive meaning of arbitrage is that it is possible to make a (positive) profit, whereas losses are impossible, also called absence of *downside risk*. The formal definition is as follows and given in terms of the discounted value process V, an equivalent definition in terms of the non-discounted process W is obvious.

Definition 8.6 A self-financing trading strategy is called an *arbitrage opportunity* if its discounted value process V satisfies $V_0 \leq 0$, $\mathbb{P}(V_T \geq 0) = 1$ and $\mathbb{P}(V_T > 0) > 0$. A market is called arbitrage free, if no arbitrage opportunities exist.

As in Section 1, absence of arbitrage in the market is necessary to obtain a fair and sensible pricing system. We first give a characterization of existence of arbitrage. Later on we alternatively characterize absence of arbitrage.

Proposition 8.7 *An arbitrage opportunity exists iff there is a* $t \in \{1, \ldots, T\}$ *and a* $\mathcal{F}_{t-1}$*-measurable random vector* η_t *such that* $\mathbb{P}(\eta_t \cdot \Delta X_t \geq 0) = 1$ *and* $\mathbb{P}(\eta_t \cdot \Delta X_t > 0) > 0$. *As a consequence, in an arbitrage free market, for all* $t \in \{1, \ldots, T\}$ *one has* $\mathbb{P}(\eta_t \cdot \Delta X_t = 0) = 1$ *as soon as* $\mathbb{P}(\eta_t \cdot \Delta X_t \geq 0) = 1$ *for an* $\mathcal{F}_{t-1}$*-measurable random vector* η_t.

Proof Let $\bar{\xi}$ be an arbitrage opportunity and V the corresponding discounted value process. Put

$$t = \min\{k \geq 1 : \mathbb{P}(V_k \geq 0) = 1 \text{ and } \mathbb{P}(V_k > 0) > 0\}.$$

Then $1 \leq t \leq T$ and $\mathbb{P}(V_{t-1} < 0) > 0$ or $\mathbb{P}(V_{t-1} > 0) = 0$. In the first case, let $\eta_t = \xi_t \mathbf{1}_{\{V_{t-1} < 0\}}$. Then η_t is $\mathcal{F}_{t-1}$-measurable and

$$\eta_t \cdot \Delta X_t = \Delta V_t \mathbf{1}_{\{V_{t-1} < 0\}} = (V_t - V_{t-1}) \mathbf{1}_{\{V_{t-1} < 0\}} \geq -V_{t-1} \mathbf{1}_{\{V_{t-1} < 0\}},$$

and the requirements are met. In the other case, we take $\eta_t = \xi_t$ and then $\xi_t \cdot \Delta X_t = \Delta V_t \geq V_t$ a.s. and again the requirements are met, by definition of t.

Conversely, assume that η_t with the stipulated properties exists. Define the trading strategy ξ by $\xi_s = \eta_t \mathbf{1}_{\{t\}}(s)$ and complete it by choosing $V_0 = 0$ and ξ^0 as in Remark 8.5 such that $\bar{\xi}$ is self-financing. Then $V_T = \eta_t \cdot \Delta X_t$ and we have an arbitrage property. □

We have seen in Section 1 that absence of arbitrage was equivalent with the existence of a risk-neutral measure $\mathbb{P}^*$, that by definition had the property, using the current notation, that $\mathbb{E}^*X_1 = X_0$, which is in fact the martingale property of the pair (X_0, X_1), since $\mathcal{F}_0$ is trivial. This makes the next definition understandable.

Definition 8.8 A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ is called a *martingale measure* or a *risk-neutral measure* if the process X is a martingale under $\mathbb{Q}$. If a martingale measure $\mathbb{P}^*$ is equivalent to $\mathbb{P}$ on $\mathcal{F}_T$, then it is called an *equivalent martingale measure*. The set of all equivalent martingale measures is denoted by $\mathcal{P}$.

There are various ways to characterize martingale measures. We use the following

Theorem 8.9 *For a self-financing strategy $\bar{\xi}$ the discounted value process is denoted V. Let $\mathbb{Q}$ be a probability measure on $(\Omega, \mathcal{F}_T)$. Equivalent are*

(i) $\mathbb{Q}$ is a martingale measure.

(ii) If $\bar{\xi}$ is self-financing, bounded, then V is a $\mathbb{Q}$-martingale.

(iii) If $\bar{\xi}$ is self-financing and $\mathbb{E}_{\mathbb{Q}}V_T^- < \infty$, then V is a $\mathbb{Q}$-martingale.

(iv) If $\bar{\xi}$ is self-financing and $\mathbb{Q}(V_T \geq 0) = 1$, then $\mathbb{E}_{\mathbb{Q}}V_T = V_0$.

Proof (i) $\Rightarrow$ (ii): It follows that V_t is $\mathbb{Q}$-integrable for each t, since ξ is bounded. From Proposition 8.4 and ξ being predictable, we have for $t \geq 1$

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[\Delta V_t|\mathcal{F}_{t-1}] &= \xi_t \cdot \mathbb{E}_{\mathbb{Q}}[\Delta X_t|\mathcal{F}_{t-1}] \\ &= 0,\end{aligned}$$

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

(ii) $\Rightarrow$ (iii): As a first step in the proof, we show, for $t \in \{1, \dots, T\}$,

$$\mathbb{E}_{\mathbb{Q}}V_t^- < \infty \Rightarrow \mathbb{E}_{\mathbb{Q}}[V_t|\mathcal{F}_{t-1}] = V_{t-1}\ \mathbb{Q}\text{-a.s.}. \tag{8.1}$$

Since $\mathbb{E}_{\mathbb{Q}}V_t^- < \infty$, the (generalized) conditional expectation $\mathbb{E}_{\mathbb{Q}}[V_t|\mathcal{F}_{t-1}]$ is well defined. Fix $a > 0$ and put $\xi_t^a = \xi_t \mathbf{1}_{\{|\xi_t| \leq a\}}$. Then $\xi_t^a \cdot \Delta X_t$ is the increment of a martingale, since ξ_t^a is bounded, so $\mathbb{E}_{\mathbb{Q}}[\xi_t^a \cdot \Delta X_t|\mathcal{F}_{t-1}] = 0$. Hence

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[V_t|\mathcal{F}_{t-1}]\mathbf{1}_{\{|\xi_t| \leq a\}} &= \mathbb{E}_{\mathbb{Q}}[V_t\mathbf{1}_{\{|\xi_t| \leq a\}}|\mathcal{F}_{t-1}] \\ &= \mathbb{E}_{\mathbb{Q}}[V_t\mathbf{1}_{\{|\xi_t| \leq a\}}|\mathcal{F}_{t-1}] - \mathbb{E}_{\mathbb{Q}}[\xi_t^a \cdot \Delta X_t|\mathcal{F}_{t-1}] \\ &= \mathbb{E}_{\mathbb{Q}}[(V_{t-1} + \xi_t \cdot \Delta X_t)\mathbf{1}_{\{|\xi_t| \leq a\}} - \xi_t^a \cdot \Delta X_t|\mathcal{F}_{t-1}] \\ &= V_{t-1}\mathbf{1}_{\{|\xi_t| \leq a\}}.\end{aligned}$$

By letting $a \to \infty$, one obtains (8.1). Use this equation for $t = T$ to get

$$\mathbb{E}_{\mathbb{Q}}V_{T-1}^- = \mathbb{E}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[V_T|\mathcal{F}_{T-1}])^- \leq \mathbb{E}_{\mathbb{Q}}(\mathbb{E}_{\mathbb{Q}}[V_T^-|\mathcal{F}_{T-1}]) = \mathbb{E}_{\mathbb{Q}}V_T^-,$$

by Jensen's inequality for conditional expectations applied to the convex function $x \mapsto x^-$. From the assumption, we get that $\mathbb{E}_\mathbb{Q} V_{T-1}^- < \infty$. Iterating this procedure, one obtains $\mathbb{E}_\mathbb{Q} V_t^- < \infty$ for all t and by (8.1) also $\mathbb{E}_\mathbb{Q} V_t = \mathbb{E}_\mathbb{Q} V_0 = V_0$, which is a finite quantity. It follows that all V_t are integrable and combined with (8.1) this makes that V is a martingale.

(iii) ⇒ (iv): We clearly have $\mathbb{E}_\mathbb{Q} V_T^- = 0$ and by the fact that V is then a martingale, $\mathbb{E}_\mathbb{Q} V_T = \mathbb{E}_\mathbb{Q} V_0 = V_0$.

(iv) ⇒ (i): We have to show that X is a $\mathbb{Q}$-martingale, for which we shall select convenient trading strategies. First we show that every element X_t^i of X_t is $\mathbb{Q}$-integrable. Let $\xi_s^i = \mathbf{1}_{\{s \leq t\}}$ and $\xi^j = 0$ if $1 \leq j \neq i$. Let $V_0 = X_0^i$ and choose ξ^0 such that $\bar{\xi}$ is self-financing, see Remark 8.5. It follows that now $V_T = X_t^i \geq 0$. Using the assumption, we get that X_t^i has finite expectation, in fact

$$\mathbb{E}_\mathbb{Q} X_t^i = X_0^i. \tag{8.2}$$

Next we show that $\mathbb{E}_\mathbb{Q}[\Delta X_t^i | \mathcal{F}_{t-1}] = 0$, $\mathbb{Q}$-a.s., equivalently, $\mathbb{E}_\mathbb{Q}[\mathbf{1}_A \Delta X_t^i] = 0$ for every $A \in \mathcal{F}_{t-1}$, by selecting another appropriate trading strategy. Given such A, we define $\xi_s^i = \mathbf{1}_{s \leq t} - \mathbf{1}_A \mathbf{1}_{\{s=t\}}$ and $\xi_s^j = 0$ if $1 \leq j \neq i$. Let $V_0 = X_0^i \geq 0$ and complement ξ by ξ^0 to obtain a self-financing strategy (note that it is indeed predictable). A simple computation gives

$$V_T = X_t^i - \mathbf{1}_A \Delta X_t^i = \mathbf{1}_{A^c} X_t^i + \mathbf{1}_A X_{t-1}^i \geq 0.$$

The assumption $\mathbb{E}_\mathbb{Q} V_T = V_0$ now reads $\mathbb{E}_\mathbb{Q}(X_t^i - \mathbf{1}_A \Delta X_t^i) = \mathbb{E}_\mathbb{Q} X_t^i$, in view of (8.2). It follows that $\mathbb{E}_\mathbb{Q}[\mathbf{1}_A \Delta X_t^i] = 0$. □

Remark 8.10 Suppose that $\mathbb{P}$ itself is a martingale measure and that a risk averse investor uses the same probability measure to decide whether or not to invest in products with a certain expected pay-off. According to Example 4.8, he will invest all his capital in a riskless product. Moreover, the market is arbitrage free. Indeed, if $\bar{\xi}$ is a self-financing strategy and $V_0 \leq 0$, then we obtain that $V_T \geq 0$ $\mathbb{P}$-a.s. implies that $\mathbb{E} V_T = 0$ and hence $V_T = 0$ $\mathbb{P}$-a.s.

The main theorem of this section is Theorem 8.12 below, the dynamic version of Theorem 1.6. To prove it, we need a lemma that concerns a static one period model as in Section 1, but now with *random* initial prices. We single out one time step of the dynamic model, the one from $t-1$ to t. Below we write $L^0(\Omega, \mathcal{G}, \mathbb{P})$ for the set of $\mathcal{G}$-measurable random variables, with the identification of $\mathbb{P}$-a.s. equal random variables to be the same.

The *space* generated by the discounted net gains from $t-1$ to t with $t \in \{1, \ldots, T\}$ is

$$\mathcal{K}_t(\mathbb{P}) = \{\xi \cdot \Delta X_t : \xi^i \in L^0(\Omega, \mathcal{F}_{t-1}, \mathbb{P}), i = 1, \ldots, d\}. \tag{8.3}$$

Hence, by Proposition 8.7 an arbitrage-free market can be characterized by the relation

$$\mathcal{K}_t(\mathbb{P}) \cap L_+^0(\Omega, \mathcal{F}_t, \mathbb{P}) = \{0\}, \forall t \in \{1, \ldots, T\}. \tag{8.4}$$

If $\mathbb{P}'$ is a probability measure on $\mathcal{F}_T$ that is equivalent to $\mathbb{P}$, then the no arbitrage condition (8.4) can be replaced with the equivalent condition

$$\mathcal{K}_t(\mathbb{P}') \cap L^0_+(\Omega, \mathcal{F}_t, \mathbb{P}') = \{0\}, \, \forall t \in \{1, \ldots, T\}.$$

Below we sometimes need this equivalent formulation. We will mostly write $\mathcal{K}_t$ for $\mathcal{K}_t(\mathbb{P})$ and keep in mind that $\mathcal{K}_t = \mathcal{K}_t(\mathbb{P}')$ for $\mathbb{P}' \sim \mathbb{P}$. The following is of *fundamental* importance, a cornerstone in the proof of Theorem 8.12.

Lemma 8.11 *Let $t \in \{1, \ldots, T\}$. The following statements are equivalent.*

(i) The intersection $\mathcal{K}_t \cap L^0_+(\Omega, \mathcal{F}_t, \mathbb{P}) = \{0\}$.

(ii) There exists a probability measure $\mathbb{P}^*_t$ on $\mathcal{F}_t$, equivalent to $\mathbb{P}$, with a bounded $\mathcal{F}_t$-measurable density $Z_t = \frac{\mathrm{d}\mathbb{P}^*_t}{\mathrm{d}\mathbb{P}}$, such that $\mathbb{E}_{\mathbb{P}^*_t}[\Delta X_t | \mathcal{F}_{t-1}] = 0$. The measure $\mathbb{P}^*_t$ can be trivially extended to a probability measure on $\mathcal{F}$ by putting $\mathbb{P}^*_t(F) = \mathbb{E}_{\mathbb{P}} \mathbf{1}_F Z_t$ for all $F \in \mathcal{F}$.

Proof See Section 8.3, where this lemma is alternatively formulated as Corollary 8.24. □

We return to the dynamic setting. The next theorem is the first *Fundamental Theorem of Asset Pricing* for a dynamic market in discrete time.

Theorem 8.12 *The market is free of arbitrage iff there exists an equivalent martingale measure $\mathbb{P}^*$ on $\mathcal{F}_T$ with bounded Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}$.*

Proof Assume that a risk-neutral measure $\mathbb{P}^*$ exists. Let $\bar{\xi}$ be any self-financing trading strategy with $V_0 \leq 0$ and $\mathbb{P}^*(V_T \geq 0) = 1$. Theorem 8.9 yields $0 \leq \mathbb{E}^* V_T = V_0 \leq 0$, hence $\mathbb{P}^*(V_T = 0) = 1$ and an arbitrage opportunity doesn't exist under $\mathbb{P}^*$, and then also not under $\mathbb{P}$ by equivalence of the two measures.

Conversely, assume that the market is free of arbitrage. Let $t \in \{1, \ldots, T\}$ and let $\mathcal{K}_t$ be as in (8.3). Recall that by Proposition 8.7 it holds that $\mathcal{K}_t \cap L^0_+(\Omega, \mathcal{F}_t, \mathbb{P}) = \{0\}$ for all t. Consider $t = T$, then Lemma 8.11 applies with $t = T$ and we conclude to the existence of a probability measure $\mathbb{P}^*_T$ on $\mathcal{F}_T = \mathcal{F}$, with $\mathbb{P}^*_T \sim \mathbb{P}$ and $\mathbb{E}_{\mathbb{P}^*_T}[\Delta X_T | \mathcal{F}_{T-1}] = 0$. Moreover $Z_T = \frac{\mathrm{d}\mathbb{P}^*_T}{\mathrm{d}\mathbb{P}}$ is bounded.

We proceed by *backward* induction. Suppose that for $t < T$ a probability measure $\mathbb{P}^*_{t+1}$ on $\mathcal{F}$ is found such that $\mathbb{P}^*_{t+1} \sim \mathbb{P}$, with bounded $\mathcal{F}_t$-measurable $\frac{\mathrm{d}\mathbb{P}^*_{t+1}}{\mathrm{d}\mathbb{P}}$, and

$$\mathbb{E}_{\mathbb{P}^*_{t+1}}[\Delta X_k | \mathcal{F}_{k-1}] = 0, \text{ for } t+1 \leq k \leq T, \tag{8.5}$$

in other words, the process $(X_k)_{k \in \{t, \ldots, T\}}$ is a martingale under $\mathbb{P}^*_{t+1}$. By equivalence we also have $\mathcal{K}_t \cap L^0_+(\Omega, \mathcal{F}_t, \mathbb{P}^*_{t+1}) = \{0\}$. Then, we apply Lemma 8.11 again to obtain existence of a probability measure $\mathbb{P}^*_t$ on $\mathcal{F}$, equivalent to $\mathbb{P}^*_{t+1}$, with bounded density $Z_t = \frac{\mathrm{d}\mathbb{P}^*_t}{\mathrm{d}\mathbb{P}^*_{t+1}}$ which is $\mathcal{F}_t$-measurable, and such that

$$\mathbb{E}_{\mathbb{P}^*_t}[\Delta X_t | \mathcal{F}_{t-1}] = 0.$$

Our next aim is to show for $t+1 \leq k \leq T$ the equality $\mathbb{E}_{\mathbb{P}^*_t}[\Delta X_k|\mathcal{F}_{k-1}] = 0$, equivalently $\mathbb{E}_{\mathbb{P}^*_t}[\mathbf{1}_A \Delta X_k] = 0$, for $A \in \mathcal{F}_{k-1}$. Take such an A and compute, using $\mathcal{F}_t$-measurability of Z_t and (8.5),

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}^*_t}[\mathbf{1}_A \Delta X_k] &= \mathbb{E}_{\mathbb{P}^*_{t+1}}[\mathbf{1}_A \Delta X_k Z_t] \\
&= \mathbb{E}_{\mathbb{P}^*_{t+1}} \mathbb{E}_{\mathbb{P}^*_{t+1}}[\mathbf{1}_A \Delta X_k Z_t|\mathcal{F}_{k-1}] \\
&= \mathbb{E}_{\mathbb{P}^*_{t+1}}(Z_t \mathbf{1}_A \mathbb{E}_{\mathbb{P}^*_{t+1}}[\Delta X_k|\mathcal{F}_{k-1}]) \\
&= 0.
\end{aligned}$$

Hence Equation (8.5) remains true with the substitution $t+1 \to t$. Moreover,

$$\frac{\mathrm{d}\mathbb{P}^*_t}{\mathrm{d}\mathbb{P}} = Z_t \frac{\mathrm{d}\mathbb{P}^*_{t+1}}{\mathrm{d}\mathbb{P}}$$

is bounded as well. By iteration, we conclude that the procedure yields a probability measure $\mathbb{P}^* = \mathbb{P}^*_1$ with the desired properties. □

We close this section by studying what happens under a *change of numéraire*. Absence of arbitrage is defined as the impossibility to have a $\mathbb{P}$-almost sure profit. Clearly, we can replace in this statement $\mathbb{P}$ with any risk-neutral measure $\mathbb{P}^*$, since these measures define the same null sets and the role of the process S^0 is not relevant to describe arbitrage. But any particular $\mathbb{P}^*$ is such that the price processes discounted by the numéraire process S^0 are, by definition, $\mathbb{P}^*$-martingales. Hence, if one prefers to take the price process of another asset as a discount factor, there will be another risk-neutral measure. So, the set of risk-neutral measures depends on the choice of numéraire and it is interesting to investigate how different risk-neutral measures resulting from different numéraires are related.

Suppose one takes the process S^1 as a numéraire. It is assumed that S^1 is $\mathbb{P}$-a.s. strictly positive. Put

$$\bar{Y}_t = \frac{\bar{S}_t}{S^1_t},$$

then $\bar{Y}_t X^1_t = \bar{X}_t$, $t \in \{0, \ldots, T\}$. Let $\tilde{\mathcal{P}}$ denote the set of all probability measures $\tilde{\mathbb{P}}$ equivalent to $\mathbb{P}$ that are such that $\bar{Y}$ is a $\tilde{\mathbb{P}}$-martingale. Absence of arbitrage is then equivalent to $\tilde{\mathcal{P}} \neq \emptyset$, by virtue of Theorem 8.12.

Proposition 8.13 *A probability measure $\tilde{\mathbb{P}}$ belongs to $\tilde{\mathcal{P}}$ iff there exists a probability measure $\mathbb{P}^* \in \mathcal{P}$ such that $\tilde{\mathbb{P}} \sim \mathbb{P}^*$ and*

$$\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}^*} = \frac{X^1_T}{X^1_0}. \tag{8.6}$$

In this case one also has

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\tilde{\mathbb{P}}} = \frac{Y^0_T}{Y^0_0}.$$

Proof Let $\mathbb{P}^*$ be given. The random variables $\frac{X^1_t}{X^1_0}$ form a martingale under $\mathbb{P}^*$, with $\mathbb{E}^* \frac{X^1_T}{X^1_0} = 1$. Hence, if we define $\tilde{\mathbb{P}}$ by (8.6), then it is a probability measure, equivalent to $\mathbb{P}^*$ and for $t > s$ one has

$$
\begin{aligned}
\mathbb{E}_{\tilde{P}}[\bar{Y}_t|\mathcal{F}_s] &= \frac{\mathbb{E}^*[\bar{Y}_t \frac{X^1_t}{X^1_0}|\mathcal{F}_s]}{\mathbb{E}^*[\frac{X^1_t}{X^1_0}|\mathcal{F}_s]} \\
&= \frac{\mathbb{E}^*[\bar{Y}_t X^1_t|\mathcal{F}_s]}{X^1_s} \\
&= \frac{\mathbb{E}^*[\bar{X}_t|\mathcal{F}_s]}{X^1_s} \\
&= \frac{\bar{X}_s}{X^1_s} \\
&= \bar{Y}_s.
\end{aligned}
$$

Hence $\tilde{P}$ is a martingale measure for $\bar{Y}$, or $\tilde{\mathbb{P}} \in \tilde{\mathcal{P}}$. To prove the other implication, one just swaps the roles of X and Y in the previous part. □

Proposition 8.14 *Suppose that X^1_T is not degenerate under $\mathbb{P}$. Then the sets $\mathcal{P}$ and $\tilde{\mathcal{P}}$ have empty intersection.*

Proof Exercise 8.2. □

8.3 Proof of Lemma 8.11

This section extends the proof of existence of an equivalent martingale measure (FTAP, Theorem 1.6) to the situation of a non-trivial initial history, $\mathcal{F}_0$ is *not necessarily the trivial σ-algebra*, with the aim to ultimately prove Theorem 8.12 for which Lemma 8.11 is a key result.

The background is a multi-period model as in Section 8.2 in which *we single out one arbitrary time step*, from $t-1$ to t, for some $t \in \{1, \dots, T\}$. The prices S^i_t are $\mathcal{F}_t$-measurable nonnegative random variables and the portfolio choices ξ^i_t are $\mathcal{F}_{t-1}$-measurable. Note that $\mathcal{F}_{t-1}$ is usually not the trivial σ-algebra for $t > 1$. By a *time shift*, we may as well consider a one period model with $t = 0, 1$ as in Section 1, but with the generalization that $\mathcal{F}_0$ *is no longer assumed to be trivial*. Having done so we can extend the results below to an arbitrary step in a multi-period setting, which eventually leads to Theorem 8.12.

Here is some notation for this section, in agreement with what has been previously introduced. We use L^p as an abbreviation of $L^p(\Omega, \mathcal{F}_1, \mathbb{P})$, for $p \geq 0$. For $p = 0$, we make L^0 a metric space by using the metric d defined by $d(X, Y) = \mathbb{E}[|X - Y| \wedge 1]$. It then holds that $d(X_n, X) \to 0$ iff $X_n \xrightarrow{\mathbb{P}} X$. By L^p_+ we denote the nonnegative elements of L^p.

We adopt the following standing assumption throughout this section. *The $(d+1)$-dimensional price process S is assumed to be adapted and such that* *the discounted prices $X_t^i = S_t^i/S_t^0$ have finite expectation for all $1 \leq i \leq d$ and* $t = 0, 1$. The integrability assumption can be circumvented more or less as in Exercise 1.3. Furthermore, we require (non-random) $S_0^0 > 0$ and $S_1^0 > 0$ to have the X_t^i well defined.

A portfolio is a $(d+1)$-dimensional random vector that is $\mathcal{F}_0$-measurable and thus not necessarily constant. As usual we denote by ξ the investments in the risky assets, now a d-dimensional $\mathcal{F}_0$-measurable random vector. The vector of net gains Y is also defined as usual, but adapted to the current situation we have

$$Y = X_1 - X_0.$$

The random vector Y is $\mathcal{F}_1$-measurable and (component wise) integrable under the standing assumption. Recall from Corollary 1.4 that a market is arbitrage free if for any $\xi \in \mathcal{F}_0$ one has that the discounted portfolio gain $\xi \cdot Y \geq 0$ a.s. implies $\xi \cdot Y = 0$ a.s. The characterization of an arbitrage free market now becomes $\mathcal{K} \cap L^0_+ = \{0\}$, where $\mathcal{K} = \{\xi \cdot Y : \xi^i \in \mathcal{F}_0, i = 1, \ldots, d\}$. In Lemma 8.15 we use the notation $A - B$ for two subsets A and B of some vector space to denote the set $\{a - b : a \in A, b \in B\}$.

Lemma 8.15 *There is equivalence between $\mathcal{K} \cap L^0_+ = \{0\}$ and $(\mathcal{K} - L^0_+) \cap L^0_+ = \{0\}$.*

Proof Assume $\mathcal{K} \cap L^0_+ = \{0\}$ and let $Z \in \mathcal{K} - L^0_+$, $Z = \xi \cdot Y - U$ say. If $Z \in L^0_+$ too, i.e. $Z \geq 0$, then also $\xi \cdot Y \geq 0$ and by the hypothesis $\xi \cdot Y = 0$, which yields $Z = -U \leq 0$. So $Z = 0$. The converse implication follows from $\mathcal{K} \subset \mathcal{K} - L^0_+$. $\square$

In all what follows we let

$$\mathcal{C} = (\mathcal{K} - L^0_+) \cap L^1.$$

Note that $\mathcal{C}$ is a cone, i.e. $W \in \mathcal{C}$ implies $\lambda W \in \mathcal{C}$ for all $\lambda \geq 0$. The concept of martingale measure, adapted to the present situation, is as follows.

Definition 8.16 A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_1)$ is called a martingale measure, or risk-neutral measure, if $\mathbb{E}_{\mathbb{Q}}[Y|\mathcal{F}_0] = 0$ $\mathbb{Q}$-a.s. It is called *equivalent martingale measure*, if moreover $\mathbb{Q} \sim \mathbb{P}$. The set of equivalent martingale measures is denoted $\mathcal{P}$.

In the proof of the next lemma we use the formula for conditional expectations under an absolutely continuous change of measure, see Proposition B.39. For convenience we recall the result here. Let $\mathbb{Q} \ll \mathbb{P}$ with Radon-Nikodym derivative Z. If $\mathbb{E}|XZ| < \infty$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$, then

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}] = \frac{\mathbb{E}[XZ|\mathcal{G}]}{\mathbb{E}[Z|\mathcal{G}]} \ \mathbb{Q}\text{-a.s.} \tag{8.7}$$

Lemma 8.17 *Suppose there is $Z \in L^\infty$ such that $\mathbb{E}(ZW) \leq 0$ for all $W \in \mathcal{C}$. Then $Z \geq 0$ a.s. and if $\mathbb{E}Z = 1$, then $\mathrm{d}\mathbb{Q} = Z\,\mathrm{d}\mathbb{P}$ defines a martingale measure $\mathbb{Q}$.*

Proof Note that $W = -\mathbf{1}_{\{Z<0\}} \in \mathcal{C}$. Hence $\mathbb{E}(-Z\mathbf{1}_{\{Z<0\}}) \leq 0$ and it follows that $Z\mathbf{1}_{\{Z<0\}} = 0$ a.s., hence $Z \geq 0$ a.s. Under the condition $\mathbb{E}Z = 1$, $\mathbb{Q}$ is a probability measure, absolutely continuous w.r.t. $\mathbb{P}$.

Let ξ be bounded, $\mathcal{F}_0$-measurable and $\lambda \in \{-1, 1\}$. Since $\xi \cdot Y \in \mathcal{K}$, also $\lambda\xi \cdot Y \in \mathcal{K} \subset \mathcal{K} - L^0_+$. Because ξ is bounded, we also have $(\lambda\xi) \cdot Y \in L^1$, hence $(\lambda\xi) \cdot Y \in \mathcal{C}$ and therefore $\lambda\,\mathbb{E}(\xi \cdot Y)Z \leq 0$. But since $\lambda \in \{-1, 1\}$ is arbitrary, we must have $\mathbb{E}(\xi \cdot Y)Z = 0$. One then has $0 = \mathbb{E}(\xi \cdot Y)Z = \mathbb{E}[\mathbb{E}[(\xi \cdot Y)Z|\mathcal{F}_0]] = \mathbb{E}[\xi \cdot \mathbb{E}[YZ|\mathcal{F}_0]]$ for all bounded ξ. But then (why?) $\mathbb{E}[YZ|\mathcal{F}_0] = 0$ a.s. Equation (8.7) yields

$$\mathbb{E}_{\mathbb{Q}}[Y|\mathcal{F}_0] = \frac{\mathbb{E}[YZ|\mathcal{F}_0]}{\mathbb{E}[Z|\mathcal{F}_0]} = 0,$$

whence $\mathbb{Q}$ is a martingale measure according to Definition 8.16. □

Remark 8.18 The use of Equation (8.7) above to prove that $\mathbb{Q}$ is a martingale measure can be circumvented by computing for every $F \in \mathcal{F}_0$

$$\mathbb{E}_{\mathbb{Q}}[Y\mathbf{1}_F] = \mathbb{E}[Y\mathbf{1}_F Z] = \mathbb{E}[\mathbf{1}_F \mathbb{E}[YZ|\mathcal{F}_0]] = 0.$$

We proceed with further steps on our way to prove Lemma 8.11. Let

$$\mathcal{Z} = \{Z \in \mathcal{F}_1 : 0 \leq Z \leq 1, \mathbb{E}Z > 0 \text{ and } \mathbb{E}(ZW) \leq 0, \forall W \in \mathcal{C}\}. \tag{8.8}$$

If the set $\mathcal{Z}$ is nonempty, we can choose $Z \in \mathcal{Z}$ and the normalization $\zeta = Z/\mathbb{E}Z$ can then serve as a Radon-Nikodym derivative of a martingale measure w.r.t. $\mathbb{P}$. We shall see that under the additional condition that the market is arbitrage free, the set $\mathcal{Z}$ is indeed non-empty and one can even select a Z^* from it satisfying $\mathbb{P}(Z^* > 0) = 1$, which yields the existence of an *equivalent* martingale measure. The technical property that we need is that the set $\mathcal{C}$ is *closed in* L^1, Proposition 8.21, our next aim. The proof of this requires quite some work.

To accomplish this we need two technical results, a decomposition of L^0 into suitable 'orthogonal' subspaces and a version of the Bolzano-Weierstraß theorem for sequences of random variables, presented next. Note that a random variable X can be viewed as a collection of real numbers $X(\omega)$ for $\omega \in \Omega$ and is thus in general an infinite dimensional object. So a straightforward application of the classical Bolzano-Weierstraß theorem for sequences in a finite-dimensional Euclidean space is not possible. Here we go.

Lemma 8.19 *Let (ξ_n) be a sequence of d-dimensional random vectors defined on some $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\liminf |\xi_n| < \infty$ a.s. Then exists a sequence of strictly increasing random variables σ_m and an a.s. finite random vector ξ such that $\xi_{\sigma_m} \stackrel{\text{a.s.}}{\to} \xi$.*

Proof Let $L = \liminf |\xi_n|$. Then $\mathbb{P}(L = \infty) = 0$ and for definiteness we define $\sigma_m = m$ on $\{L = \infty\}$. From now on we work on the set $F = \{L < \infty\} \in \mathcal{F}$. Put $\sigma^0_1 = 1$ and define recursively for $m \geq 1$

$$\sigma^0_{m+1}(\omega) = \inf A_m(\omega),$$

where

$$A_m(\omega) = \{n > \sigma^0_m(\omega) : ||\xi_n(\omega)| - L(\omega)| < \frac{1}{m}\}.$$

Note that $A_m(\omega)$ contains infinite many elements for every m and $\omega \in F$ by the hypothesis and that all σ^0_m are $\mathcal{F}$-measurable (verify this!). It follows that also the $\xi_{\sigma^0_m}$ are $\mathcal{F}$-measurable. The details are left as Exercise 8.11. Write ξ^1_n for the first component of the vector ξ_n and define $\xi^1 = \liminf \xi^1_{\sigma^0_m}$. Since the $\xi^1_{\sigma^0_m}(\omega)$ converge along a subsequence it makes sense to define $\sigma^1_1 = 1$ and recursively

$$\sigma^1_{m+1}(\omega) = \inf\{\sigma^0_n(\omega) > \sigma^1_m(\omega) : |\xi^1_{\sigma^0_n(\omega)}(\omega) - \xi^1(\omega)| < \frac{1}{m}\}.$$

We conclude that $\xi^1_{\sigma^1_m} \to \xi^1$ on F, which is the desired behavior for the first component of the ξ_n. The further idea is to thin the sequence of σ^1_m in order to obtain a subsequence for which also the second components converge. Thereto one first defines the candidate limit $\xi^2 = \liminf \xi^2_{\sigma^1_m}$ and finds a sequence (σ^2_m) by mimicking the above procedure. Go on like this with subsequent thinning until also the last component converges. □

We proceed with the announced 'orthogonal' decomposition of L^0. Recall the present one-period setting, in particular ξ and η below are always d-dimensional $\mathcal{F}_0$-measurable random vectors.

Lemma 8.20 *Let $N = \{\eta \in L^0(\Omega, \mathcal{F}_0, \mathbb{P})^d : \eta \cdot Y = 0 \text{ a.s.}\}$ and $N^\perp = \{\xi \in L^0(\Omega, \mathcal{F}_0, \mathbb{P})^d : \xi \cdot \eta = 0 \text{ a.s.}, \forall \eta \in N\}$. Then N and $N^\perp$ are closed subsets of $L^0(\Omega, \mathcal{F}_0, \mathbb{P})^d$, $N \cap N^\perp = \{0\}$ and $L^0(\Omega, \mathcal{F}_0, \mathbb{P})^d = N + N^\perp$.*

Proof Let $(\eta_n) \subset N$ such that $\eta_n \xrightarrow{\mathbb{P}} \eta$. Since almost sure convergence holds along a subsequence we must also have $\eta \cdot Y = 0$ a.s. Closedness of $N^\perp$ is proved similarly. If $\eta \in N \cap N^\perp$, then $\eta \cdot \eta = 0$ a.s. and hence $\eta = 0$ a.s.

The final assertion, every vector in $L^0(\Omega, \mathcal{F}_0, \mathbb{P})^d$ can be written as a sum of vectors in N and $N^\perp$, we first prove for the non-random standard basis vectors e_i of $\mathbb{R}^d$ by a projection argument. Note that every e_i belongs to the Hilbert space $H = L^2(\Omega, \mathcal{F}_0, \mathbb{P})^d$. Moreover, $N \cap H$ and $N^\perp \cap H$ are both closed subspaces of H (why?) and have trivial intersection. By using the orthogonal projections on these subspaces we should have $e_i = \eta_i + \xi_i$, with $\eta_i \in N \cap H$ and $\xi \in N^\perp \cap H$. Note that this is not immediately guaranteed, since we don't know yet that $(N^\perp \cap H) + (N^\perp \cap H) = H$. We proceed as follows. Let η_i be the orthogonal projection of e_i onto $N \cap H$ and define $\xi_i = e_i - \eta_i$, the projection error, which is orthogonal to N by construction and has $\mathbb{E}|\xi_i|^2 < \infty$.

Suppose that $\xi_i \notin N^\perp$. Then there must be $\eta \in N$ such that $\xi_i \cdot \eta \neq 0$ a.s., say $\mathbb{P}(\xi_i \cdot \eta > 0) > 0$. The truncated random vector $\tilde{\eta} := \eta \mathbf{1}_{\{\xi_i \cdot \eta > 0, |\eta| \leq c\}}$ also belongs to N, as well as to H for every $c > 0$. Now $\tilde{\eta} \cdot \xi_i = \eta \cdot \xi_i \mathbf{1}_{\{\xi_i \cdot \eta > 0, |\eta| \leq c\}}$ is positive with positive probability for c large enough and it follows that $\mathbb{E}(\tilde{\eta} \cdot \xi_i) > 0$ (the truncation guarantees that this expectation exists) contradicting that ξ_i is orthogonal to N.

Having established the decomposition for the basis vectors e_i, we now turn to the general case. Every $\mathcal{F}_0$-measurable random vector V can be written as $V = \sum_{i=1}^d V_i e_i$, with $\mathcal{F}_0$-measurable random variables V_i. Since $e_i = \xi_i + \eta_i$ with $\eta_i \in N$ and $\xi_i \in N^\perp$, we have $V = \sum_{i=1}^d V_i \xi_i + \sum_{i=1}^d V_i \eta_i$. One verifies that along with the $\eta_i \in N$ also the $V_i \eta_i \in N$, since the V_i are $\mathcal{F}_0$-measurable. Likewise the $V_i \xi_i$ belong to $N^\perp$. Since both spaces N and $N^\perp$ are closed under addition, we have established a decomposition of V. Uniqueness follows from $N \cap N^\perp = \{0\}$. □

Having done all these preparations, we can show the closedness property of $\mathcal{C}$.

Proposition 8.21 *Under the no arbitrage condition $\mathcal{K} \cap L^0_+ = \{0\}$ it holds that $\mathcal{K} - L^0_+$ is closed in L^0 and hence $\mathcal{C}$ is closed in L^1.*

Proof It is sufficient to show the first assertion, the latter being its direct consequence (verify this!). Let (W_n) be a sequence in $\mathcal{K} - L^0_+$ with W as its limit in probability. Along a subsequence, again denoted (W_n), we have a.s. convergence to W. Since $W_n \in \mathcal{K} - L^0_+$, we can write $W_n = \xi_n \cdot Y - U_n$, with $U_n \geq 0$ a.s. Moreover, we may even assume $\xi_n \in N^\perp$. Indeed, by virtue of Lemma 8.20, every $\mathcal{F}_0$-measurable ξ_n can be decomposed as $\xi_n = \xi_n' + \eta_n$ with $\xi_n' \in N^\perp$ and $\eta_n \in N$. But then $\xi_n \cdot Y = \xi_n' \cdot Y$.

In order to apply Lemma 8.19, we first show that $\liminf |\xi_n| < \infty$ a.s. Consider the $\zeta_n := \xi_n / |\xi_n|$ (well defined if $|\xi_n| > 0$, which is w.l.o.g. true on the set I below), these form a bounded sequence. Invoking Lemma 8.19, we can choose an increasing sequence of $\mathcal{F}_0$-measurable random integers τ_n such that $\zeta_{\tau_n} \stackrel{\text{a.s.}}{\to} \zeta$ for some $\mathcal{F}_0$-measurable random vector ζ with norm one. Since the W_n converge a.s. to a finite limit, we have *on the set* $I = \{\liminf |\xi_n| = \infty\}$

$$0 \leq \frac{U_{\tau_n}}{|\xi_{\tau_n}|} = \zeta_{\tau_n} \cdot Y - \frac{W_n}{|\xi_{\tau_n}|} \stackrel{\text{a.s.}}{\to} \zeta \cdot Y.$$

Since $\mathcal{K} \cap L^0_+ = \{0\}$, we conclude that $\zeta \cdot Y = 0$ a.s. on $\{\liminf |\xi_n| = \infty\}$, so $\mathbf{1}_I \zeta \cdot Y = 0$ a.s. Furthermore, since the $\xi_n \in N^\perp$, we have for every $\eta \in N$ that also $\zeta_{\tau_n} \cdot \eta = 0$ a.s. Because $N^\perp$ is closed under a.s. convergence, it follows that $\zeta \in N^\perp$, but then also $\mathbf{1}_I \zeta \in N^\perp$, because $\mathbf{1}_I$ is $\mathcal{F}_0$-measurable. Together with the previously established fact $\mathbf{1}_I \zeta \cdot Y = 0$ (so $\mathbf{1}_I \zeta \in N$), we conclude $\mathbf{1}_I \zeta = 0$ a.s. Since $|\zeta| = 1$ a.s., this can only happen if $\mathbb{P}(I) = 0$.

Having established $\liminf |\xi_n| < \infty$ a.s., we invoke Lemma 8.19 again to obtain an a.s. finite random vector ξ and a sequence of strictly increasing $\mathcal{F}_0$-measurable integer valued random variables σ_n such that $\xi_{\sigma_n} \to \xi$ a.s. Note (verify!) that also $W_{\sigma_n} \to W$ a.s. Hence

$$0 \leq U_{\sigma_n} = \xi_{\sigma_n} \cdot Y - W_{\sigma_n} \to \xi \cdot Y - W =: U \text{ a.s.}$$

Hence we have $W = \xi \cdot Y - U$ with $U \in L^0_+$, i.e. W belongs to $\mathcal{K} - L^0_+$. □

Having proved that $\mathcal{C}$ is closed in L^1, we shall show the existence of a $Z^* \in \mathcal{Z}$ (recall that $\mathcal{Z}$ is defined in (8.8)) that is strictly positive $\mathbb{P}$-a.s, Theorem 8.23 below. We need another auxiliary result.

Lemma 8.22 *Assume that* $(\mathcal{K}-L^0_+)\cap L^1_+ = \{0\}$. *If* U *is a non-negative element of* L^1 *and* $\mathbb{P}(U>0)>0$, *then there exists* $Z\in\mathcal{Z}$ *such that* $\mathbb{E}(UZ)>0$.

Proof In this proof we shall use the Hahn-Banach theorem in the version of Corollary A.9. Suppose that $U\in\mathcal{C}=(\mathcal{K}-L^0_+)\cap L^0_+$. Since it is given that $U\in L^1_+$, we then have $U\in(\mathcal{K}-L^0_+)\cap L^1_+=\{0\}$, which would imply $U=0$ a.s. This contradicts $\mathbb{P}(U>0)>0$, and so we conclude that U cannot be an element of the nonempty *convex* set $\mathcal{C}$, which is closed in L^1 by Proposition 8.21. Hence we apply Corollary A.9 which says that there exists a $Z'\in L^\infty$ with $\sup\{\mathbb{E}(WZ'):W\in\mathcal{C}\}<\mathbb{E}(UZ')<\infty$. Since $0\in\mathcal{C}$, it follows that $\mathbb{E}(UZ')>0$. Moreover we also have $\sup\{\mathbb{E}(WZ'):W\in\mathcal{C}\}\le\mathbb{E}(UZ')<\infty$. For $W\in\mathcal{C}$ we also have $\lambda W\in\mathcal{C}$ for every $\lambda>0$, and thus $\lambda\mathbb{E}(WZ')\le\mathbb{E}(UZ')$, hence $\mathbb{E}(WZ')\le\mathbb{E}(UZ')/\lambda$, for every $\lambda>0$. Hence $\mathbb{E}(WZ')\le 0$ for every $W\in\mathcal{C}$. It now follows from Lemma 8.17 that $Z'\ge 0$ a.s. Moreover, $Z=0$ a.s. would contradict $\mathbb{E}(UZ')>0$ and so $||Z'||_\infty>0$. We conclude that $Z:=Z'/||Z'||_\infty$ belongs to $\mathcal{Z}$ and has the property $\mathbb{E}(UZ)>0$. □

Theorem 8.23 *Assume that* $(\mathcal{K}-L^0_+)\cap L^1_+=\{0\}$. *Then there exists a* $Z^*\in\mathcal{Z}$ *with* $\mathbb{P}(Z^*>0)=1$.

Proof Let $\alpha:=\sup\{\mathbb{P}(Z>0):Z\in\mathcal{Z}\}\le 1$. By definition of α, there exists a sequence $(Z_n)\subset\mathcal{Z}$ such that $\mathbb{P}(Z_n>0)\uparrow\alpha$. Let $Z^*:=\sum_{n\ge1}2^{-n}Z_n$. Check, use the dominated convergence theorem, that this infinite sum belongs to $\mathcal{Z}$ as well. Since for every n it holds that $\mathbb{P}(Z^n>0)\le\mathbb{P}(Z^*>0)$, it follows that $\mathbb{P}(Z^*>0)=\alpha$.

To show that $\alpha=1$, we assume the contrary, $\mathbb{P}(Z^*=0)>0$ and construct a $Z'\in\mathcal{Z}$ with $\mathbb{P}(Z'>0)>\alpha$. So let $\mathbb{P}(Z^*=0)>0$, then $U=\mathbf{1}_{\{Z^*=0\}}$ is nonnegative and $\mathbb{P}(U>0)>0$. Lemma 8.22 yields the existence of $Z\in\mathcal{Z}$ such that $\mathbb{E}(\mathbf{1}_{\{Z^*=0\}}Z)>0$ and we must have $\mathbb{P}(\mathbf{1}_{\{Z^*=0\}}Z>0)>0$, so $\mathbb{P}(Z^*=0,Z>0)>0$. Let now $Z'=\frac{1}{2}(Z+Z^*)$. One verifies that $Z'\in\mathcal{Z}$ and

$$\begin{aligned}\mathbb{P}(Z'>0)&=\mathbb{P}(Z+Z^*>0,Z^*>0)+\mathbb{P}(Z+Z^*>0,Z^*=0)\\&=\mathbb{P}(Z^*>0)+\mathbb{P}(Z>0,Z^*=0)>\alpha,\end{aligned}$$

a contradiction. □

Here is Lemma 8.11, formulated in agreement with the terminology and notation of the present section.

Corollary 8.24 *If the one-period market is arbitrage free,* $\mathcal{K}\cap L^0_+=\{0\}$, *there exists an equivalent martingale measure* $\mathbb{P}^*$ *such that* $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}$ *is bounded. Conversely, if there exists an equivalent martingale measure* $\mathbb{P}^*$, *the market is arbitrage free.*

Proof Assume that the market is arbitrage free. In view of Lemma 8.15, we have $(\mathcal{K}-L^0_+)\cap L^0_+=\{0\}$. But then also $\mathcal{C}\cap L^1_+=(\mathcal{K}-L^0_+)\cap L^1_+=\{0\}$.

Theorem 8.23 yields the existence of $Z^* \in \mathcal{Z}$ such that $\mathbb{P}(Z^* > 0) = 1$. Then $\mathbb{P}^*$ defined by

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = \frac{Z^*}{\mathbb{E}Z^*}$$

is a probability measure equivalent to $\mathbb{P}$ and a martingale measure in view of Lemma 8.17. Since $Z^* \in \mathcal{Z}$, it is bounded.

Conversely, take a risk-neutral measure $\mathbb{P}^*$ and an arbitrary $\mathcal{F}_0$-measurable ξ such that $\xi \cdot Y \in \mathcal{K} \cap L^0_+(\Omega, \mathcal{F}, \mathbb{P})$. Then for all $m > 0$ the random variable $\mathbf{1}_{\{|\xi| \le m\}} \xi \cdot Y$ belongs to $\mathcal{K} \cap L^1_+(\Omega, \mathcal{F}, \mathbb{P}^*)$, because its expectation $\mathbb{E}^* \mathbf{1}_{\{|\xi| \le m\}} \xi \cdot Y$ is well defined. Hence we have we have

$$\mathbb{E}^* \mathbf{1}_{\{|\xi| \le m\}} \xi \cdot Y = \mathbb{E}^* (\mathbf{1}_{\{|\xi| \le m\}} \xi \cdot \mathbb{E}^*[Y | \mathcal{F}_0]) = 0.$$

It follows that $\mathbf{1}_{\{|\xi| \le m\}} \xi \cdot Y = 0$, $\mathbb{P}^*$-a.s. for all m and therefore $\xi \cdot Y = 0$, $\mathbb{P}^*$-a.s. By equivalence, $\xi \cdot Y = 0$, $\mathbb{P}$-a.s. too. Hence ξ does not yield an arbitrage opportunity. □

8.4 European contingent claims

In this section we study the valuation problem for European contingent claims. *The standing assumption is that the market is arbitrage-free.*

Definition 8.25 A *contingent claim* C is a nonnegative $\mathcal{F}_T$-measurable random variable. It is called a *derivative* of the underlying assets, if C is measurable w.r.t. the σ-algebra $\sigma(S_0, \ldots, S_T)$.

If a contingent claim C is a derivative, then there exists a Borel function $f : (\mathbb{R}^{d+1})^{T+1} \to \mathbb{R}$ such that $C = f(S_0, \ldots, S_T)$.

Here are some examples of contingent claims. The first one is $C = (S^i_T - K)^+$, the European call option on S^i with maturity T and strike price K. An Asian option is for instance the claim $C = (\frac{1}{T+1} \sum_{t=0}^T S^i_t - K)^+$. A knock-in option is for instance $C = \mathbf{1}_{\{\max_{0 \le t \le T} S^i_t \ge B\}}$, where B is a (nonnegative) constant.

Definition 8.26 A contingent claim C is called attainable if there exists a self-financing trading strategy $\bar{\xi}$ such that $C = \bar{\xi}_T \cdot \bar{S}_T$. Such a strategy is called a *replicating* or *hedge* strategy for C.

The discounted value of a claim C is given by

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

Let the claim C be attainable with replicating strategy $\bar{\xi}$. The discounted value process of this self-financing strategy is given in terms of the discounted gains process G, see Proposition 8.4, by

$$V_t = V_0 + G_t = V_0 + \sum_{s=1}^t \xi_s \cdot \Delta X_s, \tag{8.9}$$

and since C is attainable, we have for its discounted value H the relation

$$H = \bar{\xi}_T \cdot \bar{X}_T = V_T = V_0 + \sum_{t=1}^{T} \xi_t \cdot \Delta X_t.$$

We will also say that H is attainable. Notice that $H \geq 0$ a.s.

Proposition 8.27 *Let $\mathbb{P}^*$ be any equivalent martingale measure and H an attainable claim. Then $\mathbb{E}^* H < \infty$. If $\bar{\xi}$ is a replicating strategy, then its discounted value process V satisfies*

$$V_t = \mathbb{E}^*[H|\mathcal{F}_t] \text{ a.s.,} \tag{8.10}$$

for all $t = 0, \ldots, T$, hence V is a nonnegative martingale under $\mathbb{P}^*$.

Proof This follows from Theorem 8.9 (iv), since $H = V_T \geq 0$. □

Notice that this proposition concerns the discounted value of the claim. Of course, if S^0_T is deterministic, also C has finite expectation under each equivalent martingale measure. Moreover, it has two important consequences. The first one is that V_t, although it can be viewed as a conditional expectation, is the same for every equivalent martingale measure in view of (8.9). The second one is that every replicating strategy for H has the same value process.

Equation (8.10) can be rewritten as

$$\bar{\xi}_t \cdot \bar{S}_t = S^0_t \, \mathbb{E}^*[\frac{C}{S^0_T}|\mathcal{F}_t],$$

which for $t = 0$ yields the initial investment to purchase the replicating strategy,

$$V_0 = S^0_0 \, \mathbb{E}^* \frac{C}{S^0_T} = \mathbb{E}^* \frac{C}{S^0_T}.$$

Note that V_0 is the same for every equivalent martingale measure. This number can be interpreted as the fair price (at $t = 0$) of the undiscounted claim C. Any other price would result in an arbitrage opportunity, see the arguments for the corresponding statement in Section 1, realizing that $V_T = H$.

For non-attainable claims we have the following formal definition (compare also to Definition 1.11) of a fair price.

Definition 8.28 A nonnegative real number π^H is called an arbitrage-free price (at $t = 0$) of a discounted contingent claim H, if there exists an adapted process X^{d+1} such that a.s.

$$\begin{aligned} X^{d+1}_0 &= \pi^H, \\ X^{d+1}_t &\geq 0, \text{ for } t = 1, \ldots, T-1, \\ X^{d+1}_T &= H, \end{aligned}$$

and if the extended market with (discounted) price process $(X^1, \ldots, X^{d+1})$ is arbitrage-free. The set of all arbitrage-free prices is denoted by $\Pi(H)$.

It is mathematically more convenient to define a price for the discounted claim H. But of course, this is equivalent to a similar definition of an arbitrage price π^C for the undiscounted claim C. One has $\pi^C = S_0^0 \pi^H$. Since one usually takes $S_0^0 = 1$, it follows that $\pi^C = \pi^H$.

Definition 8.28 is the dynamic counterpart of Definition 1.11. Note that for an attainable discounted claim H, one can take $X_t^{d+1} = \mathbb{E}^*[H|\mathcal{F}_t]$, which is equal to the value V_t of a replication strategy, to see that the fair price of H is equal to V_0. This is in agreement with Proposition 8.27 and the discussion after it.

Our first result in the valuation of claims is presented below; it extends Theorem 1.12 for the static situation to the present dynamic setting.

Theorem 8.29 *The set $\Pi(H)$ is non-empty and one has*

$$\Pi(H) = \{\mathbb{E}^* H : \mathbb{P}^* \in \mathcal{P},\, \mathbb{E}^* H < \infty\}. \tag{8.11}$$

Moreover, the upper and lower bounds of $\Pi(H)$ are given by $\sup\{\mathbb{E}^* H : \mathbb{P}^* \in \mathcal{P}\}$ and $\inf\{\mathbb{E}^* H : \mathbb{P}^* \in \mathcal{P}\}$ respectively.

Proof First we show that the set on the right hand side of (8.11) is non-empty. Define a probability measure $\mathbb{P}'$ on $\mathcal{F}_T$ by

$$\frac{\mathrm{d}\mathbb{P}'}{\mathrm{d}\mathbb{P}} = \frac{c}{H+1},$$

where c is the normalization constant. Then $\mathbb{P}' \sim \mathbb{P}$, hence under $\mathbb{P}'$ the market is arbitrage-free, and $\mathbb{E}_{\mathbb{P}'} H < \infty$. According to Theorem 8.12, there exists a risk-neutral measure $\mathbb{P}^*$ such that $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}'}$ is bounded. But then $\mathbb{E}^* H < \infty$, and thus belongs to $\{\mathbb{E}^* H : \mathbb{P}^* \in \mathcal{P},\, \mathbb{E}^* H < \infty\}$.

Next we prove (8.11). Take $\pi^H \in \Pi(H)$, recall Definition 8.28 and apply Theorem 8.12 to the extended market. This yields the existence of a probability measure $\mathbb{P}^*$ on $\mathcal{F}_T$ such that the X^i become martingales for $i = 1, \ldots, d+1$. But this implies that $\mathbb{P}^* \in \mathcal{P}$ and $\pi^H = X_0^{d+1} = \mathbb{E}^* X_T^{d+1} = \mathbb{E}^* H$. So $\pi^H \in \{\mathbb{E}^* H : \mathbb{P}^* \in \mathcal{P}, \mathbb{E}^* H < \infty\}$.

Conversely, take $\mathbb{P}^* \in \mathcal{P}$ such that $\mathbb{E}^* H < \infty$. Define $X_t^{d+1} = \mathbb{E}^*[H|\mathcal{F}_t]$. Then $\mathbb{P}^*$ is an equivalent martingale measure for the extended market, which is thus arbitrage free, and the requirements of Definition 8.28 are met with $\pi^H = \mathbb{E}^* H$. By the first part of the proof we now also know that $\Pi(H) \neq \emptyset$.

That $\inf \Pi(H) = \inf\{\mathbb{E}^* H : \mathbb{P}^* \in \mathcal{P}\}$ is trivial. To show the companion statement with suprema instead of infima, we note that we only have to consider the case in which $\{\mathbb{E}^* H : \mathbb{P}^* \in \mathcal{P}\}$ differs from $\Pi(H)$, which happens if there exists some $\mathbb{P}_\infty \in \mathcal{P}$ such that $\mathbb{E}_{\mathbb{P}_\infty} H = \infty$. The desired equality follows, as soon as we can show that for all $c > 0$, there exists a $\mathbb{P}_c \in \mathcal{P}$ such that $\infty > \mathbb{E}_{\mathbb{P}_c} H > c$. Indeed, in this case we have by the first part of the theorem that $\mathbb{E}_{\mathbb{P}_c} H \in \Pi(H)$ and it follows that $\sup \Pi(H) = \infty$.

First we note that for given $c > 0$, by monotone convergence, there exists n such that $\pi_n := \mathbb{E}_{\mathbb{P}_\infty}(H \wedge n) > c$. Put $X_t^{d+1} = \mathbb{E}_{\mathbb{P}_\infty}[H \wedge n|\mathcal{F}_t]$ and note that

$\pi_n = X_0^{d+1}$. The measure $\mathbb{P}_\infty$ becomes an equivalent martingale measure in the market extended with the additional asset $H \wedge n$. This extended market is free of arbitrage, when the price vector is extended with π_n. Application of the first part of the theorem to the extended market then shows that for any contingent claim in the extended market, in particular for H, there exists a $\mathbb{P}_c$, equivalent to $\mathbb{P}_\infty$, such that $\mathbb{E}_{\mathbb{P}_c} H < \infty$. But then this $\mathbb{P}_c$ is also an equivalent martingale measure for the original market and thus $\mathbb{E}_{\mathbb{P}_c} H \in \Pi(H)$. On the other hand, the price process X^{d+1} is a martingale under $\mathbb{P}_c$ as well, and so $\mathbb{E}_{\mathbb{P}_c} X_T^{d+1} = X_0^{d+1}$. Using this fact, we have

$$\mathbb{E}_{\mathbb{P}_c} H \geq \mathbb{E}_{\mathbb{P}_c}(H \wedge n) = \mathbb{E}_{\mathbb{P}_c} X_T^{d+1} = X_0^{d+1} = \pi_n > c,$$

which finishes the proof for the supremum. □

We extend more results of Section 1 to a dynamic setting. Recall Proposition 1.19; its dynamic version is the next theorem.

Theorem 8.30 *Assume the market to be arbitrage free. Let H be a discounted claim. If H is attainable, $\Pi(H)$ consists of one element, the value at $t = 0$ of any replicating portfolio. If H is not attainable, then $\Pi(H)$ is an open interval.*

Proof If H is attainable, then the assertion follows from Theorem 8.29 combined with the discussion after Proposition 8.27.

The proof of the other case is much more involved. As in the proof of Proposition 1.19 we observe that $\Pi(H)$ is convex and thus an interval. We will show that it is open. To that end, let $\pi \in \Pi(H)$. It is sufficient to show that there are $\pi_0, \pi_1 \in \Pi(H)$ such that $\pi_0 < \pi < \pi_1$. We first construct π_1.

Take $\mathbb{P}^* \in \mathcal{P}$ such that $\mathbb{E}^* H = \pi$ and let $M_t = \mathbb{E}^*[H|\mathcal{F}_t]$. Then

$$H = M_0 + \sum_{t=1}^{T} \Delta M_t.$$

Since H is not attainable, there must be some $t \in \{1, \ldots, T\}$ such that ΔM_t can not be written as $\xi_t \cdot \Delta X_t$, for some $\mathcal{F}_{t-1}$-measurable ξ_t with $\xi_t \cdot \Delta X_t \in L^1(\Omega, \mathcal{F}_t, \mathbb{P}^*)$. It follows from the proof of Proposition 8.21 (take all $U_n = 0$ there) that the collection $\mathcal{C}_t$ of all random variables that are a.s. equal to such a $\xi_t \cdot \Delta X_t$ is a closed linear subspace of $L^1(\Omega, \mathcal{F}_t, \mathbb{P}^*)$, and thus convex as well. We apply the infinite dimensional version of the separating hyperplane theorem of Corollary A.9 to conclude that there exists a $Z \in L^\infty(\Omega, \mathcal{F}_t, \mathbb{P}^*)$ such that

$$\sup_{W \in \mathcal{C}_t} \mathbb{E}^* W Z < \mathbb{E}^* \Delta M_t Z < \infty.$$

If we replace in the above inequality W with αW for arbitrary $\alpha \in \mathbb{R}$, then by linearity the inequality can only be preserved if

$$\mathbb{E}^* W Z = 0, \forall W \in \mathcal{C}_t. \tag{8.12}$$

We conclude that

$$\mathbb{E}^*\Delta M_t Z > 0. \tag{8.13}$$

As Z is bounded $\mathbb{P}^*$-a.s., by multiplying Z by a sufficiently small positive number, if necessary, we may assume that (8.12) and (8.13) are true for a random variable Z with $\mathbb{P}^*(|Z| < \frac{1}{2}) = 1$. Let

$$Z_t = 1 + Z - \mathbb{E}^*[Z|\mathcal{F}_{t-1}].$$

Then $\mathbb{P}^*(0 < Z_t < 2) = 1$, $\mathbb{E}^* Z_t = 1$ and $\frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\mathbb{P}^*} = Z_t$ defines a probability measure $\mathbb{P}_t \sim \mathbb{P}^*$ on $\mathcal{F}_T$ with Radon-Nikodym derivative Z_t that is $\mathcal{F}_t$-measurable. We compute

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_t} H &= \mathbb{E}^* H Z_t \\
&= \mathbb{E}^* H + \mathbb{E}^*(\mathbb{E}^*[H|\mathcal{F}_t]Z) - \mathbb{E}^*(H\mathbb{E}^*[Z|\mathcal{F}_{t-1}]) \\
&= \mathbb{E}^* H + \mathbb{E}^* M_t Z - \mathbb{E}^*(\mathbb{E}^*[H|\mathcal{F}_{t-1}]\mathbb{E}^*[Z|\mathcal{F}_{t-1}]) \\
&= \mathbb{E}^* H + \mathbb{E}^* M_t Z - \mathbb{E}^*(M_{t-1}\mathbb{E}^*[Z|\mathcal{F}_{t-1}]) \\
&= \mathbb{E}^* H + \mathbb{E}^* M_t Z - \mathbb{E}^* M_{t-1} Z \\
&= \mathbb{E}^* H + \mathbb{E}^*(\Delta M_t Z) \\
&> \mathbb{E}^* H,
\end{aligned}$$

where the inequality follows from (8.13). Since $\mathbb{E}_{\mathbb{P}_t} H = \mathbb{E}^* H Z_t \leq 2\mathbb{E}^* H < \infty$, we can take $\pi_1 = \mathbb{E}_{\mathbb{P}_t} H$ and then

$$\pi_1 > \mathbb{E}^* H. \tag{8.14}$$

Hence we have reached our aim, provided that $\mathbb{P}_t$ is a martingale measure (and thus belongs to $\mathcal{P}$), which we are going to prove now. We discern three cases.

Let $k > t$, the first case. Since Z_t is $\mathcal{F}_t$-measurable and hence $\mathcal{F}_{k-1}$-measurable, we have (recall Proposition B.39)

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_t}[\Delta X_k|\mathcal{F}_{k-1}] &= \frac{\mathbb{E}^*[\Delta X_k Z_t|\mathcal{F}_{k-1}]}{\mathbb{E}^*[Z_t|\mathcal{F}_{k-1}]} \\
&= \mathbb{E}^*[\Delta X_k|\mathcal{F}_{k-1}] = 0.
\end{aligned}$$

For $k = t$, the second case, we now show $\mathbb{E}^*[\Delta X_t Z|\mathcal{F}_{t-1}] = 0$. Let $F \in \mathcal{F}_{t-1}$ arbitrary. Because of $\mathbf{1}_F \in \mathcal{F}_{t-1}$ and (8.12), it holds that $\mathbb{E}^*(\mathbf{1}_F e_i \cdot \Delta X_t Z) = 0$ for every basis vector e_i of $\mathbb{R}^d$. Hence the vector $\mathbb{E}^*(\mathbf{1}_F \Delta X_t Z) = 0$ and since $F \in \mathcal{F}_{t-1}$ was arbitrary, this is equivalent to $\mathbb{E}^*[\Delta X_t Z|\mathcal{F}_{t-1}] = 0$. Note also that $\mathbb{E}^*[Z_t|\mathcal{F}_{t-1}] = 1$, straight from the definition of Z_t. But then

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_t}[\Delta X_t|\mathcal{F}_{t-1}] &= \mathbb{E}^*[\Delta X_t Z_t|\mathcal{F}_{t-1}] \\
&= \mathbb{E}^*[\Delta X_t(1 - \mathbb{E}^*[Z|\mathcal{F}_{t-1}])|\mathcal{F}_{t-1}] + \mathbb{E}^*[\Delta X_t Z|\mathcal{F}_{t-1}] \\
&= (1 - \mathbb{E}^*[Z|\mathcal{F}_{t-1}])\mathbb{E}^*[\Delta X_t|\mathcal{F}_{t-1}] + 0 = 0.
\end{aligned}$$

The third case, $k < t$, is easy. Since one now has $\mathbb{E}^*[Z_t|\mathcal{F}_k] = 1$, one obtains $\mathbb{E}^*[\frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\mathbb{P}^*}|\mathcal{F}_k] = 1$, so the measures $\mathbb{P}_t$ and $\mathbb{P}^*$ coincide on $\mathcal{F}_k$: $\mathbb{P}_t(A) = \mathbb{E}^*\mathbb{E}^*[\mathbf{1}_A Z_t|\mathcal{F}_k] = \mathbb{E}^*(\mathbf{1}_A \mathbb{E}^*[Z_t|\mathcal{F}_k]) = \mathbb{P}^*(A)$ for $A \in \mathcal{F}_k$. Hence

$$\mathbb{E}_{\mathbb{P}_t}[\Delta X_k|\mathcal{F}_{k-1}] = \mathbb{E}^*[\Delta X_k|\mathcal{F}_{k-1}] = 0.$$

Combining the three cases, we conclude that $\mathbb{P}_t$ is an equivalent martingale measure and hence $\pi_1 \in \Pi(H)$.

We turn to the construction of π_0. Let

$$\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}^*} = 2 - Z_t.$$

Then $\mathbb{P}^*(0 < \frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}^*} < 2) = 1$ and $\mathbb{E}^* \frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}^*} = 1$. Hence also $\mathbb{P}_0$ is a probability measure, equivalent to $\mathbb{P}^*$, and a martingale measure as well. The latter follows from the just proven fact that $\mathbb{P}_t$ is a martingale measure, Exercise 8.15. Moreover,

$$\mathbb{E}_{\mathbb{P}_0} H = \mathbb{E}^*(2 - Z_t)H = 2\mathbb{E}^*H - \mathbb{E}^*Z_tH = 2\mathbb{E}^*H - \pi_1 < \mathbb{E}^*H = \pi,$$

by (8.14). Taking $\pi_0 = \mathbb{E}_{\mathbb{P}_0} H$ completes the proof. □

8.5 Complete markets

The definition of a complete market looks the same as for the static case, Definition 1.20, but it involves the more subtle notion of attainability in the multi period setting as in Definition 8.26.

Definition 8.31 An arbitrage-free market is *complete*, if every contingent claim is attainable.

A consequence of a market being complete is that every contingent claim has a unique price, in view of Theorem 8.30. We now present what is also known as the *Second Fundamental Theorem of Asset Pricing*, see also Theorem 1.22.

Theorem 8.32 *An arbitrage-free market is complete iff there exists a unique equivalent martingale measure. The number of atoms of* $(\Omega, \mathcal{F}, \mathbb{P})$ *in case of a complete market is at most* $(d+1)^T$*. Moreover,* $\dim L^0(\Omega, \mathcal{F}, \mathbb{P}) \leq (d+1)^T$ *and* Ω *can be decomposed in at most* $(d+1)^T$ *atoms.*

Proof If the market is complete, we argue as in the proof of Theorem 1.22. Every claim $\mathbf{1}_F$, with $F \in \mathcal{F} = \mathcal{F}_T$ has a unique price. Hence there is a unique $\mathbb{P}^*$. Conversely, if there exists only one equivalent martingale measure, the result follows from Theorem 8.30.

We turn to the number of atoms. We have seen the statement to be true for $T = 1$ in Theorem 1.22 and we proceed by induction. Suppose that the assertion is true for a time horizon $T - 1$. By completeness, every claim can be replicated. So if H is a bounded nonnegative discounted claim, there is a replicating strategy $\bar{\xi}$ with value process V such that

$$H = V_{T-1} + \xi_T \cdot \Delta X_T.$$

Since V_{T-1} and ξ_T are $\mathcal{F}_{T-1}$-measurable, they are constant on atoms A that belong to $\mathcal{F}_{T-1}$. Consider for such A the restricted probability space $(A, \mathcal{F}_T^A, \mathbb{P}^A)$, where $\mathcal{F}_T^A = \{F \cap A : F \in \mathcal{F}_T\}$, and $\mathbb{P}^A$ the conditional probability $\mathbb{P}(\cdot|A)$ restricted to $\mathcal{F}_T^A$. As we just said, on this restricted probability space V_{T-1} and ξ_T are constant. Hence Theorem 1.22 applies and so the dimension of $L^\infty(A, \mathcal{F}_T^A, \mathbb{P}^A)$ is at most $d+1$. Then Proposition 1.21 implies that $(A, \mathcal{F}_T^A, \mathbb{P}^A)$ has at most $d+1$ atoms.

Every atom of $(\Omega, \mathcal{F}_T, \mathbb{P})$ is an atom of one and only one $(A, \mathcal{F}_T^A, \mathbb{P}^A)$. Indeed, if B is an atom of $(\Omega, \mathcal{F}_T, \mathbb{P})$ and the different atoms of $\mathcal{F}_{T-1}$ are labelled A^i, then $\mathbb{P}(B) = \sum_i \mathbb{P}(B \cap A_i)$. But $B \cap A_i \subset B$ and an element of $\mathcal{F}_T$. Hence there is only one $A := A_i$ such that $\mathbb{P}(B) = \mathbb{P}(B \cap A)$. Hence we can w.l.o.g. consider B as an atom in $(A, \mathcal{F}_T^A, \mathbb{P}^A)$. Applying the induction hypothesis, we know that there are at most $(d+1)^{T-1}$ of such restricted probability spaces. The conclusion follows by multiplication, and it implies the assertion on the dimension. □

Consider the set $\mathcal{Q}$ of all martingale measures as in Definition 8.8, it is a convex set. Likewise the set of equivalent martingale measures $\mathcal{P}$ is convex. We will see below that complete markets can be characterized by extreme points of those convex sets. Recall that an extreme point of a convex set is such that it doesn't admit a non-trivial convex combination of points in the convex set.

Theorem 8.33 *Let $\mathbb{P}^* \in \mathcal{P}$. The following are equivalent.*

(i) $\mathcal{P} = \{\mathbb{P}^*\}$ *(the market is complete).*

(ii) $\mathbb{P}^*$ *is an extreme point of* $\mathcal{P}$.

(iii) $\mathbb{P}^*$ *is an extreme point of* $\mathcal{Q}$.

(iv) If M is a martingale under $\mathbb{P}^*$, then there exists a d-dimensional predictable process ξ, such that

$$M_t = M_0 + \sum_{k=1}^{t} \xi_k \cdot \Delta X_k,\ t \in \{0, \dots, T\}.$$

Proof (i) $\Rightarrow$ (iii): Write $\mathbb{P}^* = \alpha\mathbb{Q}_1 + (1-\alpha)\mathbb{Q}_2$ for $\alpha \in (0,1)$ and $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{Q}$. Then $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are necessarily absolutely continuous w.r.t. $\mathbb{P}^*$. But also $\mathbb{P}_i = \frac{1}{2}(\mathbb{Q}_i + \mathbb{P}^*)$ $(i = 1,2)$, being convex combinations of martingale measures, are martingales measures too, and equivalent to $\mathbb{P}^*$. From the assumption it follows that $\mathbb{P}_1 = \mathbb{P}_2 = \mathbb{P}^*$ and then also $\mathbb{Q}_1 = \mathbb{Q}_2 = \mathbb{P}^*$.

(iii) $\Rightarrow$ (ii): Trivial, since $\mathcal{P} \subset \mathcal{Q}$.

(ii) $\Rightarrow$ (i): Let $\mathbb{P}_0^* \in \mathcal{P}$, different from $\mathbb{P}^*$. We first show that we can find a $\mathbb{P}_1^* \in \mathcal{P}$ different from $\mathbb{P}^*$ such that $\frac{d\mathbb{P}_1^*}{d\mathbb{P}^*}$ is bounded. Since $\mathbb{P}_0^* \neq \mathbb{P}^*$, there must be a set $E \in \mathcal{F}_T$ such that $\mathbb{P}^*(E) \neq \mathbb{P}_0^*(E)$. Extend the market with the $\mathbb{P}_0^*$-martingale X^{d+1} given by $X_t^{d+1} = \mathbb{E}_{\mathbb{P}_0^*}[\mathbf{1}_E|\mathcal{F}_t]$. Choose $(\Omega, \mathcal{F}_T, \mathbb{P}^*)$ to be the underlying probability space. It follows by construction, that $\mathbb{P}_0^*$ is an equivalent martingale measure for the extended market, and so the extended market is also arbitrage free under $\mathbb{P}^*$. Theorem 8.12 then provides the existence of a probability measure $\mathbb{P}_1^*$ equivalent to $\mathbb{P}^*$ such that $\frac{\mathrm{d}\mathbb{P}_1^*}{\mathrm{d}\mathbb{P}^*}$ is bounded by some $B > 0$, which is a risk-neutral measure for the extended market. Since

$$X_0^{d+1} = \mathbb{P}_0^*(E) \neq \mathbb{P}^*(E) = \mathbb{E}^* X_T^{d+1},$$

$\mathbb{P}^*$ is not a martingale measure for the extended market, and hence $\mathbb{P}_1^*$ must be different from $\mathbb{P}^*$.

Choose $\varepsilon < 1/B$ and put

$$Z = 1 + \varepsilon - \varepsilon \frac{\mathrm{d}\mathbb{P}_1^*}{\mathrm{d}\mathbb{P}^*}.$$

Notice that $\varepsilon \leq Z \leq 1 + \varepsilon$ and $\mathbb{E}^* Z = 1$. Hence

$$\frac{\mathrm{d}\mathbb{P}_2^*}{\mathrm{d}\mathbb{P}^*} = Z$$

defines a probability measure $\mathbb{P}_2^* \in \mathcal{P}$ (that $\mathbb{P}_2^*$ is a martingale measure is Exercise 8.9) with bounded density Z. Moreover, $\mathbb{P}^*$ turns out to be a convex combination,

$$\mathbb{P}^* = \frac{\varepsilon}{1+\varepsilon}\mathbb{P}_1^* + \frac{1}{1+\varepsilon}\mathbb{P}_2^*,$$

which contradicts that $\mathbb{P}^*$ is extreme.

(i) $\Rightarrow$ (iv): Let M be a positive $\mathbb{P}^*$-martingale. Then we can see M_T as a discounted contingent claim, which is attainable by assumption and Theorem 8.32. Let $\bar{\xi}$ be a replicating strategy. Then, $\mathbb{P}^*$-a.s.,

$$M_T = V_0 + \sum_{k=1}^{T} \xi_k \cdot \Delta X_k.$$

By Proposition 8.27, the corresponding value process V is a martingale under $\mathbb{P}^*$ and satisfies $V_t = \mathbb{E}^*[M_T|\mathcal{F}_t]$. Hence $M_t = V_t$ for all t. But we also know from (8.9) that $V_t = V_0 + \sum_{k=1}^{t} \xi_k \cdot \Delta X_k$, which proves the assertion for positive martingales as $V_0 = M_0$. The general case follows by the decomposition $M_T = M_T^+ - M_T^-$.

(iv) $\Rightarrow$ (i): Pick $E \in \mathcal{F}_T$ and put $M_t = \mathbb{E}^*[\mathbf{1}_E|\mathcal{F}_t]$, $t \in \{0, \ldots, T\}$. Then the assumption implies that $\mathbf{1}_E$ is an attainable claim. According to Theorem 8.30, it has a unique arbitrage free price, $\mathbb{P}^*(E)$. Hence for all $\mathbb{P}^* \in \mathcal{P}$, we have that $\mathbb{P}^*(E)$ is one and the same number. Since E is arbitrary, $\mathcal{P}$ must be a singleton. But then the market is complete in view of Theorem 8.32. □

Remark 8.34 Property (iv) of Theorem 8.33 is also called the discrete time Martingale Representation Theorem, similar to a theorem for so called Brownian martingales in continuous time.

8.6 CRR model

In this section we consider the Cox-Ross-Rubinstein (CRR) model, a popular model of a financial market in discrete time. Apart from its tractability and that of related pricing issues, it is also interesting, because pricing formulas tend to Black-Scholes related formulas under the right kind of asymptotics. We will not treat this aspect in the present course.

In the CRR model, there is a riskless asset, whose price evolves according to

$$S_t^0 = (1+r)^t,\ t \in \{0, \dots, T\},$$

for some $r \in (-1, \infty)$, although usually $r \geq 0$. There is only one risky asset with price process $S^1 =: S$, whose *relative returns*

$$R_t := \frac{\Delta S_t}{S_{t-1}}\ t \in \{1, \dots, T\} \tag{8.15}$$

are random variables greater than -1. Equation (8.15) is equivalent to $S_t = (1+R_t)S_{t-1}$ for $t \geq 1$ and two useful relations follow, $S_t = S_0 \prod_{k=1}^t (1+R_k)$ and $S_T = S_t \prod_{k=t+1}^T (1+R_k)$.

It is assumed that R_t at each time t assumes only two values, which are even the same for all $t \geq 1$, say a and b, with $a < b$. The simplest probability space that carries all random variables below, assuming a finite time horizon T, is $\Omega = \{a, b\}^T$. The obvious filtration is such that $\mathcal{F}_t = \sigma(R_1, \dots, R_t)$, $t \in \{1, \dots, T\}$ and $\mathcal{F}_0$ trivial. In this case, any sensible probability measure on $\mathcal{F}_T$ must be such that all singletons have positive probability. The totality of all these conventions will be referred to as the CRR model. We will see that absence of arbitrage has a simple characterization in terms of the parameters a, b and r.

We use X to denote the discounted price process of the risky asset, so

$$X_t = \frac{S_t}{S_t^0},$$

and note that for $t \in \{1, \dots, T\}$

$$\begin{aligned} X_t &= \frac{1+R_t}{1+r} X_{t-1}, \\ \Delta X_t &= \frac{R_t - r}{1+r} X_{t-1}. \end{aligned}$$

Proposition 8.35 *The CRR model is arbitrage-free iff $a < r < b$. Moreover, if it is arbitrage-free, it is also complete. The unique equivalent martingale measure is such that the R_t become i.i.d. random variables, whose common distribution is determined by*

$$\mathbb{P}^*(R_t = b) = \frac{r-a}{b-a} =: p^*.$$

Proof First we show that if a martingale measure exists, it is necessarily unique. Let $\mathbb{Q}$ be a martingale measure and $t \in \{1, \dots, T\}$. Then

$$X_{t-1} = \mathbb{E}_{\mathbb{Q}}[X_t|\mathcal{F}_{t-1}] = X_{t-1}\mathbb{E}_{\mathbb{Q}}[\frac{1+R_t}{1+r}|\mathcal{F}_{t-1}].$$

Since X_{t-1} is positive $\mathbb{Q}$-a.s., we can divide this equation by it and conclude

$$\mathbb{E}_{\mathbb{Q}}[R_t|\mathcal{F}_{t-1}] = r. \tag{8.16}$$

Let $q = \mathbb{Q}(R_t = b|\mathcal{F}_{t-1}) = 1 - \mathbb{Q}(R_t = a|\mathcal{F}_{t-1})$. Then we can rewrite (8.16) as $qb + (1-q)a = r$, which yields

$$q = \mathbb{Q}(R_t = b|\mathcal{F}_{t-1}) = \frac{r-a}{b-a}.$$

This implies that R_t is, under $\mathbb{Q}$, independent of $\mathcal{F}_{t-1}$ and that its unconditional distribution is also given by $\mathbb{Q}(R_t = b) = q$. It follows that, necessarily, the R_t are i.i.d. under $\mathbb{Q}$, and hence $\mathbb{Q}$ must be unique. Note that we have $q = p^*$. For $\mathbb{Q}$ to be a probability measure, we need $p^* \in [0, 1]$, which is equivalent to $r \in [a, b]$. To have that $\mathbb{Q}$ is *equivalent* to $\mathbb{P}$, $\mathbb{Q} \sim \mathbb{P}$, $p^* \in \{0, 1\}$ is to be excluded. In that case $a < r < b$.

Let the market be arbitrage free. Then there exists an equivalent martingale measure $\mathbb{P}^*$. By the above reasoning, we necessarily have that $\mathbb{P}^*$ is as asserted. The market is then also complete in view of Theorem 8.33.

If the condition $a < r < b$ holds true, then we can define the measure $\mathbb{P}^*$ on Ω, by putting

$$\mathbb{P}^*(\{\omega\}) = (p^*)^{k(\omega)}(1-p^*)^{T-k(\omega)},$$

where $k(\omega)$ denotes the number of b's in ω. Clearly we have $\mathbb{P}^* \sim \mathbb{P}$, independence of the R_t follows and we also see that the marginal distribution of each R_t is the same as for the others. We have seen above that $\mathbb{P}^*$ defines a martingale measure. □

We turn to the pricing of contingent claims. Recall that they have a unique price by completeness of the market. Consider a discounted claim H. Since H is $\mathcal{F}_T$-measurable, there exists, see Proposition B.10, a function $h : \Omega \to \mathbb{R}$ such that

$$H = h(R_1, \dots, R_T). \tag{8.17}$$

The value process for H is, whatever replicating strategy (but in Exercise 8.8 it is shown to be unique), given by

$$V_t = \mathbb{E}^*[H|\mathcal{F}_t].$$

In what follows we need a property of conditional expectations, sometimes called the independence lemma, part (iv) of Theorem B.34, which we recall here for convenience.

Lemma 8.36 *If $\mathcal{G}$ is a σ-algebra, X is a $\mathcal{G}$-measurable random variable (or vector), Y is independent of $\mathcal{G}$ and f is a measurable function such that the expectations below exist, then $\mathbb{E}[f(X,Y)|\mathcal{G}] = \hat{f}(X)$, where $\hat{f}(x) = \mathbb{E}f(x,Y)$.*

Proof See measure theory for the general case and Exercise 8.16 for a simple special case. □

Note that in Lemma 8.36 the conditional expectation is obtained by taking expectation w.r.t. Y, i.e. integrating out the variable Y only, and leaving X untouched.

We continue with the CRR model. Let $r_j \in \{a,b\}$, $j \in \{1,\dots,T\}$, put $v_T(r_1,\dots,r_T) = h(r_1,\dots,r_T)$ and for $t \in \{1,\dots,T-1\}$

$$v_t(r_1,\dots,r_t) = \mathbb{E}^* h(r_1,\dots,r_t,R_{t+1},\dots,R_T),$$

and $v_0 = \mathbb{E}^* H$. Exploiting the independence of the R_t and using the *independence lemma*, Lemma 8.36, we get for all $t \in \{0,\dots,T\}$ that

$$V_t = v_t(R_1,\dots,R_t).$$

Moreover, using the martingale property of V under $\mathbb{P}^*$, we similarly obtain the backward recursion

$$v_{t-1}(r_1,\dots,r_{t-1}) = p^* v_t(r_1,\dots,r_{t-1},b) + (1-p^*) v_t(r_1,\dots,r_{t-1},a).$$

If the discounted claim H only depends on the terminal price S_T, then we have $H = k(S_T)$, for some function k. The relation between k and the above h is

$$k\big(S_0(1+r_1)\cdots(1+r_T)\big) = h(r_1,\dots,r_T).$$

Put $w_t(s) = \mathbb{E}^* k\big(s\,(1+R_{t+1})\cdots(1+R_T)\big)$. Then we can alternatively write

$$w_t(s) = \mathbb{E}^* k\big(s\frac{S_T}{S_t}\big), \tag{8.18}$$

from which (use the independence lemma again, $\frac{S_T}{S_t}$ is independent of $\mathcal{F}_t$) it follows that

$$w_t(S_t) = \mathbb{E}^*[k\big(S_T\big)|\mathcal{F}_t].$$

Between v_t and w_t one has the relation

$$v_t(r_1,\dots,r_t) = w_t(S_0(1+r_1)\cdots(1+r_t)),$$

and hence $V_t = w_t(S_t)$. See also Exercises 8.7 and 8.8.

8.7 Exercises

8.1 Prove Proposition 8.3.

8.2 Prove Proposition 8.14. *Hint:* Apply Jensen's inequality to $\mathbb{E}_{\tilde{\mathbb{P}}} \frac{X_T^0}{X_T^1}$.

8.3 Fix the time horizon at T and assume the initial σ-algebra $\mathcal{F}_0$ to be trivial. Let S_t^0 be identically equal to 1 and let $Z_t = \log \frac{S_t^1}{S_{t-1}^1}$. Suppose that the market that is described by the pair of processes S^0, S^1 is arbitrage-free. Suppose that $\mathbb{P}$ is such that the Z_t are i.i.d. with a common normal $N(\mu, \sigma^2)$ distribution. Give a relation between the parameters if $\mathbb{P}$ is a martingale measure.

8.4 Consider an arbitrage-free market with one risky asset. Let S^1 be its price process and S^0 the deterministic price process of the riskless asset. Consider a European call option with discounted payoff

$$H = \frac{(S_T^1 - K)^+}{S_T^0},$$

for some $K > 0$. Assume that S_T^1 has a density w.r.t. Lebesgue measure under any risk-neutral measure. Let π^* be an arbitrage-free price of the call option under some risk-neutral measure $\mathbb{P}^*$. Obviously π^* depends on K and S_0^1, so we write $\pi^* = \pi^*(K, S_0^1)$. Show that

$$0 < \frac{\partial \pi^*}{\partial S_0^1} < 1$$
$$\frac{\partial \pi^*}{\partial K} = -\frac{1}{S_T^0}(1 - F^*(K)),$$

where F^* is the distribution function of S_T^1 under $\mathbb{P}^*$. To show the first assertion you may make additional assumptions, e.g. that S_T^1 is increasing in S_0^1, or even more explicit, $S_T^1 = S_0^1 R_T$, with R_T a positive random variable.

8.5 Consider a market with underlying $\Omega = \{1, 2, 3, 4\}$. Assume that $T = 2$ and that $S_t^0 = 1$ for $t = 0, 1, 2$, the price of the riskless asset is constant. Let the evolution for the price S_t of the *single* risky asset be as given in the table.

ω	$S_0(\omega)$	$S_1(\omega)$	$S_2(\omega)$
1	5	8	9
2	5	8	6
3	5	4	6
4	5	4	3

(a) Assume that $\mathbb{P}$ gives positive probability to each singleton. Show that the market is complete and that $\mathbb{P}^*$ as represented by the vector $(\frac{1}{6}, \frac{1}{12}, \frac{1}{4}, \frac{1}{2})$ is the unique equivalent martingale measure.

(b) Let H be the European call option $H = (S_2 - 5)^+$. Let $\bar{\xi}$ be the replicating strategy. Show that $\bar{\xi}_2(\omega) = (-5, 1)$ if $\omega = 1, 2$ and $\bar{\xi}_2(\omega) = (-1, \frac{1}{3})$ if $\omega = 3, 4$.

(c) Compute V_1 and show that $\bar{\xi}_1(\omega) = (-\frac{7}{3}, \frac{2}{3})$ for all ω. What is the value of the claim at $t = 0$?

(d) As an alternative you can use the self-financing property in the form $V_t = V_{t-1} + \Delta G_t$, $t = 1, 2$. Use this to compute the replicating strategy anew.

(e) Suppose that the riskless interest rate is r. For which possible values of r do we still have an arbitrage free market?

8.6 Consider the market of Exercise 8.5. Compute the value of the claim $H = (\frac{1}{3}(S_0 + S_1 + S_2) - 5)^+$.

8.7 Give an explicit formula for $w_t(s)$, see (8.18), as a sum involving the probabilities of the Binomial distribution with parameters $T - t$ and p^*.

8.8 Let H be a claim as in (8.17). Show that the hedge strategy is given by

$$\xi_t = (1+r)^t \frac{v_t(R_1, \ldots, R_{t-1}, b) - v_t(R_1, \ldots, R_{t-1}, a)}{S_{t-1}(b-a)}.$$

Give also an expression for ξ_t, if $H = h(S_T)$. What is the explicit resulting strategy if $H = (1+r)^{-T} S_T$?

8.9 Show that the probability measure $\mathbb{P}_2^*$ in the proof of Theorem 8.33 is a martingale measure.

8.10 Let $Y_1, \ldots, Y_T$ be *iid* random variables on some $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E} Y_t = 0$ for all t and let $\mathcal{F}_t = \sigma(Y_1, \ldots, Y_t)$. Let $X_t = \sum_{k=1}^t Y_k$ for $t \leq T$. Obviously, the X_t form a martingale. Consider an *insider trader*, a trader whose information pattern is given by the σ-algebras $\tilde{\mathcal{F}}_t := \sigma(\mathcal{F}_t \cup \sigma(X_T))$, i.e. at any time $t \leq T$ she 'knows' the final value X_T.

(a) Show that the X_t don't result in a martingale w.r.t. the enlarged filtration of the $\tilde{\mathcal{F}}_t$.

(b) Let $\tilde{X}_t = X_t - \sum_{k=0}^{t-1} \frac{X_T - X_k}{T-k}$, $t \leq T$. Show that the $\tilde{X}_t$ yield a martingale w.r.t. enlarged filtration. *Hint: use that* $\mathbb{E}[X_t | X_T] = \frac{t}{T} X_T$ *and independence of the* Y_k.

(c) Construct a self-financing strategy of investments $\tilde{\xi}_t$ w.r.t. the enlarged filtration (so the $\tilde{\xi}_t$ are $\tilde{\mathcal{F}}_{t-1}$-measurable) such that $\mathbb{E} \sum_{t=1}^T \xi_t (X_t - X_{t-1})$ is positive. This should follow from maximization of the expected gain $\mathbb{E} \sum_{t=1}^T \tilde{\xi}_t (X_t - X_{t-1})$ over all self-financing strategies such that $|\tilde{\xi}_t| \leq 1$.

8.11 Show that the σ_m^0 in the proof of Lemma 8.19 are $\mathcal{F}$-measurable, as well as the $\xi_{\sigma_m^0}$. Same question for the σ_m^1 and the $\xi_{\sigma_m^1}$. Finish the proof of Lemma 8.19.

8.12 The proof of Proposition 8.21 is a lot simpler if $\mathcal{F}_0$ is the trivial σ-algebra $\{\emptyset, \Omega\}$. In this case all ξ_n is the proof are just vectors in $\mathbb{R}^d$ and the spaces N and $N^\perp$ are closed linear subspaces of $\mathbb{R}^d$. Rewrite (and shorten) the proof under this additional assumption and make clear that Lemma 8.19 and Lemma 8.20 can be circumvented by using standard analysis arguments instead.

8.13 Show that it follows from the proof of Proposition 8.21 that under the same assumption also $\mathcal{K}$ is closed in L^0.

8.14 If one drops the no arbitrage assumption in Proposition 8.21, the assertion is no longer true in general. This exercise contains an example. Assume that market contains only one risky asset ($d = 1$). Let in $(\Omega, \mathcal{F}, \mathbb{P})$, $\Omega = [0, 1]$, $\mathcal{F}_1$ the Borel σ-algebra, and $\mathbb{P}$ the Lebesgue measure. Furthermore, $\mathcal{F}_0$ is the trivial sigma-algebra on $[0, 1]$. Assume that $Y : \Omega \to \mathbb{R}$ is given by $Y(\omega) = \omega$.

(a) Show that the no arbitrage condition is violated.

(b) Let $Z \geq 1$ be a constant. Show that Z cannot belong to $\mathcal{C}$, and conclude that $\mathcal{C}$ is not all of L^1.

(c) Let $Z \in L^1$ and define $Z_n = (Z^+ \wedge n)\mathbf{1}_{[\frac{1}{n},1]} - Z^-$. Show that $Z_n \in \mathcal{C}$ (establish first that $(Z^+ \wedge n)\mathbf{1}_{[\frac{1}{n},1]} \leq c_n Y$ for some constant c_n) and that $Z^n \to Z$ in L^1 for $n \to \infty$.

(d) Conclude that $\mathcal{C}$ is not closed.

8.15 Show that the probability measure $\mathbb{P}_0$ in the proof of Theorem 8.30 is a martingale measure.

8.16 Prove Lemma 8.36 for the special case in which X and Y are discrete random variables.

9 Optimization in dynamic models

In this section we study portfolio optimization over a nontrivial horizon. We thus extend the results of Section 6 to a dynamic case. We will present two methods to find an optimal portfolio, one is based on *Dynamic Programming*, the other is based on first finding an optimal random pay-off and then to construct a trading strategy that replicates this pay-off. That method is known as the *martingale method* or as the *risk neutral approach*.

9.1 Dynamic programming

Dynamic programming is a main tool in optimization for dynamic models, especially useful if the relevant underlying processes are defined by recursive models, or else have a model describing their time dependent behavior. This could mean e.g. that they are Markov processes or martingales. We first explain the two key ideas behind dynamic programming and then proceed with a more formal treatment. The proofs of the results in this section are either exercises or given in Appendix A.5

We give the ideas underlying dynamic programming in its most rudimentary form. Suppose one wants to maximize a function V of two variables, of the specific form

$$V(u_1, u_2) = V_1(u_1) + V_2(f(u_1), u_2).$$

This maximization problem can be carried out as the iterated maximization

$$\max_{u_1,u_2} V(u_1, u_2) = \max_{u_1}(V_1(u_1) + \max_{u_2} V_2(f(u_1), u_2)),$$

with the underlying idea that the functions V_i are to be interpreted as 'rewards' at 'times' $i = 1, 2$. The maximization on the right hand side over u_2, with any u_1 fixed, yields (assuming a maximizer exists and is unique) an optimal

$$u_2^*(u_1) = g(u_1),$$

for some function g. Substitution of this relation for u_2 in $V(u_1, u_2)$ yields a function of u_1 only,

$$V(u_1, g(u_1)) = V_1(u_1) + V_2(u_1, g(u_1)),$$

and maximizing over u_1 yields an optimal solution u_1^*, that in turn yields $u_2^* = g(u_1^*)$.

So, the (first) idea is to optimize over the second variable and then over the first one, and that, given that the first one yields an optimal value, the optimum of the second step is immediately known. Hence, if one views the pair (u_1^*, u_2^*) as some kind of optimal path to reach one's goal, then the second part of the path, u_2^*, is optimal once the 'starting value' u_1^* is given. This reflects the second idea behind dynamic programming, also called *Bellman's optimality principle*.

We move on to a random dynamic setting, where all random variables and processes are defined on some $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that one has an $\mathbb{R}^d$-valued stochastic process $X = (X_0, \ldots, X_T)$, in fact it is going to be a family of processes as we shall soon see, that obey the recursion

$$X_{t+1} = f_t(X_t, U_t, \varepsilon_t),\ t = 0, \ldots, T-1, \tag{9.1}$$

and that start in some value X_0.

Here the random quantities $X_0, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1}$ are given and they are assumed to be independent. The m-dimensional random variables U_t are supposed to be of the form

$$U_t = u_t(X_0, \ldots, X_t), \tag{9.2}$$

for certain measurable functions $u_t : (\mathbb{R}^d)^{t+1} \to \mathbb{R}^m$, for which we use the notation $u_t \in \mathcal{B}((\mathbb{R}^d)^{t+1}, \mathbb{R}^m)$. If the ε_t are k-dimensional, the f_t in (9.1) are defined on (appropriate subsets of) $\mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^k$, and are also assumed to be jointly measurable in their arguments. As a filtration we take the family of σ-algebras $\mathcal{F}_t = \sigma(X_0, \varepsilon_0, \ldots, \varepsilon_{t-1})$. Then the processes X and U are adapted. Moreover, if $U_t = u_t(X_t)$ (as we shall see below, this is an important case), the resulting process X is even Markov.

Lemma 9.1 *The process X is Markov w.r.t. the filtration specified above, if U_t depends on X_t only, $U_t = u_t(X_t)$ say, where the u_t are measurable functions $u_t : \mathbb{R}^d \to \mathbb{R}^m$.*

Proof Exercise 9.1. □

Remark 9.2 Under the conditions of Lemma 9.1, we have $X_{t+1} = F_t(X_t, \varepsilon_t)$ for $F_t(x, y) = f_t(x, u_t(x), y)$. But a similar structure is also valid, if e.g. $U_t = u_t(X_t, X_{t-1})$ by suitable rewriting. Indeed, let $\mathbf{X}_t = (X_t, X_{t-1})$, $x = (x_1, x_2)$. Then we obtain $\mathbf{X}_{t+1} = \mathbf{F}_t(\mathbf{X}_t, \varepsilon_t)$, where $\mathbf{F}_t(x, y) = (f_t(x_1, u_t(x_1, x_2), y), x_1)$. This shows that the setup of (9.1) and (9.2) is more general then it may appear at first glance, and it includes k-step Markov processes as well.

Our aim is to solve the following problem.

Problem 9.3 Let $g_0, \ldots, g_{T-1} : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ be measurable functions, as well as $g_T : \mathbb{R}^d \to \mathbb{R}$. The problem is to maximize over $U = (U_0, \ldots, U_{T-1})$ the expectation

$$J(U) := \mathbb{E}\big(\sum_{t=0}^{T-1} g_t(X_t, U_t) + g_T(X_T)\big),$$

with each U_t as in (9.2). This problem is thus equivalent to the finding of measurable functions u_t such that the constraint (9.2) holds. With $u = (u_0, \ldots, u_{T-1})$ we also write $J(u)$ instead of $J(U)$ to emphasize that J depends on the *functions u_t*. Usually, the functions have to satisfy certain constraints. These will be clear in the appropriate context and not always explicitly mentioned. For instance, it is tacitly assumed that all random quantities involved are such that the expectations exist.

Definition 9.4 A sequence of functions $u^* = (u_0^*, \dots, u_{T-1}^*)$ is called optimal if $J(u^*) = \sup J(u)$ holds, where the supremum is taken over sequences $u = (u_0, \dots, u_{T-1})$ with u_t in $\mathcal{B}((\mathbb{R}^d)^{t+1}, \mathbb{R}^m)$ for $t = 0, \dots, T-1$.

Definition 9.5 Let u be a sequence of functions $(u_0, \dots, u_{T-1})$ and let the process X^u be defined by (9.1) and (9.2), the notation expresses the dependence of X on u. Note that $X_0^u = X_0$. Then, we define $J_T(u) = g_T(X_T^u)$, and for $t < T$,

$$J_t(u) := \mathbb{E}[\sum_{s=t}^{T-1} g_s(X_s^u, U_s) + g_T(X_T^u)|\mathcal{F}_t].$$

Then, for $t < T-1$,

$$\begin{aligned}
J_t(u) &= g_t(X_t^u, U_t) + \mathbb{E}[\sum_{s=t+1}^{T-1} g_s(X_s^u, U_s) + g_T(X_T^u)|\mathcal{F}_t] \\
&= g_t(X_t^u, U_t) + \mathbb{E}[\mathbb{E}[\sum_{s=t+1}^{T-1} g_s(X_s^u, U_s) + g_T(X_T^u)|\mathcal{F}_{t+1}]|\mathcal{F}_t] \\
&= g_t(X_t^u, U_t) + \mathbb{E}[J_{t+1}(u)|\mathcal{F}_t].
\end{aligned}$$

Note that $\mathbb{E}J_0(u) = J(u)$. The J_t can be interpreted as expected future rewards, given the past up to time t.

Define also for certain given measurable functions $v_0, \dots, v_T : \mathbb{R}^d \to \mathbb{R}$

$$\hat{v}_{t+1}(x, y) = \mathbb{E}v_{t+1}(f_t(x, y, \varepsilon_t)),\ t = 0, \dots, T-1, \tag{9.3}$$

so $\hat{v}_{t+1} : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$. Note that these functions are measurable in x and y.

Important is the situation in which all the U_t are such that $U_t = u_t(X_t)$, for some measurable functions $u_t : \mathbb{R}^d \to \mathbb{R}^m$. Denote the class of sequences $u = (u_0, \dots, u_{T-1})$ of such functions by $\mathcal{M}$. Here is the main theorem of this section.

Theorem 9.6 *Define recursively the functions v_t, $t = 0, \dots, T$, by*

$$\begin{aligned}
v_T(x) &= g_T(x) \\
v_t(x) &= \sup_y \{g_t(x, y) + \hat{v}_{t+1}(x, y)\},\ t = 0, \dots, T-1.
\end{aligned} \tag{9.4}$$

Assume that the v_t are measurable functions. Then the following hold true.

(i) For any sequence $u \in \mathcal{M}$ of functions u_t one has

$$v_t(X_t^u) \geq J_t(u) \text{ a.s.}$$

and $\mathbb{E}v_0(X_0) \geq J(u)$.

(ii) Let $u^* \in \mathcal{M}$. Then u^* is optimal iff the supremum in (9.4) is attained for $y = u_t^*(x)$. If this happens, then $v_t(X_t^{u^*}) = J_t(u^*)$ and $\sup_u J(u) = J(u^*) = \mathbb{E}v_0(X_0)$.

Proof See Section A.5. □

The condition in Theorem 9.6 that the functions v_t are measurable, is satisfied under the additional assumption that both $g_t(x,\cdot)$ and $\hat{v}_{t+1}(x,\cdot)$ are continuous for all t and x. Theorem 9.6 also provides an algorithm that yields the optimal functions u_t^*.

Algorithm 9.7 (Dynamic programming) Suppose that the suprema in Equation (9.4) are attained for all t. Define

$$\begin{aligned} v_T(x) &= g_T(x) \\ u_{T-1}^*(x) &= \arg\sup_y \{g_{T-1}(x,y) + \hat{v}_T(x,y)\}, \end{aligned}$$

and by backwards recursion for $t \in \{0,\ldots,T-1\}$

$$\begin{aligned} v_t(x) &= \sup_y \{g_t(x,y) + \hat{v}_{t+1}(x,y)\} \\ &= g_t(x,u_t^*(x)) + \hat{v}_{t+1}(x,u_t^*(x)) \\ u_{t-1}^*(x) &= \arg\sup_y \{g_{t-1}(x,y) + \hat{v}_t(x,y)\}. \end{aligned}$$

This yields the sequence of functions $v_T, u_{T-1}^*, v_{T-1}, u_{T-2}^*, \ldots, v_1, u_0^*, v_0$ where the u_t^* constitute the optimal sequence u^* and $\mathbb{E}v_0(X_0) = J(u^*)$. The functions v_t are called the *(optimal) value functions.*

Proposition 9.8 (Optimality principle) *Let $u^* = (u_0^*,\ldots,u_{T-1}^*)$ be the optimal sequence for Problem 9.3 as obtained from Algorithm 9.7. Then the sequence $(u_t^*,\ldots,u_{T-1}^*)$ is optimal for the corresponding optimization problem over the time set $\{t,\ldots,T\}$, when starting in $X_t = X_t^{u^*}$. In this case the optimal value is equal to $\mathbb{E}v_t(X_t^{u^*})$.*

Proof Exercise 9.2. □

Remark 9.9 The results above strongly depend on the fact that $\hat{v}_{t+1}(x,y)$ is equal to the conditional expectation $\mathbb{E}[v_{t+1}(f_t(x,y,\varepsilon_t))|\mathcal{F}_t]$, which follows from the assumed independence of X_0 and the ε_t, see part (iv) of Theorem B.34. In Section 9.2 however, we will come across situations, where this assumption is often violated. We proceed with giving some results for a more general setting.

From here on, *we drop the assumption that the ε_t are independent.* One can still define 'functions' $\hat{v}_{t+1}$, but now we alter the definition of (9.3) into

$$\hat{v}_{t+1}(x,y) = \mathbb{E}[v_{t+1}(f_t(x,y,\varepsilon_t))|\mathcal{F}_t]. \tag{9.5}$$

Then the $\hat{v}_{t+1}(x,y)$ are in general not deterministic anymore, but become $\mathcal{F}_t$-measurable random variables. However, many of the above results continue to hold. For instance, the technical result (A.7) is still correct. On the other hand, we need an alternative to Lemma A.15, used in the proof of Theorem 9.6, and to the theorem itself. The following proposition uses the concept of *essential supremum*, see Section A.3. We will consider $\mathcal{B}(\mathbb{R}^d) \times \mathcal{F}_t$-measurable functions v_t, meaning that the mapping $(x, \omega) \mapsto v_t(x, \omega)$ is $\mathcal{B}(\mathbb{R}^d) \times \mathcal{F}_t$-measurable. As usual, dependence on ω is often suppressed and then we write $v_t(x)$ for the random variable $\omega \mapsto v_t(x, \omega)$.

Proposition 9.10 *Suppose that $\mathcal{B}(\mathbb{R}^d) \times \mathcal{F}_t$-measurable functions v_t (for $t = 0, \dots, T$) a.s. satisfy*

$$
\begin{aligned}
v_T(x) &= g_T(x) \\
v_t(x) &= \operatorname{ess\,sup}\{g_t(x, y) + \mathbb{E}[v_{t+1}(f_t(x, y, \varepsilon_t))|\mathcal{F}_t] : y \in \mathcal{F}_t\},\ t \leq T - 1.
\end{aligned}
$$

Then, for any sequence $u = (u_0, \dots, u_{T-1})$ it holds that $v_t(X_t^u) \geq J_t(u)$ for $t = 0, \dots, T$. Optimality is obtained for the $\mathcal{F}_t$-measurable random variables $u_t^*(x)$ for which the essential supremum is attained.

Proof Essentially as the proofs of Lemma A.15 and Theorem 9.6. □

There is a major difference between the above proposition and previous results. The optimizing $u_t^*(x)$ are now $\mathcal{F}_t$-measurable random variables, and to emphasize this, we should write $u_t^*(x) = u_t^*(x, \omega)$. Writing $u_t^*(X_t^{u^*})$ suggests explicit dependence of an optimal control through $X_t^{u^*}$ only, but this is in general not the case. The dependence on ω, is usually through $X_0^u, \dots, X_t^u$ for $u = u^*$. This is particularly true if $\mathcal{F}_t = \mathcal{F}_t^{X^u} := \sigma(X_0^u, \dots, X_t^u)$ for $u = u^*$, or if $\mathcal{F}_t$ can be replaced with $\mathcal{F}_t^{X^u}$ in (9.5) and in Proposition 9.10. Note that these σ-algebras in principle depend on the controls u. The good news is that in many relevant practical situations this doesn't pose a problem and that it is still possible to explicitly compute the optimal strategy, see for instance Exercise 9.4 for a problem in a financial context.

9.2 Optimal portfolios via dynamic programming

We apply the general results of the previous section to an expected utility maximization problem. We assume that the assumptions of Section 6 are in force. An investor has an initial capital w at his disposal, also called initial wealth. He can invest in shares in a market described by the model of Section 8. Without further explanation, below we use the notation of that section. The first problem we consider is the problem of maximizing the expected utility of terminal wealth, similar to Equation (6.1). So we want to maximize

$$\mathbb{E}\tilde{u}(W_T),$$

subject to the budget constraint $W_0 \leq w$ and to the constraint that W_T results from a self-financing trading strategy $\bar{\xi}$ with $W_T = \bar{\xi}_T \cdot \bar{S}_T$. Writing $W_T = S_T^0 V_T$ and $V_T = V_0 + G_T$, we see that we have to maximize

$$\mathbb{E}\tilde{u}(S_T^0(w + G_T)),$$

where of course we have taken $V_0 = W_0 = w$, since again it can never be optimal to use only a fraction of the initial capital w. As before, we assume that the process S^0 is deterministic, non-risky. Then we can define a new utility function u by putting $u(x) = \tilde{u}(S^0_T x)$. Hence we have to find

$$\max \mathbb{E}u(w + G_T),$$

where $G_T = \sum_{t=1}^T \xi_t \cdot \Delta X_t$ and the ξ_t are d-dimensional $\mathcal{F}_{t-1}$-measurable random vectors, also briefly denoted $\xi_t \in \mathcal{F}_{t-1}$. Notice that the ξ_t will be our (random) decision variables. In principle we'd like to apply the dynamic programming algorithm 9.7, with the proper substitutions and change of notation. For instance, we have that the functions g_t are zero for $t \leq T-1$ and the recursion $X_{t+1} = f_t(X_t, U_t, \varepsilon_t)$ in the present setting reads $V_{t+1} = V_t + \xi_{t+1} \cdot \Delta X_{t+1}$. Note that the control variables U_t in the general setting now become the ξ_t. Note however too that, unlike the ε_t, the ΔX_t are in general not independent, which spoils the fact that the optimal decisions at time t not only depend on V_t, but also on past values. Therefore the version of the dynamic programming algorithm that we need below is taken from Proposition 9.10. Define

$$\tilde{v}_T(x) = \tilde{u}(x)$$

and then, recursively, for $t \in \{0, \ldots, T-1\}$

$$\tilde{v}_t(x) = \operatorname{ess\,sup}\{\mathbb{E}[\tilde{v}_{t+1}(S^0_{t+1}(\frac{x}{S^0_t} + \xi \cdot \Delta X_{t+1}))|\mathcal{F}_t] : \xi \in \mathcal{F}_t\}. \tag{9.6}$$

Assume that for every t the essential supremum is attained at some $\xi^*_{t+1} = \xi^*_{t+1}(x) \in \mathcal{F}_t$. This eventually gives rise to a self-financing strategy $\bar{\xi}$ by the choices for the ξ^0_t as in Remark 8.5, the investments in the non-risky asset. At the final step of the algorithm ($t = 0$), we find ξ_0 and Theorem 9.6, or rather Proposition 9.10, tells us that $\mathbb{E}\tilde{v}_0(w)$ is the optimal value for our problem. Conditions for which the suprema are attained, for instance at an interior point of the domains of the $\tilde{v}_t$, can be derived from Theorem 6.6, although this theorem in general has to be applied 'ω-wise', which involves some subtleties.

In terms of the modified utility functions u, we can recast the optimal value functions resulting from Dynamic Programming as

$$v_T(x) = u(x)$$
$$v_t(x) = \operatorname{ess\,sup}\{\mathbb{E}[v_{t+1}(x + \xi \cdot \Delta X_{t+1})|\mathcal{F}_t] : \xi \in \mathcal{F}_t\},\ t \leq T-1. \tag{9.7}$$

One can show that this leads to the same optimum, if it exists (Exercise 9.5).

We generalize the above problem as to include *consumption*.

Definition 9.11 A *consumption process* $C = (C_0, \ldots, C_T)$ is a nonnegative adapted process. A consumption-investment plan is a pair $(C, \bar{\xi})$, where C is a consumption process and $\bar{\xi}$ a trading strategy. Such a plan is called *self-financing*, if

$$W_t = C_t + \bar{\xi}_{t+1} \cdot \bar{S}_t,\ t \in \{0, \ldots, T-1\}. \tag{9.8}$$

where W_t is as before, $W_t = \bar{\xi}_t \cdot \bar{S}_t$. Such a plan is called *admissible* if $C_T \leq W_T$ a.s.

Remark 9.12 The self-financing condition above is equivalent to the following two relations.

$$\Delta W_t = \bar{\xi}_t \cdot \Delta \bar{S}_t - C_{t-1},\ t \in \{1, \ldots, T\}$$
$$\Delta V_t = \xi_t \cdot \Delta X_t - \gamma_{t-1},\ t \in \{1, \ldots, T\},$$

where γ_{t-1} is discounted consumption at time $t-1$, $\gamma_{t-1} = \frac{C_{t-1}}{S^0_{t-1}}$. Note that

$$V_t - V_0 = \sum_{s=1}^{t} (\xi_s \cdot \Delta X_s - \gamma_{s-1})$$

and under an equivalent martingale measure one has

$$\mathbb{E}^*[\Delta V_t | \mathcal{F}_{t-1}] = \mathbb{E}^*[\xi_t \cdot \Delta X_t - \gamma_{t-1} | \mathcal{F}_{t-1}] = -\gamma_{t-1}.$$

The problem we are going to address is the maximization of

$$\mathbb{E} \sum_{t=1}^{T} \alpha^t u(C_t),$$

where $\alpha \in (0, 1]$ (a discount factor) and u a utility function, subject to the constraints that C forms together with a trading strategy $\bar{\xi}$ an admissible consumption-investment plan. Note that, for simplicity, we use a single utility function at all times t, although it can be made time dependent too.

It is immediately clear that for the optimum one must have $C_T = W_T$, since u is increasing. This makes the first step in the dynamic programming approach easy, $v_T = u$. In order to motivate the resulting backward recursion, we now consider the problem at time $T-1$, assuming that we have (optimally) invested and consumed up to that time. This can be viewed as a one period problem. We have to maximize

$$u(C_{T-1}) + \alpha\, \mathbb{E}[v_T(W_T) | \mathcal{F}_{T-1}] \tag{9.9}$$

subject to constraints that we now derive. Let W_{T-1} denote the wealth at time $T-1$. By the self-financing condition (9.8) for $t = T-1$ we have

$$W_{T-1} = C_{T-1} + \bar{\xi}_T \cdot \bar{S}_{T-1},$$

whereas

$$W_T = \bar{\xi}_T \cdot \bar{S}_T.$$

To assure that the (optimal) consumption process is non-negative, we assume *without loss of generality* that $u(x) = -\infty$ for $x < 0$. Using the self-financing characterization as in Remark 9.12 for $t = T$, we can rewrite (9.9) with $W'_{T-1} = \frac{W_{T-1}}{S^0_{T-1}}$ as

$$u(C_{T-1}) + \alpha\,\mathbb{E}[v_T(S^0_T(W'_{T-1} - \gamma_{T-1} + \xi_T \cdot \Delta X_T))|\mathcal{F}_{T-1}],$$

which we have to maximize over $\xi_T \in \mathcal{F}_{T-1}$. Of course we can iterate this procedure to get the dynamic programming equation for every $t \in \{1, \ldots, T\}$

$$\begin{aligned}(9.10)\quad v_{t-1}(w') = \operatorname{ess\,sup}\{&u(C_{t-1})\\ &+ \alpha\,\mathbb{E}[v_t(S^0_t(w' - \gamma_{t-1} + \xi_t \cdot \Delta X_t))|\mathcal{F}_{t-1}] : \xi_t, \gamma_{t-1} \in \mathcal{F}_{t-1}\},\end{aligned}$$

where w' denotes the value of the discounted wealth W'_{t-1}. If they exist, this yields the a.s. optimal ξ^*_t, γ^*_{t-1}, the maximizers at time $t-1$ of (9.10). At the final step, now denoting as before the initial capital by w, one obtains $v_0(w') = v_0(w)$, which will then yields the optimal value $\mathbb{E}v_0(w)$. The optimal processes C^* (and its discounted version γ^*) and ξ^* –assuming that they exist– have to be complemented by the process ξ^0 to obey the self-financing restriction. Similar to Remark 8.5, in the present context one has the recursion

$$\xi^0_{t+1} = \xi^0_t - (\xi^*_{t+1} - \xi^*_t) \cdot X_t - \gamma^*_t,$$

with initial value, given V_0,

$$\xi^0_1 = V_0 - \xi^*_1 \cdot X_0 - \gamma^*_0.$$

9.3 Consumption-investment and the martingale method

In this section we treat the martingale method (also called the risk-neutral approach) to solve consumption-investment optimization problems. In Section 9.3.1 we consider a one-period problem to introduce this method, which then forms the basis of the approach in the dynamic context of Section 9.3.2.

We mostly assume that *the market is complete*. For some of the problems that we introduce later on, this assumption is not always needed. This will be explained, when those are treated.

9.3.1 The static case

To illustrate the underlying principles of the martingale method, we first recall a static one-period problem, which is the maximization of $\mathbb{E}u(W_1)$. Here W_1 represents the value of a portfolio $\bar{\xi}$ at $t = 1$. Hence the maximization takes place over all $\bar{\xi} \in \mathbb{R}^{d+1}$, that of course have to satisfy the budget constraint $\bar{\xi} \cdot \bar{\pi} = w$, where w is the initially available capital.

The central idea behind the martingale method is to break down the optimization problem into two subproblems. The first one is to identify the optimal (random) pay-off W^*_1, given the budget constraint. The second one is then to identify the optimal portfolio $\bar{\xi}$, i.e. the portfolio whose terminal wealth is equal to W_1^*. The standing assumption is that the market is arbitrage free, a necessary and sufficient condition for the existence of an optimizer when considering expected utility maximization, see Theorem 6.5. For simplicity and focussing on the main ideas, we assume for the time being that *the price of the riskless asset is constant and equal to one*. Let $\mathbb{P}^*$ be the unique risk-neutral measure, recall the assumption of a complete market. Then the budget constraint is given by $\mathbb{E}^* W_1 = w$. Let $\mathcal{B}_w$ be the set of random variables that are integrable w.r.t. $\mathbb{P}^*$ with corresponding expectation equal to w. Hence, the first problem becomes the maximization of $\mathbb{E}u(W)$, subject to $W \in \mathcal{B}_w$, where we write W instead of W_1. Let $\phi = \frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}$. The budget constraint can then alternatively be expressed as $\mathbb{E}(\phi W) = w$.

One way to solve this problem is to employ a *Lagrange multiplier* (similar, but alternative to what we did in Section 7.1). So one likes to maximize

$$L(W, \lambda) = \mathbb{E}u(W) - \lambda(\mathbb{E}(\phi W) - w).$$

Differentiation of L w.r.t. λ yields the budget constraint. Next, one would like to differentiate w.r.t. W, which is in principle, as a function on some Ω, an infinite dimensional object. This would lead to some *variational problem* as in Section 7.1. To circumvent a full treatment of this approach, we think for a while that the underlying Ω is finite, $\Omega = \{\omega_1, \ldots, \omega_n\}$ with positive probabilities p_j of all singletons. Then we can represent W by a finite dimensional vector with a generic element $w_j = W(\omega_j)$ and

$$\begin{aligned} L(W, \lambda) &= \sum_j p_j u(w_j) - \lambda(\sum_j p_j \phi(\omega_j) w_j) - w) \\ &= \sum_j p_j u(W(\omega_j)) - \lambda(\sum_j p_j \phi(\omega_j) W(\omega_j) - w). \end{aligned}$$

Assume that u is differentiable and write W^* for the optimal payoff. Differentiation w.r.t. w_j of L can now be carried out under the expectation and yields in the optimum, after dividing $\frac{\partial L}{\partial w_j}$ by p_j,

$$u'(W^*(\omega_j)) - \lambda \phi(\omega_j) = 0. \tag{9.11}$$

Since this equation has to hold for every ω_j, we can multiply by p_j and sum to obtain, using that $\mathbb{E}\phi = 1$, for the Lagrange multiplier

$$\lambda = \mathbb{E}u'(W^*). \tag{9.12}$$

Compare this and the rest of this paragraph to the results in Section 7.1. Let I denote the inverse of u', which is assumed to exist (otherwise, one should work with the function I^+ as in Section 7.1). Then, from (9.11),

$$W^*(\omega_j) = I(\lambda \phi(\omega_j)),$$

or, in short,

$$W^* = I(\lambda \phi). \tag{9.13}$$

Dropping the assumption that Ω is finite, we conjecture that Equations (9.12) and (9.13) are needed to obtain the optimal claim. Theorem 7.2 justifies the optimality of (9.13), since the constant c there is nothing else but λ here. The budget restriction tells us that also for the optimal W^* it must hold that $\mathbb{E}^*W^* = w$, so we obtain the equation

$$\mathbb{E}^*I(\lambda\phi) = w. \tag{9.14}$$

Under conditions as in Corollary 7.5, a unique solution of (9.14) exists, which is then the optimal λ^*.

Knowing the optimal contingent claim, one then has to find a corresponding hedge strategy. Under the assumption that the market is complete one can find $\bar{\xi}^*$ such that $W^* = \bar{\xi}^* \cdot \bar{S}$ a.s. The resulting $\bar{\xi}^*$ should of course coincide with the solution of Theorem 6.6.

Let us now consider a consumption investment problem. First we pin down the admissible consumption-investment plans, see Definition 9.11, and an initial capital w. Recall that W_t is the notation for the undiscounted value of a portfolio at time t, which, under *the temporary assumption that* $S^0_t \equiv 1$, is equal to V_t. The plan is such that $C_0 + W_0 = w$ (spend all the initial capital) and $C_1 = W_1$, there is no life after $T = 1$ so all available wealth has to be consumed at time $T = 1$. Note that it follows that C_1 can be considered an attainable claim (see also Definition 9.13 below in the dynamic setting). Therefore, for any risk-neutral measure $\mathbb{P}^*$ it holds that $\mathbb{E}^*C_1 = W_0$ and we get

$$\mathbb{E}^*C_1 + C_0 = w. \tag{9.15}$$

There is also a converse reasoning. Suppose a consumption plan $C = (C_0, C_1)$ is fixed, as well as an initial capital w, such that (9.15) holds for any risk-neutral measure. Then the price $\mathbb{E}^*C_1$ of the consumption C_1 is unique and it follows from Proposition 1.19 that C_1 is an attainable claim, hence $C_1 = \bar{\xi} \cdot \bar{S}_1$, for some $\bar{\xi}$ and also $C_1 = W_1$.

A consumption-investment optimization problem is often formulated as the maximization of

$$u(C_0) + \alpha \mathbb{E}u(C_1),$$

subject to the constraints that $C_0, C_1 \geq 0$ a.s. and $C_0 + \mathbb{E}^*C_1 = w$. Note that the utility of C_0 and C_1 is represented by the same utility function u, but of course different choices for each of them are equally conceivable.

We adopt again the Lagrange multiplier approach to solve this problem. So we want to maximize

$$L(C_0, C_1, \lambda) = u(C_0) + \alpha \mathbb{E}u(C_1) - \lambda(C_0 + \mathbb{E}(\phi\, C_1) - w).$$

For the optimal consumption pair C^*_0, C^*_1 we then get, by the same token as we used before and assuming that $I = (u')^{-1}$ is well defined on $(0, \infty)$,

$$C^*_0 = I(\lambda), \tag{9.16}$$

(9.17) $$C_1^* = I(\lambda\phi/\alpha),$$

whereas the optimal λ^* has to solve the equation $I(\lambda)+\mathbb{E}^*I(\lambda\phi/\alpha) = w$ in view of (9.15). Optimality of C_0^* and C_1^* can be formally shown by an extension of the proof of Theorem 7.2.

9.3.2 The dynamic case

The approach in Section 9.3.1 extends to a consumption-investment problem with a time horizon $T > 1$ and also takes discounting into account. As we shall see, here we don't always need market completeness (due to the possibility of consumption), but of course we can't dispense with the requirement that the market is free of arbitrage.

First we consider the problem of determining the optimal final wealth, resulting from investments only. As in the static case, we first determine an $\mathcal{F}_T$-measurable random variable W^* that is such that $\mathbb{E}u(W)$ is maximal, subject to the constraint that $\mathbb{E}^*V = w$, where $V = W/S^0_T$ and w is an initially available capital. Mimicking the static case and taking care of the discount factor, we maximize, with ϕ equal to the Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}$, where the probability measures are now defined on $\mathcal{F}_T$, the Lagrangian

$$L(W,\lambda) = \mathbb{E}u(W) - \lambda(\mathbb{E}\phi\frac{W}{S^0_T} - w).$$

This leads to the following results, similar to (9.13) and (9.14),

$$W^* = I(\lambda\frac{\phi}{S^0_T}),$$
$$w = \mathbb{E}^*I(\lambda\frac{\phi}{S^0_T})/S^0_T.$$

If the market is assumed to be complete, we can in principle find a replicating strategy that yields W^* as its terminal value. What one in principle has to do is to find $\bar{\xi}_T \in \mathcal{F}_{T-1}$ such that $\bar{\xi}_T \cdot \bar{S}_T = W^*$. As soon as this has happened, we know by the fact that $\bar{\xi}$ is a self-financing strategy, that $\bar{\xi}_T \cdot \bar{S}_{T-1} = \bar{\xi}_{T-1} \cdot \bar{S}_{T-1}$, from which one has to determine $\bar{\xi}_{T-1} \in \mathcal{F}_{T-2}$. See Exercise 9.7 for an example that shows how to carry out this programme in a concrete situation.

We turn to consumption-investment problems. The market is, as always, assumed to be arbitrage-free, so a risk-neutral measures $\mathbb{P}^*$ exist.

Definition 9.13 A consumption process is called *attainable* with initial wealth w, if there exists a trading strategy $\bar{\xi}$ such that $(C,\bar{\xi})$ is self-financing, admissible and satisfies $C_T = W_T$. It is then said that $\bar{\xi}$ *replicates* C *(at time* T*).*

Let us first characterize attainable consumption processes. Recall Remark 9.12. Since the gains process G, $G_t = \sum_{s=1}^t \xi_s \cdot \Delta X_s$, is a martingale with expectation zero under any risk-neutral measure (see Theorem 8.9 for a precise statement), it follows that for each $t \le T$ one has, with $\gamma_s = \frac{C_s}{S_s^0}$ and $V_0 = w$,

$$\mathbb{E}^* V_t + \sum_{s=0}^{t-1} \mathbb{E}^* \gamma_s = w. \tag{9.18}$$

If the consumption process is attainable, $\gamma_T = V_T$, we get from (9.18) for $t = T$

$$\sum_{t=0}^{T} \mathbb{E}^* \gamma_t = w, \text{ for any } \mathbb{P}^* \in \mathcal{P}, \tag{9.19}$$

which is the dynamic counterpart of (9.15) under discounting.

Proposition 9.14 *Assume a complete market. Given any initial wealth $w \ge 0$, a consumption process is attainable iff (9.19) holds. In this case the corresponding value process of the replicating portfolio, and in particular its initial value, is nonnegative.*

Proof Necessity has already been proved. We turn to sufficiency. Any of the C_t and its discounted value γ_t in (9.19) is nonnegative and attainable at time t since the market is complete. So, for each of them, there exists a replicating self-financing strategy $\bar{\xi}^{(t)} = (\bar{\xi}_0^{(t)}, \ldots, \bar{\xi}_T^{(t)})$. Note that $\bar{\xi}_s^{(t)} = 0$ for $s > t$, since the claim C_t expires at time t and becomes worthless afterwards.

Put $\bar{\xi} = \sum_{k=0}^{T} \bar{\xi}^{(k)}$ and note that $\bar{\xi}_s = \sum_{k=s}^{T} \bar{\xi}_s^{(k)}$, in particular $\bar{\xi}_T = \bar{\xi}_T^{(T)}$. It trivially follows that $C_T = \bar{\xi}_T^{(T)} \cdot \bar{S}_T = \bar{\xi}_T \cdot \bar{S}_T$ and $\gamma_T = \bar{\xi}_T^{(T)} \cdot \bar{X}_T = \bar{\xi}_T \cdot \bar{X}_T = V_T$. Moreover, the C_t can be used to rebalance the portfolio in a self-financing way, i.e. $C_t + \bar{\xi}_{t+1} \cdot \bar{S}_t = W_t$, or, in discounted terms, $\gamma_t + \bar{\xi}_{t+1} \cdot \bar{X}_t = V_t$, $t \le T - 1$. This can be seen (for the non-discounted version) as below, where we use the special form of $\bar{\xi}_t$ and $\bar{\xi}_{t+1}$, the ordinary self-financing property of each of the processes $\xi^{(k)}$, and the fact that at time t, the claim C_t is hedged by the strategy $\bar{\xi}^{(t)}$. Write

$$\begin{aligned} W_t = \bar{\xi}_t \cdot \bar{S}_t &= \sum_{k=t}^{T} \bar{\xi}_t^{(k)} \cdot \bar{S}_t \\ &= \sum_{k=t+1}^{T} \bar{\xi}_t^{(k)} \cdot \bar{S}_t + \bar{\xi}_t^{(t)} \cdot \bar{S}_t \\ &= \sum_{k=t+1}^{T} \bar{\xi}_{t+1}^{(k)} \cdot \bar{S}_t + C_t = \bar{\xi}_{t+1} \cdot \bar{S}_t + C_t. \end{aligned}$$

The plan is thus self-financing in the sense of Definition 9.11, admissible, and the consumption process is attainable.

The martingale property of the gains process yields as in Remark 9.12

$$\mathbb{E}^*[\Delta V_s | \mathcal{F}_{s-1}] + \gamma_{s-1} = 0.$$

Summing this equation for $s = t+1, \ldots, T$ and taking conditional expectation given $\mathcal{F}_t$ yields

$$\mathbb{E}^*[V_T - V_t + \sum_{s=t}^{T-1} \gamma_s | \mathcal{F}_t] = 0.$$

Together with $V_T = \gamma_T$ this gives

$$V_t = \mathbb{E}^*[\sum_{s=t}^{T} \gamma_s | \mathcal{F}_t] \geq 0, \text{ a.s.}$$

In particular, $V_0 \geq 0$. Moreover, V_0 is equal to w, the initial capital, by (9.19). The consumption process is thus attainable with initial wealth w. □

After this intermediate result we are in the position to state the optimization problem properly. It is the maximization of

$$\mathbb{E}[\sum_{t=0}^{T} \alpha^t u(C_t)], \tag{9.20}$$

subject to the constraint that C is a nonnegative adapted process and the budget constraint (9.19). We will also assume that u is continuous on $[0, \infty)$, differentiable on $(0, \infty)$, $\lim_{x \downarrow 0} u'(x) = \infty$, and moreover $u(x) = -\infty$ for $x < 0$ when needed. These conditions are sufficient to obtain an a.s. strictly positive optimal consumption process. In view of Proposition 9.14, the resulting optimal consumption process C^* will be attainable. Once we have found this, we have to find the replicating strategy. Let us first focus on the finding of C^*. We will use the following lemma.

Lemma 9.15 *Let $\mathbb{P}^* \in \mathcal{P}$ and $\phi = \frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}}$, the Radon-Nikodym derivative on $\mathcal{F}_T$, and $\phi_t = \mathbb{E}[\phi | \mathcal{F}_t]$, $t = 0, \ldots, T$. Consider also $\mathbb{P}^*_t$ and $\mathbb{P}_t$, the restrictions of $\mathbb{P}^*$ and $\mathbb{P}$ to $\mathcal{F}_t$. Then $\phi_t = \frac{\mathrm{d}\mathbb{P}^*_t}{\mathrm{d}\mathbb{P}_t}$ and we can replace the constraint (9.19) with*

$$\mathbb{E} \sum_{t=0}^{T} \gamma_t \phi_t = w. \tag{9.21}$$

Proof Let $F \in \mathcal{F}_t$. The Radon-Nikodym property of ϕ_t follows from $\mathbb{P}^*_t(F) = \mathbb{E}^* \mathbf{1}_F = \mathbb{E}\phi \mathbf{1}_F = \mathbb{E}(\mathbb{E}[\phi \mathbf{1}_F | \mathcal{F}_t]) = \mathbb{E}(\mathbf{1}_F \mathbb{E}[\phi | \mathcal{F}_t]) = \mathbb{E}\mathbf{1}_F \phi_t$. In a similar way, (9.21) follows from $\mathbb{E}^* \gamma_t = \mathbb{E}\phi\gamma_t = \mathbb{E}(\mathbb{E}[\phi\gamma_t | \mathcal{F}_t]) = \mathbb{E}(\gamma_t \mathbb{E}[\phi | \mathcal{F}_t]) = \mathbb{E}\gamma_t \phi_t$, valid for all $t \in \{0, \ldots, T\}$. □

We will perform the maximization of (9.20) by again using a Lagrange multiplier approach, taking the alternative budget contraint (9.21) into account. We maximize

$$L(C_0, \ldots, C_T, \lambda) = \mathbb{E} \sum_{t=0}^{T} \alpha^t u(C_t) - \lambda \mathbb{E}(\sum_{t=0}^{T} \gamma_t \phi_t - w). \tag{9.22}$$

The necessary conditions for a maximum become, recall $\gamma_t = \frac{C_t}{S_t^0}$,

$$\alpha^t u'(C_t) - \lambda\tilde{\phi}_t = 0,\ t \in \{0, \dots, T\},$$

where $\tilde{\phi}_t = \frac{\phi_t}{S_t^0}$. Then, if $I = (u')^{-1}$ is well-defined everywhere, we obtain the optimal

$$C_t^* = I(\lambda\alpha^{-t}\tilde{\phi}_t),\ t \in \{0, \dots, T\},$$

and observe that C_t^* is $\mathcal{F}_t$-measurable. Of course the optimal λ^* has to satisfy

$$\mathbb{E}\sum_{t=0}^{T} I(\lambda^*\alpha^{-t}\tilde{\phi}_t)\tilde{\phi}_t = w.$$

Under the usual monotonicity and continuity assumptions, this equation has a unique solution λ^*, as in Corollary 7.5.

Here is an example that illustrates the above procedure.

Example 9.16 Let $u(x) = \log x$, $x > 0$. Then we have $I(x) = \frac{1}{x}$. We obtain

$$C_t^* = \frac{\alpha^t}{\lambda\tilde{\phi}_t},$$

and

$$w = \mathbb{E}\sum_{t=0}^{T}\frac{\alpha^t}{\lambda\tilde{\phi}_t}\tilde{\phi}_t = \sum_{t=0}^{T}\frac{\alpha^t}{\lambda},$$

from which it follows that

$$\lambda = \begin{cases} \frac{T+1}{w} & \text{if } \alpha = 1 \\ \frac{1-\alpha^{T+1}}{w(1-\alpha)} & \text{if } \alpha < 1. \end{cases}$$

In the first of these two cases, one can compute the optimal $C_t^* = \frac{w}{(T+1)\tilde{\phi}_t}$ and $\gamma_t^* = \frac{w}{(T+1)\phi_t}$. The maximal value of the objective function becomes $(T+1)\log\frac{w}{T+1} - \sum_{t=0}^{T}\mathbb{E}\log\tilde{\phi}_t$.

We continue to study an optimization problem that combines the previous two, we want to maximize expected utility derived from both consumption and terminal wealth. The chief difference with the previous problem is that we don't require $W_T = C_T$ anymore. Therefore, we have to assume again that the market is complete. We denote by $\mathcal{A}_w$ the set of all admissible consumption-investment plans that have w as initial wealth and that satisfy the terminal condition $C_T \leq W_T$. We will assume that two utility functions u_1 and u_2, with $u_i \in C^1(0,\infty)$, are involved. The function u_1 describes the utility directly derived from consumption and u_2 the utility derived from terminal wealth as well. Again, we assume that the u_i can be extended to the whole of $\mathbb{R}$ by setting $u_i(x) = -\infty$ for $x < 0$, u right-continuous at $x = 0$ and moreover

$\lim_{x\downarrow 0} u_i'(x) = \infty$. The aim is to maximize for $(C, \bar{\xi}) \in \mathcal{A}_w$ the cumulative expected utility

$$\mathbb{E}[\sum_{t=0}^{T} \alpha^t u_1(C_t) + \alpha^T u_2(W_T - C_T)]. \tag{9.23}$$

There is a variation on this problem conceivable, for instance by replacing the last utility term by $u_2(W_T)$ and/or having in the summation an upper limit equal to $T-1$. We don't consider these possibilities further.

Paralleling the reasoning that led us to Proposition 9.14, we obtain

Proposition 9.17 *Given an initial wealth $w \geq 0$ and an admissible consumption-investment plan $(C, \bar{\xi})$, it holds that*

$$\mathbb{E}^*[\sum_{t=0}^{T-1} \gamma_t + V_T] = w. \tag{9.24}$$

Conversely, given $w \geq 0$ and a consumption process C with $C_T \leq W_T$, there exists a trading strategy $\bar{\xi}$ such that $(C, \bar{\xi})$ is admissible if relation (9.24) holds.

Proof Similar to the proof of Proposition 9.14. □

It follows that we can recast the optimization problem as the maximization of (9.23) subject to the constraints $W_T \in \mathcal{F}_T$, C a (nonnegative) adapted process such that $C_T \leq W_T$ and (9.24). Actually the assumptions that $u_i'(x) \to \infty$ as $x \downarrow 0$ will guarantee that the optimal consumption process C^* is positive a.s. for all t and that the optimal terminal wealth is such that $W_T^* > C_T^*$ a.s. Hence the constraint $C_T \leq W_T$ will be automatically satisfied in the optimum and is therefore redundant.

Recall the definition of the ϕ_t in Lemma 9.15, its 'discounted' analogue $\tilde{\phi}_t = \frac{\phi_t}{S_t^0}$ and Equation (9.22). In the present situation we maximize the Lagrangian

$$L(C_0, \ldots, C_T, W_T, \lambda) =$$
$$\mathbb{E}[\sum_{t=0}^{T} \alpha^t u_1(C_t) + \alpha^T u_2(W_T - C_T) - \lambda(\sum_{t=0}^{T-1} C_t\tilde{\phi}_t + W_T\tilde{\phi}_T - w)].$$

As before, one can write down the first order necessary conditions, by computing partial derivatives. Solving these equations and assuming that I_1 and I_2 are properly defined inverse functions of u_1' and u_2' respectively, we obtain

$$C_t^* = I_1(\alpha^{-t}\lambda\tilde{\phi}_t),\ t = 0, \ldots, T$$
$$W_T^* = I_1(\alpha^{-T}\lambda\tilde{\phi}_T) + I_2(\alpha^{-T}\lambda\tilde{\phi}_T).$$

The optimal value λ^* follows by inserting the optimal solution into Equation (9.24), provided that the resulting equation has a unique solution. One can show that this is for instance the case if $u_1 = u_2$ and of HARA type.

As an aside, we mention that the dynamic programming approach to the problem of maximizing (9.23) results in a recursion formula that coincides with (9.10) for $t < T$. However, there is a crucial difference with the initialization of the dynamic programming algorithm. At time T the final utility is $\alpha^T(u_1(C_T)+u_2(W_T-C_T))$ and so one has to divide terminal wealth into what is kept (for future investments for instance, there is life after T) and what is consumed. Therefore the proper initialization becomes

$$v_T(w) = \max\{u_1(c) + u_2(w-c) : 0 \le c \le w\}.$$

Note that under the additional condition $u_1(x) = u_2(x) = -\infty$ for $x < 0$ the constraint $0 \le c \le w$ is superfluous.

9.4 Exercises

9.1 Prove Lemma 9.1.

9.2 Prove Proposition 9.8.

9.3 Consider the optimization problem at the beginning of Section 9.2. Show (by a heuristic argument) that

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}} = \frac{S_T^0 \tilde{u}'(W_T^*)}{\mathbb{E} S_T^0 \tilde{u}'(W_T)}$$

defines a risk-neutral measure on $\mathcal{F}_T$, analogous to Proposition 6.7. Here W_T^* stands for the wealth resulting from the optimal strategy, and is assumed to be such that u' is defined at it. *Hint: The expected future rewards, see Definition 9.5 are maximized by the optimal strategy.*

9.4 Consider a market with underlying $\Omega = \{1, 2, 3, 4\}$. Assume that $T = 2$ and that $S_t^0 = 1$ for $t = 0, 1, 2$, the price of the riskless asset is constant and equal to one. Let the evolution for the price S_t of the *single* risky asset be as given in the table. We further assume that all singletons have probability $\frac{1}{4}$.

ω	$S_0(\omega)$	$S_1(\omega)$	$S_2(\omega)$
1	5	8	9
2	5	8	6
3	5	4	6
4	5	4	3

The aim is to maximize $\mathbb{E}u(W_T)$, as in Section 9.2, for $u(x) = 1 - \exp(-x)$.

(a) Let $t = 2$. Show that the optimal ξ_2^* satisfies $\xi_2^*(\omega) = -\frac{1}{3}\log 2$, if $\omega \in \{1, 2\}$ and that $\xi_2^*(\omega) = \frac{1}{3}\log 2$, if $\omega \in \{3, 4\}$.

(b) Compute the optimal ξ_1^* (a constant!).

(c) What is the resulting optimal expected utility?

9.5 Show that the recursions (9.6) and (9.7) lead to the same optimum at $t = 0$.

9.6 Consider a CRR model, in which the returns are *iid* with $\mathbb{P}(R_t = b) = p$ (where p is not necessarily equal to the risk neutral value p^*). Consider the maximization of $\mathbb{E}u(W_T)$, with $u(x) = \log x$. Compute the optimal trading strategy ξ^* via dynamic programming. *Hint*: show that $\tilde{v}_t(x) = \log x + k_t$ for certain *constants* k_t.

9.7 Consider a CRR model as in Exercise 9.6, so with a parameter p that determines the probability measure $\mathbb{P}$. Let p^* be as in Proposition 8.35. Assume that $u(x) = \log x$ and that the initial capital is w.

(a) Show that the optimal attainable terminal wealth is given by

$$W_T^* = w(1+r)^T \big(\frac{p}{p^*}\big)^{B_T} \big(\frac{1-p}{1-p^*}\big)^{T-B_T},$$

where B_T is the number of 'up-movements' of the stock.

(b) Assume that at time $T-1$ and that B_{T-1} 'up-movements' have been observed. Show, using the risk-neutral approach, that for the optimal replicating strategy one has

$$\xi_T^1 = w(1+r)^T \big(\frac{p}{p^*}\big)^{B_{T-1}} \big(\frac{1-p}{1-p^*}\big)^{T-1-B_{T-1}} \frac{p-p^*}{S_{T-1}(b-a)p^*(1-p^*)}$$

$$\xi_T^0 = w\big(\frac{p}{p^*}\big)^{B_{T-1}} \big(\frac{1-p}{1-p^*}\big)^{T-1-B_{T-1}} \big(\frac{p^* - p + bp^*(1-p) - ap(1-p^*)}{(b-a)p^*(1-p^*)}\big).$$

(c) Show that $W_{T-1}^* = w(1+r)^{T-1}\big(\frac{p}{p^*}\big)^{B_{T-1}}\big(\frac{1-p}{1-p^*}\big)^{T-1-B_{T-1}}$ and that the fraction of the wealth W_{T-1}^* that is invested in the risky asset is equal to

$$\frac{(1+r)(p-p^*)}{(b-a)p^*(1-p^*)}.$$

(d) Conjecture what the fraction of the capital W_t^* is, that is invested in the risky asset at $t < T-1$.

9.8 Suppose that at each time $t \in \{0, \ldots, T\}$ an investor has a certain capital W_t at her disposal. She consumes part of this, $C_t \geq 0$ say, and invests the remaining $W_t - C_t \geq 0$. The latter she does partly, a deterministic fraction π_t, in a riskless asset with fixed return $1 + r =: R$ and the remaining money in a risky asset with random yield S_t. This leads to the evolution

$$W_{t+1} = (W_t - C_t)\left(\pi_t R + (1-\pi_t)S_t\right), \, t = 0, \ldots, T-1.$$

It is assumed that the S_t are *i.i.d.* random variables, all having the same distribution as a random variable S, and that all relevant expectations exist and are finite. Consider the utility function $u(x) = \frac{1}{\gamma}x^\gamma$, $x \geq 0$ and $\gamma < 1$, $\gamma \neq 0$. Let $\rho \in (0,1)$ be a discount factor. The aim is to maximize $\sum_{t=0}^T \rho^t \, \mathbb{E}u(C_t)$ by appropriately selecting the π_t, $t = 0, \ldots, T-1$, and the consumption C_t, $t = 0, \ldots, T$. Assume $C_T = W_T$. The purpose is to characterize the optimal consumption pattern and to derive it by dynamic programming. We need more notation. We denote by π^* the solution, assumed to exist, of the equation $\mathbb{E}(\pi R + (1-\pi)S)^{\gamma-1}(R-S) = 0$, and $\xi = \rho\mathbb{E}\big(\pi^* R + (1-\pi^*)S\big)^\gamma$.

(a) Look at the theory of dynamic programming and rewrite Algorithm 9.7 in terms of the variables of this exercise.

(b) Compute the optimal consumption C^*_{T-1} at time $T-1$, show that $C^*_{T-1} = \alpha_{T-1}W_{T-1}$, with $\alpha_{T-1} = \frac{(\beta_T\xi)^{1/(\gamma-1)}}{1+(\beta_T\xi)^{1/(\gamma-1)}}$ for $\beta_T = 1$.

(c) Show that the optimal value function at time $T-1$ is given by $v_{T-1}(w) = \rho^{T-1}\beta_{T-1}\frac{w^\gamma}{\gamma}$, where $\beta_{T-1} = \alpha^\gamma_{T-1} + \beta_T(1-\alpha_{T-1})^\gamma\xi$.

(d) Show that the optimal π^*_t are the same for all $t \leq T-1$ and that the optimal consumption is given by $C^*_t = \alpha_t W_t$, where the (nonrandom) constants $\alpha_t \in (0,1)$ are given by $\alpha_{t-1} = \frac{(\beta_t\xi)^{1/(\gamma-1)}}{1+(\beta_t\xi)^{1/(\gamma-1)}}$, and (recursively) $\beta_{t-1} = \alpha^\gamma_{t-1} + \beta_t(1-\alpha_{t-1})^\gamma\xi$. In passing you can show that the value functions are $v_t(x) = \rho^t\beta_t\frac{x^\gamma}{\gamma}$, $t = 0,\ldots,T$.

(e) Note that we can write $\alpha_{t-1} = \frac{p_t}{1+p_t}$, with $p_t = (\beta_t\xi)^{1/(\gamma-1)}$. Show by a simple computation the formula $\beta_{t-1} = (\frac{p_t}{1+p_t})^{\gamma-1}$, which equals $\alpha^{\gamma-1}_{t-1}$. Show also the backward recursion $\frac{p_{t-1}}{a} = \frac{p_t}{1+p_t}$, where $a = \xi^{1/(\gamma-1)}$.

(f) Here we do some time reversion, we put $q_k = \frac{1+p_{T-k}}{p_{T-k}} = 1 + \frac{1}{p_{T-k}}$. Show that $q_k = 1 + \frac{q_{k-1}}{a}$, leading with $q_0 = 0$ to $q_k = \sum_{j=0}^k a^{-j} = \frac{a^{k+1}-1}{a^k(a-1)}$, if $a \neq 1$ (and $k+1$ otherwise).

(g) Finally, show that the optimal consumption is given by $C^*_t = \frac{1}{q_{T-t}}W_t$, $t \leq T-1$. What is $v_0(x)$?

(h) Sketch the solution to the optimal consumption problem for the situation of logarithmic utility, $u(x) = \log x$. [For this utility function one has $u'(x) = x^{-1}$, which corresponds to $\gamma = 0$ above.]

[This exercise can be seen as a dynamic version of the situation in Proposition 4.7 and has been derived from the paper Paul A. Samuelson: Portfolio Selection By Dynamic Stochastic Programming, *The Review of Economics and Statistics* **51(3)**, 239–246, 1969.]

9.9 Show that Equation (9.24) has a unique solution λ if $u_1 = u_2$ and of HARA type and when the optimal values for γ_t and V_T are used.

9.10 Verify the validity of Equations (9.16) and (9.17) by the variational arguments of Section 7.1. You are allowed to interchange expectation and differentiation when needed.

9.11 Consider the maximization of (9.20) with the utility function $u(x) = \frac{x^\gamma}{\gamma}$ for $\gamma < 1$, $\gamma \neq 0$. Show that the optimal C^*_t is of the form $c\,\frac{\alpha^{t/(1-\gamma)}}{\tilde{\phi}_t^{1/(1-\gamma}}$ for some c (which one?) and compute the maximal value of the objective function.

A Complementary and background results

A.1 Separating hyperplanes

Theorem A.1 *Let C be a non-empty convex subset in $\mathbb{R}^n$ such that $0 \in \mathbb{R}^n \setminus C$. Then there exists a linear map $T : \mathbb{R}^n \to \mathbb{R}$ such that $Tx \geq 0$ for all x in C and $Tx_0 > 0$ for at least one $x_0 \in C$.*

Proof Assume first that $0 \notin \text{Cl}\, C$. Consider the continuous map $x \to ||x||$ and let B_n be the closed ball around zero with radius n, $n \in \mathbb{N}$. Restricted to $C_n := B_n \cap \text{Cl}\, C$, this map attains a minimum d_n at some x_n if the intersection is not empty. Obviously the d_n are decreasing, since $C_n \subset C_{n+1}$. But, if $x \in B_{n+1} \setminus B_n$, then $||x|| > ||x'||$ for all $x' \in B_n$, hence the d_n are all equal to some d. Let then $x_0 \in \text{Cl}\, C$ such that $||x_0|| = d$, note that $x_0 \neq 0$ and $d > 0$. Define $Tx = x_0 \cdot x$. Then $Tx_0 = d^2 > 0$. Let $x \in C$ and let y be the projection on the subspace spanned by x_0. Then $y = \lambda x_0$ and one easily shows, exploiting the convexity of C, that $\lambda \geq 1$, see Exercise A.1. But then $Tx = Ty = \lambda Tx_0 \geq \lambda d^2$. This shows the assertion under the extra assumption $0 \notin \text{Cl}\, C$.

To show the assertion for the general case, we may now assume that $0 \in \partial C$. We show that $\mathbb{R}^n \setminus \text{Cl}\, C \neq \emptyset$. If C is contained in a linear subspace of $\mathbb{R}^n$ with dimension less than n, the assertion is obvious. So we assume that the linear span of C is equal to $\mathbb{R}^n$ and therefore there exists a basis of $\mathbb{R}^n$ consisting of n linear independent vectors v_k in C. Let $z = -\sum_k v_k$ and suppose that $z \in \text{Cl}C$. Then there $z_m \in C$ such that $z_m \to z$ and in particular all their coordinates c_k^m w.r.t. this basis converge to -1. Hence there is an certain index m_0 such that all $c_k^{m_0}$ are negative. Let $\alpha_k = \frac{-c_k^{m_0}}{1-\sum_k c_k^{m_0}}$ for $k = 1, \ldots, n$ and $\alpha_0 = \frac{1}{1-\sum_k c_k^{m_0}}$. Then 0 is the convex combination $0 = \alpha_0 z_{m_0} + \sum_{k=1}^n \alpha_k v_k$ and thus in C, which contradicts the hypothesis. We conclude that $\text{Cl}\, C$ is not all of $\mathbb{R}^n$.

We can now choose a sequence of z_n that all have strictly positive Euclidean distance to C and $z_n \to 0$. Application of the first part of the proof yields the existence of linear functionals T_n on $\mathbb{R}^n$ such that $\inf\{T_n(x - z_n) : x \in C\} > 0$. We can represent T_n by unit vectors η_n, $T_n x = \eta_n \cdot x$. By compactness of the unit sphere, there exists a subsequence (η_{n_k}) converging to some limit vector η, with $||\eta|| = 1$. But then for all $x \in C$ one has

$$\eta \cdot x = \lim \eta_{n_k} \cdot (x - z_{n_k}) \geq 0.$$

If $\eta \cdot x$ were zero for all $x \in C$, then $\eta \cdot x = 0$ for all $x \in \mathbb{R}^n$, since $\mathbb{R}^n$ is the linear span of C. But this cannot happen, because $\eta \neq 0$. Hence there must be $x_0 \in C$ such that $\eta \cdot x_0 > 0$. □

Remark A.2 Notice that the first part of the proof shows that $\inf\{Tx : x \in C\} > 0$ if $x \notin \text{Cl}\, C$.

A.2 Hahn-Banach theorem and some ramifications

Let $\mathcal{X}$ be a (real) vector space with a norm $\|\cdot\|$. Let $T : \mathcal{X} \to \mathbb{R}$ be a linear operator. The operator norm of T is defined by $\{\|T\| := \sup\{|Tx| : \|x\| = 1\}$. We call T bounded if $\|T\| < \infty$. Recall that T is continuous iff it is bounded.

Definition A.3 Let $\mathcal{X}$ be a (real) linear space. Call a function $p : \mathcal{X} \to [0, \infty)$ a quasinorm, if p is sub-additive, $p(x + y) \leq p(x) + p(y)$ for $x, y \in \mathcal{X}$ and homogeneous, $p(tx) = tp(x)$, for $t \geq 0, x \in \mathcal{X}$.

A simple and useful example is given by $p(x) = c||x||$, if $\mathcal{X}$ is endowed with a norm $||\cdot||$ and $c > 0$.

Theorem A.4 (Hahn-Banach) *Let T_0 be a linear operator defined on a linear subspace $\mathcal{Y}$ of a real vector space $\mathcal{X}$ and p a quasinorm on $\mathcal{X}$. Suppose that $|T_0 y| \leq p(y)$ for all $y \in \mathcal{Y}$. Then T_0 admits an extension $T : \mathcal{X} \to \mathbb{R}$ such that $|Tx| \leq p(x)$ for all $x \in \mathcal{X}$. In particular, if T_0 is a bounded linear operator on a subspace of a normed space, then it extends to a bounded linear operator on $\mathcal{X}$ such that $\|T\| = ||T_0||$.*

Proof Assume that $\mathcal{Y} \neq \mathcal{X}$. Then there exists $x_1 \in \mathcal{X} \setminus \mathcal{Y}$. Let $\mathcal{Y}_1$ be the linear hull of $\{x_1\} \cup \mathcal{Y}$. Suppose that T_1 is an extension of T_0 to $\mathcal{Y}_1$. Since every $x \in \mathcal{Y}_1$ can uniquely be written as $x = \alpha x_1 + y$, for some $y \in \mathcal{Y}$ and $\alpha \in \mathbb{R}$, it must hold that $T_1 x = \alpha T_1 x_1 + T_0 y$. Write $\xi_1 := T_1 x_1$, then we have

$$T_1 x = \alpha \xi_1 + T_0 y. \tag{A.1}$$

Conversely, every T_1 defined by (A.1) for some $\xi_1 \in \mathbb{R}$ is a continuation of T_0 to $\mathcal{Y}_1$. We will show that it is possible to choose $\xi_1 \in \mathbb{R}$ such that $|T_1 x| \leq p(x)$ for all $x \in \mathcal{Y}_1$. Thereto it is sufficient to prove that

$$T_1 x \leq p(x), \ \forall x \in \mathcal{Y}_1. \tag{A.2}$$

From the proposed definition of T_1 with $\alpha = \pm 1$, it then follows from (A.2) that one necessarily has

$$\xi_1 \leq p(x_1 + y) - T_0 y, \ \forall y \in \mathcal{Y} \tag{A.3}$$

$$\xi_1 \geq -p(x_1 + y) + T_0 y, \ \forall y \in \mathcal{Y}. \tag{A.4}$$

But these two inequalities are also sufficient for (A.2) to hold. We now show that one can choose ξ_1 such that (A.3) and (A.4) hold true. Let y, y' be arbitrary elements of $\mathcal{Y}$. Then

$$T_0(y) - T_0(y') = T_0(y - y') \leq p(y + x_1 - (y' + x_1)) \leq p(y + x_1) + p(y' + x_1),$$

which yields

$$-p(y' + x_1) - T_0 y' \leq p(y + x_1) - T_0 y.$$

Taking the infimum over y and the supremum over y', we get

$$\sup\{-p(y + x_1) - T_0 y : y \in \mathcal{Y}\} \leq \inf\{p(x_1 + y) - T_0 y : y \in \mathcal{Y}\}.$$

Hence a ξ_1 satisfying (A.3) and (A.4) exists.

We finish the proof by invoking Zorn's lemma. Consider the family of all extensions T of T_0 to linear subspaces of $\mathcal{X}$ that contain $\mathcal{Y}$ and that satisfy $|Tx| \leq p(x)$ for all x in the domain of T. This family can be endowed with the partial ordering defined by $T_1 \preceq T_2$ iff T_2 is an extension of T_1. Then there exists a maximal element, T say, in this family w.r.t. this partial ordering. By the preceding part of the proof, the domain of T is all of $\mathcal{X}$. Indeed, if this were not the case, then we could take $\mathcal{Y}$ in the previous part $\mathcal{Y}$ as the domain of T, whereas we have shown that T then admits an extension to a linear subspace of $\mathcal{Y}$ that strictly contains $\mathcal{Y}$, which contradicts maximality of T. □

In the proof of the next theorem we need the concept of Minkowski functional. Let E be a subset of $\mathcal{X}$. Then one defines $\mu_E(x) = \inf\{t > 0 : t^{-1}x \in E\}$. If E is absorbing (for all $x \in \mathcal{X}$ there exists $t > 0$ such that $tx \in E$), then μ_E is finite.

Lemma A.5 *Let E be an absorbing convex subset of a linear space $\mathcal{X}$. Then μ_E is a quasinorm and $\{\mu_E < 1\} \subset E$. If E is open, then $\{\mu_E < 1\} = E$.*

Proof Exercise. □

Theorem A.6 *Let $\mathcal{Y}$ and $\mathcal{Z}$ be nonempty disjoint convex sets of a real normed linear space $\mathcal{X}$.*

(i) Assume that $\mathcal{Y}$ is open. Then there exists a continuous linear functional T on $\mathcal{X}$ and a number $\gamma \in \mathbb{R}$ such that

$$Tx < \gamma \leq Ty, \forall x \in \mathcal{Y}, y \in \mathcal{Z}.$$

(ii) If $\mathcal{Y}$ is compact and $\mathcal{Z}$ is closed, then

$$\sup\{Tx : x \in \mathcal{Y}\} < \inf\{Tx : x \in \mathcal{Z}\}.$$

Proof (i) Fix $y_0 \in \mathcal{Y}$ and $z\_0 \in \mathcal{Z}$. Put $x_0 = z_0 - y_0$ and $\mathcal{C} = \mathcal{Y} - \mathcal{Z} + x_0$. Then $\mathcal{C}$ is a convex neighbourhood of zero and $x_0 \notin \mathcal{C}$, since $\mathcal{Y}$ and $\mathcal{Z}$ are disjoint. Let p be the Minkowski functional of $\mathcal{C}$. By Lemma A.5 we know that p is a quasinorm and that $p(x_0) \geq 1$. Let $\mathcal{X}_0$ be the 1-dimensional subspace generated by x_0. Define T_0 on $\mathcal{X}_0$ by $T_0(tx_0) = t$. Then T_0 is bounded and linear on $\mathcal{X}_0$ and $T_0(tx_0) \leq tp(x_0) = p(tx_0)$ for $t \geq 0$, whereas for $t < 0$ we have $T_0(tx_0) = t < 0 \leq p(tx_0)$. By Theorem A.4, T_0 can be extended to a linear map T on $\mathcal{X}$ such that $T \leq p$. In particular, for $x \in \mathcal{C}$, we have $Tx \leq p(x) \leq 1$ and $T(-x) = -Tx \geq -1$. Hence $|Tx| \leq 1$ on $\mathcal{C} \cap (-\mathcal{C})$, so T is bounded on a neighbourhood of zero. But then T is continuous.

Let $y \in \mathcal{Y}$ and $c \in \mathcal{Z}$ be arbitrary. Since $Tx_0 = T_0x_0 = 1$, we have $Ty - Tz + 1 = T(y - z + x_0) \leq p(y - z + x_0)$. In view of Lemma A.5, $p(y - z + x_0) < 1$ since $\mathcal{C}$ is open. It follows that $Ty < Tz$ for all $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$. Define $\gamma = \sup\{Ty : y \in \mathcal{Y}\}$. Then $\gamma \leq Tz$ for all $z \in \mathcal{Z}$. But since $\mathcal{Y}$ is open, also $\{Ty : y \in \mathcal{Y}\}$ is open and the supremum is not attained. This proves part (i).

To prove part (ii), we first notice that $d(\mathcal{Y},\mathcal{Z}) > 0$, where d is the metric induced by the norm, since $\mathcal{Y}$ is compact and $\mathcal{Z}$ is closed. Hence, there exists $\delta > 0$ such that also the open δ-neighbourhood $\mathcal{Y}^\delta$ of $\mathcal{Y}$ has disjoint intersection with $\mathcal{Z}$. Application of part (i) yields that $Ty < \gamma \leq Tz$ for all $y \in \mathcal{Y}^\delta$ and $z \in \mathcal{Z}$. But, since $\mathcal{Y}$ is compact, we also have $\sup\{Ty : y \in \mathcal{Y}\} < \gamma$, which proves the second part. □

We apply Theorem A.6 to the case where $\mathcal{X} = L^1(\Omega,\mathcal{F},\mathbb{P})$. Recall that the dual space of a normed space is the linear space of all bounded linear functionals. It is well known that for $p \in [1,\infty)$ the dual space of $L^p(\Omega,\mathcal{F},\mathbb{P})$ is isomorphic to $L^q(\Omega,\mathcal{F},\mathbb{P})$ with $q = \frac{p}{p-1}$ for all $p \geq 1$, but we only need this fact for $p = 1$.

Lemma A.7 *The dual space of $L^1(\Omega,\mathcal{F},\mathbb{P})$ is $L^\infty(\Omega,\mathcal{F},\mathbb{P})$.*

Proof Let $T : L^1(\Omega,\mathcal{F},\mathbb{P}) \to \mathbb{R}$ be a bounded linear operator. Define on $\mathcal{F}$ the map ν by

$$\nu(F) = T(\mathbf{1}_F).$$

Obviously, ν is by linearity of T an additive map and by continuity of T even σ-additive. Indeed, if $F_n \downarrow \emptyset$, then $|\nu(F_n)| \leq ||T||\,\mathbb{P}(F_n) \downarrow 0$. Hence ν is a finite signed measure on $\mathcal{F}$, that is absolute continuous w.r.t. $\mathbb{P}$. It follows from the Radon-Nikodym theorem that there is $Y \in L^1(\Omega,\mathcal{F},\mathbb{P})$ such that

$$\nu(F) = \mathbb{E}[\mathbf{1}_F Y], \quad \forall F \in \mathcal{F}. \tag{A.5}$$

Next we show that Y is a.s. bounded. Let $F = \{Y > c\}$, for some $c > 0$. By continuity of T, we have

$$c\,P(Y > c) \leq \mathbb{E}[\mathbf{1}_F Y] = |T(\mathbf{1}_F)| \leq ||T||\,||\mathbf{1}_F||_1 = ||T||\,\mathbb{P}(Y > c).$$

Hence, if $\mathbb{P}(Y > c) > 0$ it follows that $||T|| \geq c$. Stated otherwise $||T|| \geq \sup\{c > 0 : \mathbb{P}(Y > c) > 0\}$. A similar argument yields $||T|| \geq \sup\{c > 0 : \mathbb{P}(Y < -c) > 0\}$. It follows that $||Y||_\infty \leq ||T|| < \infty$.

We finally show that for every $X \in L^1(\Omega,\mathcal{F},\mathbb{P})$, it holds that $T(X) = \mathbb{E}[XY]$. By the above construction this is true for X of the form $X = \mathbf{1}_F$. Hence also for (nonnegative) simple functions. Let $X \in L^1(\Omega,\mathcal{F},\mathbb{P})$ be arbitrary. Choose a sequence (X_n) of simple functions such that $X_n \to X$ a.s. and $|X_n| \leq |X|$. Then, by dominated convergence, $X_n \to X$ in $L^1(\Omega,\mathcal{F},\mathbb{P})$ as well and, since Y is a.s. bounded, we also have $X_n Y \to XY$ in $L^1(\Omega,\mathcal{F},\mathbb{P})$. But then $T(X) = \mathbb{E}[XY]$. □

Remark A.8 In the proof of Lemma A.7 one actually has that $||Y||_\infty = ||T||$. The analogous version for $p \in (1,\infty)$ can be proven as follows. Start from Equation (A.5). It follows that for every $X \in L^\infty(\Omega,\mathcal{F},\mathbb{P})$ one has $T(X) = \mathbb{E}XY$. For every $n \in \mathbb{N}$, put $X_n = \mathrm{sgn}(Y)|Y|^{q-1}\mathbf{1}_{E_n}$, where $E_n = \{|Y| \leq n\}$. Note that every X_n is bounded, $X_n Y = |Y|^q \mathbf{1}_{E_n}$ and $|X_n|^p = |Y|^q \mathbf{1}_{E_n}$. We obtain

$$\mathbb{E}|Y|^q\mathbf{1}_{E_n} = T(X_n) \leq ||T|| \cdot ||X_n||_p = ||T|| \cdot (\mathbb{E}|Y|^q\mathbf{1}_{E_n})^{1/p},$$

from which it follows that $(\mathbb{E}|Y|^q \mathbf{1}_{E_n})^{1/q} \leq ||T||$ (here it is used that $p > 1$). By letting $n \to \infty$, we obtain $||Y||_q \leq ||T|| < \infty$, so $Y \in L^q(\Omega, \mathcal{F}, \mathbb{P})$.

Finally, for $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ we put $X_n = X\mathbf{1}_{\{|X| \leq n\}}$, so that $X_n \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ and $||X - X_n||_p \to 0$. It follows by Hölder's inequality (used in the fourth step) that

$$\begin{aligned} T(X) &= T(X - X_n) + T(X_n) \\ &= T(X - X_n) + \mathbb{E}X_n Y \\ &= T(X - X_n) + \mathbb{E}(X_n - X)Y + \mathbb{E}XY \\ &\to \mathbb{E}XY. \end{aligned}$$

Moreover, one also easily shows that actually $||T|| = ||Y||_q$.

Corollary A.9 *Let $\mathcal{Y}$ and $\mathcal{Z}$ be nonempty disjoint convex sets in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathcal{Y}$ is compact and $\mathcal{Z}$ is closed, there exists an almost surely bounded random variable B such that $\sup\{\mathbb{E}[YB] : Y \in \mathcal{Y}\} < \inf\{\mathbb{E}[ZB] : Z \in \mathcal{Z}\}$.*

Proof Combine Theorem A.6 and Lemma A.7. □

A.3 Existence of essential supremum

If (f_n) is a sequence of measurable functions on some measurable space $(\Omega, \mathcal{F})$, the function f defined by $f(\omega) = \sup_n f_n(\omega)$ is measurable as well. If instead of a sequence we take an arbitrary collection $\{f_i\}_{i \in I}$, measurability of the supremum is no longer guaranteed. But even if this happens, the pointwise supremum may have undesirable properties. Take for instance $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}, \lambda)$, where λ is Lebesgue measure. Let $f_x = \mathbf{1}_{\{x\}}$, $x \in [0,1]$. Then $f = 1$, whereas each f_x is almost surely zero. Both observations trigger the following definition and theorem.

Theorem A.10 *Let $\{f_i\}_{i \in I}$ be an arbitrary collection of measurable functions on $(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists a measurable function $f : \Omega \to \mathbb{R}$ such that*

(i) $f \geq f_i$ a.s. for all $i \in I$

(ii) If f' is any measurable function with the property that $f' \geq f_i$ a.s. for all $i \in I$, then $f' \geq f$ a.s.

The function f is thus a.s. unique and called the essential supremum *of $\{f_i\}_{i \in I}$, notation: $f = \operatorname{ess\,sup}\{f_i\}_{i \in I}$.*

Proof Without loss of generality we may take the collection $F := \{f_i\}_{i \in I}$ bounded, apply an arctan-transformation if needed. Let F' be a countable sub-collection of F, then $f^{F'} := \sup\{f : f \in F'\}$ is measurable. Let

$$s := \sup\{\mathbb{E}f^{F'} : F' \subset F,\ F' \text{ countable}\}.$$

Choose a sequence of countable $F'_n \subset F$ such that $\mathbb{E}f^{F'_n} \to s$ and let $F'_\infty = \cup_n F'_n$. Then F'_∞ is countable and $\mathbb{E}f^{F'_\infty} = s$. Let $f = f^{F'_\infty}$. If (i) is not true for this f, then there exists $i^* \in I$ such that $\mathbb{P}(f < f_{i^*}) > 0$. Put $F^* = F'_\infty \cup \{f_{i^*}\}$. Then $\mathbb{E}f^{F^*} > \mathbb{E}f = s$, contradicting the definition of s. Hence we conclude that (i) holds. Let now f' be measurable such that $f' \geq f_i$ a.s. for all $i \in I$, then obviously $f' \geq f_i$ for $i \in F'_\infty$ a.s. and hence $f' \geq f$, which proves (ii). □

A.4 Results on the weak topology for measures

Let S be a complete separable metric space with metric d and let $\mathcal{S}$ be its Borel σ-field. Denote by $\mathcal{M}_1(\mathcal{S})$ the set of all probability measures on $\mathcal{S}$. By $\mathcal{D}(\mathcal{S})$ we denote the subset of $\mathcal{M}_1(\mathcal{S})$ that consists of all (finite) convex combinations of Dirac measures. Recall that a subset E of a metric space is *totally bounded*, if for every $\varepsilon > 0$, there exist finitely many open balls of radius ε whose union contains E. Recall also that a subset of a metric space is compact iff it is totally bounded and complete.

Lemma A.11 *Let μ be a probability measure on $(S, \mathcal{S})$. Then μ is tight, i.e. for each $\varepsilon > 0$ there exists a compact set K such that $\mu(K) > 1 - \varepsilon$.*

Proof For each $n \in \mathbb{N}$ there is a countable family of open balls B_{nj}, each with radius $\frac{1}{n}$, that covers S. Let $U_n = \bigcup_{j=1}^n B_{nj}$, then $U_n \uparrow S$ and hence for all $\varepsilon > 0$, there is m_n such that $\mathbb{P}(U_{nm_n}) > 1 - 2^{-n}\varepsilon$. Let $D = \bigcap_{n=1}^\infty U_{nm_n}$, then D is obviously totally bounded, and so is its closure K. But since K, being a closed subset of S, is complete, it is then also compact. Moreover, $\mu(K^c) \leq \mu(D^c) \leq \sum_{n\geq 1} \mu(U^c_{nm_m}) < \varepsilon$. □

Proposition A.12 *Let $\mu \in \mathcal{M}_1(\mathcal{S})$. Then there exists a sequence $(\mu_n) \subset \mathcal{D}(\mathcal{S})$ such that $\mu_n \to \mu$ weakly.*

Proof We first show that the assertion holds, if $\mu(K^c) = 0$ for some compact set K. For every n we can cover K by a finite set of open balls $B_{n,j}$, $j = 1, \ldots, k_n$ each having radius $\frac{1}{n}$. Put $A_{n,1} = B_{n,1} \cap K$ and for $j > 1$, recursively, $A_{n,j} = B_{n,j} \cap K \setminus (\cup_{i=1}^{j-1} A_{n,i})$. Then all $A_{n,j}$ are contained in $B_{n,j}$ and $\cup_{j=1}^{k_n} A_{n,j} = K$. Ignoring the j for which $A_{n,j}$ is empty, we select $x_{n,j} \in A_{n,j}$. Put

$$\mu_n = \sum_j \mu(A_{n,j})\delta_{x_{n,j}},$$

where $\delta_{x_{n,j}}$ is the Dirac measure concentrated at $x_{n,j}$. Note that also μ_n is concentrated on K. Let $f \in C_b(S)$. Since μ is concentrated on K, we have

$$\int f \, \mathrm{d}\mu = \sum_j \int_{A_{n,j}} f \, \mathrm{d}\mu.$$

Since f is uniformly continuous on K, we have that

$$\eta_n := \sup\{|f(x) - f(y)| : d(x, y) < \frac{1}{n}\} \to 0.$$

Hence

$$
\begin{aligned}
|\int_{A_{n,j}} f \,\mathrm{d}\mu - \mu(A_{n,j}) f(x_{n,j})| &= |\int_{A_{n,j}} (f - f(x_{n,j}))\,\mathrm{d}\mu| \\
&\leq \int_{A_{n,j}} |f - f(x_{n,j})|\,\mathrm{d}\mu \\
&\leq \eta_n \mu(A_{n,j}).
\end{aligned}
$$

By summing over j we obtain

$$
|\int f\,\mathrm{d}\mu - \int f\,\mathrm{d}\mu_n| \leq \eta_n \mu(K) = \eta_n,
$$

which yields the result.

For an arbitrary probability measure $\mu \in \mathcal{M}_1(S)$ we argue as follows. Let $\varepsilon > 0$. In view of Lemma A.11, there exists a compact set K such that $\mu(K) > 1 - \varepsilon$. Define the (conditional) probability measure μ' by $\mu'(B) = \mu(B|K)$. Let $f \in C_b(S)$ and let μ_n be the measures as in the first part of the proof (with μ' replacing μ). Then

$$
\begin{aligned}
|\int f\,\mathrm{d}\mu - \int f\,\mathrm{d}\mu_n| &\leq \varepsilon||f|| + |\mu(K)\int_K f\,\mathrm{d}\mu' - \int_K f\,\mathrm{d}\mu_n| \\
&\leq \varepsilon||f|| + |(\mu(K)-1)\int_K f\,\mathrm{d}\mu'| + |\int_K f\,\mathrm{d}\mu' - \int_K f\,\mathrm{d}\mu_n| \\
&\leq 2\varepsilon||f|| + |\int_K f\,\mathrm{d}\mu' - \int_K f\,\mathrm{d}\mu_n|.
\end{aligned}
$$

Since the last term vanishes in view of the first part of the proof, the conclusion of the theorem follows. □

Corollary A.13 *If S_0 is a countable dense subset, then we can choose the $x_{n,j}$ in the proof of Proposition A.12 in S_0. Moreover, we can approximate μ with a convex mixture of Dirac distributions, where the mixing coefficients are rational.*

Proof Obvious. □

A.5 Proofs of results of Section 9.1

This section contains proofs of results in Section 9.1. We shall use the independence lemma 8.36.

Lemma A.14 *Suppose $v_0, \ldots, v_T$ are functions on $\mathbb{R}^d$ satisfying*

$$
\begin{aligned}
&v_T(x) \geq g_T(x), \forall x \in \mathbb{R}^d, \\
&v_t(x) \geq g_t(x,y) + \hat{v}_{t+1}(x,y), \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^m, \forall t \in \{0,\ldots,T-1\}.
\end{aligned}
$$

Then for all sequences u it holds that

$$
v_t(X_t^u) \geq J_t(u)\,\text{a.s.},\ t = 0,\ldots,T, \tag{A.6}
$$

and $\mathbb{E}v_0(X_0) \geq J(u)$.

Proof Let $t = T$. Then $v_T(X_T^u) \geq g_T(X_T^u) = \mathbb{E}[g_T(X_T^u)|\mathcal{F}_T] = J_T(u)$, so (A.6) holds for $t = T$. We proceed by backward induction, so assume that (A.6) holds true at times $t+1, \ldots, T$. Notice that by independence of ε_t and $\mathcal{F}_t$ and part (iv) of Theorem B.34 it holds that

$$\hat{v}_{t+1}(X_t^u, U_t) = \mathbb{E}[v_{t+1}(f_t(X_t^u, U_t, \varepsilon_t))|\mathcal{F}_t] = \mathbb{E}[v_{t+1}(X_{t+1}^u)|\mathcal{F}_t]. \tag{A.7}$$

Then, by the assumption on v_t, (A.7), and the induction assumption,

$$\begin{aligned}
v_t(X_t^u) &\geq g_t(X_t^u, U_t) + \mathbb{E}[v_{t+1}(X_{t+1}^u)|\mathcal{F}_t] \\
&\geq \mathbb{E}[g_t(X_t^u, U_t) + J_{t+1}(u)|\mathcal{F}_t] \\
&= \mathbb{E}[g_t(X_t^u, U_t) + \mathbb{E}[\sum_{s=t+1}^{T-1} g_s(X_s^u, U_s) + g_T(X_T^u)|\mathcal{F}_{t+1}]|\mathcal{F}_t] \\
&= \mathbb{E}[\sum_{s=t}^{T-1} g_s(X_s^u, U_s) + g_T(X_T^u)|\mathcal{F}_t] \\
&= J_t(u).
\end{aligned}$$

This shows (A.6). Applying this inequality for $t = 0$ and taking expectations yields the final assertion. □

Lemma A.15 *Suppose that the U_t are such that $U_t = u_t(X_t)$, for some measurable functions $u_t : \mathbb{R}^d \to \mathbb{R}^m$, and let $\mathcal{M}$ be the class of sequences $u = (u_0, \ldots, u_{T-1})$ of such functions. Suppose that for all $u \in \mathcal{M}$ one defines functions $v_0^u, \ldots, v_T^u$ by*

$$\begin{aligned}
v_T^u(x) &= g_T(x) \\
v_t^u(x) &= g_t(x, u_t(x)) + \mathbb{E}v_{t+1}^u(f_t(x, u_t(x), \varepsilon_t)),\ t = 0, \ldots, T-1.
\end{aligned}$$

Then it holds that X^u is Markov w.r.t. to the given filtration, $v_t(X_t^u) = J_t(u)$ for $t = 0, \ldots, T$ and

$$J_t(u) = \mathbb{E}[\sum_{s=t}^{T-1} g_s(X_s^u, u_s(X_s^u)) + g_T(X_T^u)|X_t^u].$$

Proof Similar to the proof of Lemma A.14 upon noting that the conditional expectations w.r.t. $\mathcal{F}_t$ reduce to conditional expectations w.r.t. X_t^u in view of Lemma 9.1. □

Proof of Theorem 9.6 (i) This assertion directly follows from Lemma A.14.

(ii) Suppose that the supremum in (9.4) is attained at $y = u_t^*(x)$. Then, for all $t \in \{0, \ldots, T-1\}$,

$$v_t(x) = g_t(x, u_t^*(x)) + \hat{v}_{t+1}(x, u_t^*(x)).$$

We now apply Lemma A.15 to obtain $\mathbb{E}v_0(X_0) = \mathbb{E}J_0(u^*) = J(u^*)$. Since for every other $u \in \mathcal{M}$, it holds that $J(u) \leq \mathbb{E}v_0(X_0)$, it follows that u^* is optimal. Likewise one shows that $v_t(X_t^{u^*}) = J_t(u^*)$.

Conversely, assume that $u^* \in \mathcal{M}$ is optimal. Let $t = T-1$ and $X^{u^*}_{T-1}$ be the current state. Suppose that the supremum in (9.4) is not attained for $u^*_{T-1}(X^{u^*}_{T-1})$ with positive probability. From the definition of v_{T-1} it follows that there exists some $\tilde{u}_{T-1}(X^{u^*}_{T-1})$ such that a.s.

$$
\begin{aligned}
&g_{T-1}(X^{u^*}_{T-1}, \tilde{u}_{T-1}(X^{u^*}_{T-1})) + \hat{v}_T(X^{u^*}_{T-1}, \tilde{u}_{T-1}(X^{u^*}_{T-1})) \\
&\geq g_{T-1}(X^{u^*}_{T-1}, u^*_{T-1}(X^{u^*}_{T-1})) + \hat{v}_{t+1}(X^{u^*}_{T-1}, u^*_{T-1}(X^{u^*}_{T-1})),
\end{aligned}
$$

where the inequality is strict with positive probability. After taking expectations, a strict inequality emerges. We claim that $\tilde{u} = (u^*_0, \ldots, u^*_{T-2}, \tilde{u}_{T-1})$ is a sequence that yields a higher performance than u^*. Indeed $X^{\tilde{u}}_t = X^{u^*}_t$ for all $t \leq T-1$ and hence $\mathbb{E}g_t(X^{\tilde{u}}_t, \tilde{u}_t(X^{\tilde{u}}_t)) = \mathbb{E}g_t(X^{u^*}_t, u^*_t(X^{u^*}_t))$ for $t \leq T-2$, whereas

$$
\begin{aligned}
&\mathbb{E}g_{T-1}(X^{\tilde{u}}_{T-1}, \tilde{u}_{T-1}(X^{\tilde{u}}_{T-1})) + \mathbb{E}\hat{v}_T(X^{\tilde{u}}_{T-1}, \tilde{u}_{T-1}(X^{\tilde{u}}_{T-1})) \\
&> \mathbb{E}g_{T-1}(X^{u^*}_{T-1}, u^*_{T-1}(X^{u^*}_t)) + \mathbb{E}\hat{v}_T(X^{u^*}_{T-1}, u^*_{T-1}(X^{u^*}_{T-1})).
\end{aligned}
$$

But then $J(\tilde{u}) > J(u^*)$, which contradicts optimality of u^*. Hence we must have $v_{T-1}(X^{u*}_{T-1}) = J_{T-1}(u^*)$ by virtue of Lemma A.15.

One proceeds by induction. Suppose that the supremum in (9.4) is attained for $u^*_s(X^{u^*}_s)$ with probability one, for all $t+1 \leq s \leq T-1$, but that, with positive probability, this is not true for $s = t$. Then one argues as above that there exists a $\tilde{u}_t$ such that with $\tilde{u} = (u^*_0, \ldots, u^*_{t-1}, \tilde{u}_t, u^*_{t+1}, \ldots, u^*_{T-1})$ it holds that $J(\tilde{u}) > J(u^*)$, which again contradicts optimality of u^*.

In passing, we also obtain that $v_t(X^{u^*}_t) = J_t(u^*)$ and $\sup_u J(u) = J(u^*) = \mathbb{E}v_0(X_0)$. □

A.6 Exercises

A.1 The proof of the first part of Theorem A.1 says "one easily shows that $\lambda \geq 1$". You may have noticed that in the text before that, convexity of C is not used, and this is what is needed to establish $\lambda \geq 1$. Define $x_t = tx + (1-t)x_0$, $t \in [0,1]$. For which t is $\|x_t\|$ minimal? Exploit this to prove the assertion.

B Glossary of basic results in probability theory

This part is meant for readers who are not familiar with or have only marginally been exposed to measure theory. We collect a number of basic results in probability, see for details and an extensive treatment for instance the lecture notes on Measure theoretic probability.

B.1 σ-algebras and measures

Definition B.1 Let S be a non-empty set. A collection $\Sigma \subset 2^S$ is called a *σ-algebra* (on S) if

(i) $S \in \Sigma$,

(ii) $E \in \Sigma \Rightarrow E^c \in \Sigma$,

(iii) $E_1, E_2, \cdots \in \Sigma \Rightarrow \bigcup_{n=1}^\infty E_n \in \Sigma$.

If Σ is a σ-algebra on S, then (S, Σ) is called a *measurable space* and the elements of Σ are called *measurable* sets.

Notice that always $\emptyset$ belongs to a σ-algebra, since $\emptyset = S^c$. One also has $E_1, E_2, \cdots \in \Sigma \Rightarrow \bigcap_{n=1}^\infty E_n \in \Sigma$. Property (iii) is valid for finite unions too, as well as the latter intersection property.

If $\mathcal{C}$ is any collection of subsets of S, then by $\sigma(\mathcal{C})$ we denote the smallest σ-algebra containing $\mathcal{C}$. This means that $\sigma(\mathcal{C})$ is the intersection of all σ-algebras that contain $\mathcal{C}$. If $\Sigma = \sigma(\mathcal{C})$, we say that $\mathcal{C}$ generates Σ. The union of two σ-algebras Σ_1 and Σ_2 on a set S is usually not a σ-algebra. We write $\Sigma_1 \vee \Sigma_2$ for $\sigma(\Sigma_1 \cup \Sigma_2)$.

One of the most relevant σ-algebras is $\mathcal{B} = \mathcal{B}(\mathbb{R})$, the Borel sets of $\mathbb{R}$. Let $\mathcal{O}$ be the collection of all open subsets of $\mathbb{R}$ with respect to the usual topology (in which all intervals (a, b) are open). Then $\mathcal{B} := \sigma(\mathcal{O})$. Of course, one similarly defines the Borel sets of $\mathbb{R}^d$, and in general, for a topological space $(S, \mathcal{O})$, one defines the Borel-sets as $\sigma(\mathcal{O})$.

Let Σ be a σ-algebra on a set S. We consider mappings $\mu : \Sigma \to [0, \infty]$. Note that ∞ is allowed as a possible value. A mapping μ is called *σ-additive* or *countably additive*, if $\mu(\emptyset) = 0$ and if $\mu(\cup_n E_n) = \sum_n \mu(E_n)$ for every sequence (E_n) of disjoint sets of Σ.

Definition B.2 Let (S, Σ) be a measurable space. A countably additive mapping $\mu : \Sigma \to [0, \infty]$ is called a *measure*. The triple (S, Σ, μ) is called a *measure space*.

Some extra terminology follows. A measure is called finite if $\mu(S) < \infty$. It is called σ-finite, if we can write $S = \cup_n S_n$, where the S_n are measurable sets and $\mu(S_n) < \infty$. If $\mu(S) = 1$, then μ is called a *probability measure*. Here are the first elementary properties of a measure.

Proposition B.3 *Let (S, Σ, μ) be a measure space. Then the following hold true (all the sets below belong to Σ).*

(i) If $E \subset F$, then $\mu(E) \leq \mu(F)$.

(ii) $\mu(E \cup F) \leq \mu(E) + \mu(F)$.

(iii) $\mu(\cup_{k=1}^n E_k) \leq \sum_{k=1}^n \mu(E_k)$

If μ is finite, we also have

(iv) If $E \subset F$, then $\mu(F \setminus E) = \mu(F) - \mu(E)$.

(v) $\mu(E \cup F) = \mu(E) + \mu(F) - \mu(E \cap F)$.

Measures have certain continuity properties.

Proposition B.4 *Let (E_n) be a sequence in Σ.*

(i) If the sequence is increasing, with limit $E = \cup_n E_n$, then $\mu(E_n) \uparrow \mu(E)$ as $n \to \infty$.

(ii) If the sequence is decreasing, with limit $E = \cap_n E_n$ and if $\mu(E_n) < \infty$ from a certain index on, then $\mu(E_n) \downarrow \mu(E)$ as $n \to \infty$.

Corollary B.5 *Let (S, Σ, μ) be a measure space. For an arbitrary sequence (E_n) of sets in Σ, we have $\mu(\cup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty \mu(E_n)$.*

Consider a measure space (S, Σ, μ) and let $E \in \Sigma$ be such that $\mu(E) = 0$. If N is a subset of E, then it is fair to suppose that also $\mu(N) = 0$. But this can only be guaranteed if $N \in \Sigma$. Therefore we introduce some new terminology. A set $N \in S$ is called a *null set* or μ-null set, if there exists $E \in \Sigma$ with $E \supset N$ and $\mu(E) = 0$. The collection of null sets is denoted by $\mathcal{N}$, or $\mathcal{N}_\mu$ since it depends on μ. One can show that $\mathcal{N}$ is a σ-algebra and one can extend μ to $\bar{\Sigma} = \Sigma \vee \mathcal{N}$. If the extension is called $\bar{\mu}$, then we have a new measure space $(S, \bar{\Sigma}, \bar{\mu})$, which is *complete*, all $\bar{\mu}$-null sets belong to the σ-algebra $\bar{\Sigma}$.

Let (S, Σ) be a measurable space. Recall that the elements of Σ are called measurable sets. Also recall that $\mathcal{B} = \mathcal{B}(\mathbb{R})$ is the collection of all the Borel sets of $\mathbb{R}$.

Definition B.6 A mapping $h : S \to \mathbb{R}$ is called *measurable* if $h^{-1}[B] \in \Sigma$ for all $B \in \mathcal{B}$.

It is clear that this definition depends on $\mathcal{B}$ and Σ. When there are more σ-algebras in the picture, we sometimes speak of Σ-measurable functions, or $\Sigma/\mathcal{B}$-measurable functions, depending on the situation. If S is a topological space with a topology $\mathcal{T}$ and if $\Sigma = \sigma(\mathcal{T})$, a measurable function h is also called a *Borel* measurable function. We will often use the shorthand notation $\{h \in B\}$ for the set $\{s \in S : h(s) \in B\}$. Likewise we also write $\{h \leq c\}$ for the set $\{s \in S : h(s) \leq c\}$.

Proposition B.7 *Let (S, Σ) be a measurable space and $h : S \to \mathbb{R}$.*

(i) If $\mathcal{C}$ is a collection of subsets of $\mathbb{R}$ such that $\sigma(\mathcal{C}) = \mathcal{B}$, and if $h^{-1}[C] \in \Sigma$ for all $C \in \mathcal{C}$, then h is measurable.

(ii) If $\{h \leq c\} \in \Sigma$ for all $c \in \mathbb{R}$, then h is measurable.

(iii) If S is topological and h continuous, then h is measurable with respect to the σ-algebra generated by the open sets. In particular any constant function is measurable.

(iv) If h is measurable and another function $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable ($\mathcal{B}/\mathcal{B}$-measurable), then $f \circ h$ is measurable as well.

The set of measurable functions will also be denoted by Σ. This notation is of course a bit ambiguous, but it turns that no confusion can arise. Later on we often need the set of nonnegative measurable functions, denoted Σ^+.

Proposition B.8 *We have the following properties.*

(i) The collection Σ *of* Σ*-measurable functions is a vector space and products of measurable functions are measurable as well.*

(ii) Let (h_n) *be a sequence in* Σ*. Then also* $\inf h_n, \sup h_n, \liminf h_n, \limsup h_n$ *are in* Σ*, where we extend the range of these functions to* $[-\infty, \infty]$*. The set* L*, consisting of all* $s \in S$ *for which* $\lim_n h_n(s)$ *exists as a finite limit, is measurable.*

We turn to a probabilistic setting and change notation. So we consider a set Ω, to be interpreted as outcomes of some experiment (instead of S) and $\mathcal{F}$ (instead of Σ) a σ-algebra defined on it, called events. With a probability measure, denoted $\mathbb{P}$, defined on $\mathcal{F}$, one has a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. In this setting Definition B.6 takes the following form.

Definition B.9 A function $X : \Omega \to \mathbb{R}$ is called a *random variable* if it is ($\mathcal{F}$-)measurable.

Following the tradition, we denote random variables by X (or other capital letters), rather than by h, as just above. By definition, random variables are thus nothing else but measurable functions with respect to a given σ-algebra $\mathcal{F}$. Given $X : \Omega \to \mathbb{R}$, let $\sigma(X) = \{X^{-1}[B] : B \in \mathcal{B}\}$. Then $\sigma(X)$ is a σ-algebra, and X is a random variable in the sense of Definition B.9 iff $\sigma(X) \subset \mathcal{F}$. It follows that $\sigma(X)$ is the smallest σ-algebra on Ω such that X is a random variable. Sometimes we need the following result.

Proposition B.10 *Let* $X : \Omega \to \mathbb{R}$*. If* $Y : \Omega \to \mathbb{R}$ *is* $\sigma(X)$*-measurable, there exists a Borel-measurable function* $f : \mathbb{R} \to \mathbb{R}$ *such that* $Y = f \circ X$*. This result is, mutatis mutandis, also true when* X *and* Y *take values in higher dimensional Euclidean spaces.*

B.2 Lebesgue integrals and expectation

Definition B.11 A function $f : S \to [0, \infty)$ is called a nonnegative simple function, if it has a representation as a finite sum

$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}, \tag{B.8}$$

where $a_i \in [0, \infty)$ and $A_i \in \Sigma$. The class of all nonnegative simple functions is denoted by $\mathfrak{S}^+$.

Definition B.12 Let $f \in \mathfrak{S}^+$. The (*Lebesgue*) *integral* of f with respect to the measure μ is defined as

$$\int f \, \mathrm{d}\mu := \sum_{i=1}^{n} a_i \mu(A_i), \tag{B.9}$$

when f has representation (B.8).

We say that a property of elements of S holds almost everywhere (usually abbreviated by a.e. or by μ-a.e.), if the set for which this property does not hold, has measure zero. For instance, we say that two measurable functions are almost everywhere equal, if $\mu(\{f \neq g\}) = 0$. We continue with a definition, in which we use that we already know how to integrate simple functions.

Definition B.13 Let f be a nonnegative measurable function. The integral of f is defined as $\mu(f) := \sup\{\mu(h) : h \leq f, h \in \mathfrak{S}^+\}$, where $\mu(h)$ is as in Definition B.12.

Proposition B.14 *Let $f, g \in \Sigma^+$. If $f \leq g$ a.e., then $\mu(f) \leq \mu(g)$, and if $f = g$ a.e., then $\mu(f) = \mu(g)$.*

Lemma B.15 *Let $f \in \Sigma^+$ and suppose that $\mu(f) = 0$. Then $f = 0$ a.e.*

The first important limit theorem is the *Monotone Convergence Theorem.*

Theorem B.16 *Let (f_n) be a sequence in Σ^+, such that $f_{n+1} \geq f_n$ a.e. for each n. Let $f = \limsup f_n$. Then $\mu(f_n) \uparrow \mu(f) \leq \infty$.*

The next limit result is known as *Fatou's lemma.*

Lemma B.17 *Let (f_n) be an arbitrary sequence in Σ^+. Then $\liminf \mu(f_n) \geq \mu(\liminf f_n)$. If there exists a function $h \in \Sigma^+$ such that $f_n \leq h$ a.e., and $\mu(h) < \infty$, then $\limsup \mu(f_n) \leq \mu(\limsup f_n)$.*

We now extend the notion of integral to (almost) arbitrary measurable functions. Let $f \in \Sigma$. For (extended) real numbers x one defines $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$. Then, for $f : S \to [-\infty, \infty]$, one defines the functions f^+ and f^- by $f^+(s) = f(s)^+$ and $f^-(s) = f(s)^-$. Notice that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. If $f \in \Sigma$, then $f^+, f^- \in \Sigma^+$.

Definition B.18 Let $f \in \Sigma$ and assume that $\mu(f^+) < \infty$ or $\mu(f^-) < \infty$. Then we define $\mu(f) := \mu(f^+) - \mu(f^-)$. If both $\mu(f^+) < \infty$ and $\mu(f^-) < \infty$, we say that f is *integrable*. The collection of all integrable functions is denoted by $\mathcal{L}^1(S, \Sigma, \mu)$. Note that $f \in \mathcal{L}^1(S, \Sigma, \mu)$ implies that $|f| < \infty$ μ-a.e.

Proposition B.19 *Let $f, g \in \mathcal{L}^1(S, \Sigma, \mu)$.*

(i) If $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g \in \mathcal{L}^1(S, \Sigma, \mu)$ and $\mu(\alpha f + \beta g) = \alpha\mu(f) + \beta\mu(g)$. Hence μ can be seen as a linear operator on $\mathcal{L}^1(S, \Sigma, \mu)$.

(ii) If $f, g \in \mathcal{L}^1(S, \Sigma, \mu)$ and $f \leq g$ a.e., then $\mu(f) \leq \mu(g)$.

The next theorem is known as the Dominated Convergence Theorem, also called Lebesgue's Convergence Theorem.

Theorem B.20 *Let $(f_n) \subset \Sigma$ and $f \in \Sigma$. Assume that $f_n(s) \to f(s)$ for all s outside a set of measure zero. Assume also that there exists a function $g \in \Sigma^+$ such that $\sup_n |f_n| \leq g$ a.e. and that $\mu(g) < \infty$. Then $\mu(|f_n - f|) \to 0$, and hence $\mu(f_n) \to \mu(f)$.*

Consider again the probabilistic setting. The important concept to understand is that the expectation of a random variable is a Lebesgue integral. Indeed, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let X be a (real) random variable defined on it. Recall that $X : \Omega \to \mathbb{R}$ is by definition a measurable function. Making the switch between the notations (S, Σ, μ) and $(\Omega, \mathcal{F}, \mathbb{P})$, one has the following notation for the integral of X w.r.t. $\mathbb{P}$

$$\mathbb{P}(X) = \int_\Omega X \, \mathrm{d}\mathbb{P}, \tag{B.10}$$

provided that the integral is well defined. Other often used notations for this integral are $\mathbb{P}X$ and $\mathbb{E}X$. The latter is the favorite one among probabilists and one speaks of the Expectation of X. Note also that $\mathbb{E}X$ is always defined when $X \geq 0$ *almost surely*. The latter concept meaning almost everywhere w.r.t. the probability measure $\mathbb{P}$. We abbreviate almost surely by a.s.

Remark B.21 It now follows that all results for integrals of functions w.r.t. general measures are also valid for expectations.

For instance, the Monotone convergence theorem B.16 in this setting reads as follows. If (X_n) is a sequence in $\mathcal{F}^+$ such that $X_{n+1} \geq X_n$ a.s. for each n, and $X = \limsup X_n$, then $\mathbb{E}X_n \uparrow \mathbb{E}X \leq \infty$.

The Dominated convergence theorem B.20 takes the form $\mathbb{E}|X_n - X| \to 0$ and $\mathbb{E}X_n \to \mathbb{E}X$, once $X_n(\omega) \to X(\omega)$ for all ω outside a set of zero probability (also written as $X_n \overset{\text{a.s.}}{\to} X$, see Definition B.27) and there is $Y \geq 0$ s.t. $|X_n| \leq Y$ and $\mathbb{E}Y < \infty$. If it happens that all X_n are a.s. bounded by a constant $c > 0$, we can take $Y = c$.

The following proposition is known as Jensen's inequality.

Proposition B.22 *Let $g : G \to \mathbb{R}$ be convex and X a random variable with $\mathbb{P}(X \in G) = 1$. Assume that $\mathbb{E}|X| < \infty$ and $\mathbb{E}|g(X)| < \infty$. Then*

$$\mathbb{E}g(X) \geq g(\mathbb{E}X).$$

Definition B.23 Let $1 \leq p < \infty$ and X a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathbb{E}|X|^p < \infty$, we write $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ and $||X||_p = (\mathbb{E}|X|^p)^{1/p}$. One also has $||X||_\infty = \inf\{C \geq 0 : \mathbb{P}(|X| \leq C) = 1\}$. Shorthand notation $\mathcal{L}^p$ is often used.

In fact $||\cdot||_p$ $(1 \leq p \leq \infty)$ is a norm, provided that one identifies random variables that are a.s. equal. This identification can be seen as an equivalence relation, $X \sim Y$ then means $\mathbb{P}(X = Y) = 1$, and the set of equivalence classes of $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ is denoted $L^p(\Omega, \mathcal{F}, \mathbb{P})$, truly a complete space.

The following is the famous Cauchy-Schwarz inequality.

Proposition B.24 *Let $X, Y \in \mathcal{L}^2$. Then $\mathbb{E}|XY| \leq ||X||_2 \, ||Y||_2$.*

Proposition B.25 *Let $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ be independent random variables. Then $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$.*

Theorem B.26 *Let $p \in [1, \infty]$. The space $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ is* complete *in the following sense. Let (X_n) be a Cauchy-sequence in $\mathcal{L}^p$: $||X_n - X_m||_p \to 0$ for $n, m \to \infty$. Then there exists* a limit *$X \in \mathcal{L}^p$ such that $||X_n - X||_p \to 0$. The limit is unique in the sense that any other limit X' satisfies $||X - X'||_p = 0$.*

B.3 Convergence of random variables

Let $X, X_1, X_2, \ldots$ be random variables. We have the following definitions of different modes of convergence. We will always assume that the parameter n tends to infinity, unless stated otherwise.

Definition B.27 Here a three fundamental convergence concepts.

(i) If $\mathbb{P}(\omega : X_n(\omega) \to X(\omega)) = 1$, then we say that X_n converges to X almost surely (a.s.).

(ii) If $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ for all $\varepsilon > 0$, then we say that X_n converges to X in probability.

(iii) If $\mathbb{E}|X_n - X|^p \to 0$ (equivalently, $||X_n - X||_p \to 0$) for some $p \geq 1$, then we say that X_n converges to X in p-th mean, or in $\mathcal{L}^p$.

For these types of convergence we use the following notations: $X_n \stackrel{\text{a.s.}}{\to} X$, $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathcal{L}^p}{\to} X$ respectively.

The following relations hold between the types of convergence introduced in Definition B.27.

Proposition B.28 *The following implications hold.*

(i) If $X_n \stackrel{\text{a.s.}}{\to} X$, then $X_n \stackrel{\mathbb{P}}{\to} X$.

(ii) If for all $\varepsilon > 0$ the series $\sum_n \mathbb{P}(|X_n - X| > \varepsilon)$ is convergent, then $X_n \stackrel{a.s.}{\to} X$.

(iii) If $X_n \stackrel{\mathcal{L}^p}{\to} X$, then $X_n \stackrel{\mathbb{P}}{\to} X$.

(iv) If $p > q > 0$ and $X_n \stackrel{\mathcal{L}^p}{\to} X$, then $X_n \stackrel{\mathcal{L}^q}{\to} X$.

Proposition B.29 *There is equivalence between*

(i) $X_n \stackrel{\mathbb{P}}{\to} X$ and

(ii) every subsequence of (X_n) contains a further subsequence that is almost surely convergent to X.

It follows that $X_n \stackrel{\mathbb{P}}{\to} X$ implies the existence of a subsequence (X_{n_k}) such that $X_{n_k} \stackrel{\text{a.s.}}{\to} X$.

Proposition B.30 *Let $X, X_1, X_2, \ldots$ be random variables and $g : \mathbb{R} \to \mathbb{R}$ be continuous. If $X_n \stackrel{a.s.}{\to} X$, we also have $g(X_n) \stackrel{a.s.}{\to} g(X)$ and if $X_n \stackrel{\mathbb{P}}{\to} X$, then also $g(X_n) \stackrel{\mathbb{P}}{\to} g(X)$.*

B.4 Conditional expectation and martingales

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.

Definition B.31 Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. A random variable $\hat{X}$ is called a *version* of the conditional expectation $\mathbb{E}[X|\mathcal{G}]$, if it is $\mathcal{G}$-measurable and if

$$\mathbb{E}\hat{X}\mathbf{1}_G = \mathbb{E}X\mathbf{1}_G, \forall G \in \mathcal{G}. \tag{B.11}$$

Versions are a.s. unique, i.e. if X' is another $\mathcal{G}$-measurable random variable satisfying (B.11), then $\mathbb{P}(\hat{X} = X') = 1$. Usually we don't bother too much about possibly different version, and simply write $\mathbb{E}[X|\mathcal{G}]$ for $\hat{X}$, although formally $\mathbb{E}[X|\mathcal{G}]$ is an equivalence class of random variables.

If $\mathcal{G} = \sigma(Y)$, where Y is a random variable, then one usually writes $\mathbb{E}[X|Y]$ instead of $\mathbb{E}[X|\sigma(Y)]$. In this situation, Proposition B.10 guarantees the existence of a Borel-measurable f such that a given version of $\mathbb{E}[X|Y]$ is of the form $f \circ Y$.

The existence of a random variable $\hat{X}$ satisfying the requirements of Definition B.31 can for instance be shown as a consequence of the Radon-Nikodym theorem B.38.

Remark B.32 One can also define versions of conditional expectations for random variables X with $\mathbb{P}(X \in [0, \infty]) = 1$ without requiring that $\mathbb{E}X < \infty$. Again this follows from the Radon-Nikodym theorem. The definition of conditional expectation can also be extended to e.g. the case where $X = X^+ - X^-$, where $\mathbb{E}X^- < \infty$, but $\mathbb{E}X^+ = \infty$.

Proposition B.33 *The following elementary properties hold.*

(i) *If $X \geq 0$ a.s., then $\hat{X} \geq 0$ a.s. If $X \geq Y$ a.s., then $\hat{X} \geq \hat{Y}$ a.s.*

(ii) $\mathbb{E}\hat{X} = \mathbb{E}X$.

(iii) *If $a, b \in \mathbb{R}$ and if $\hat{X}$ and $\hat{Y}$ are versions of $\mathbb{E}[X|\mathcal{G}]$ and $\mathbb{E}[Y|\mathcal{G}]$, then $a\hat{X} + b\hat{Y}$ is a version of $\mathbb{E}[aX + bY|\mathcal{G}]$.*

(iv) *If X is $\mathcal{G}$-measurable, then X is a version of $\mathbb{E}[X|\mathcal{G}]$.*

Theorem B.34 *Additional properties of conditional expectations are as follows.*

(i) *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then any version of $\mathbb{E}[\hat{X}|\mathcal{H}]$ is also a version of $\mathbb{E}[X|\mathcal{H}]$ and vice versa (tower property). In short, we write $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.*

(ii) *If Z is $\mathcal{G}$-measurable such that $ZX \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, then $Z\hat{X}$ is a version of $\mathbb{E}[ZX|\mathcal{G}]$. We write $Z\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[ZX|\mathcal{G}]$.*

(iii) *Let $\hat{X}$ be a version of $\mathbb{E}[X|\mathcal{G}]$. If $\mathcal{H}$ is independent of $\sigma(X) \vee \mathcal{G}$, then $\hat{X}$ is a version of $\mathbb{E}[X|\mathcal{G} \vee \mathcal{H}]$. In short, $\mathbb{E}[X|\mathcal{G} \vee \mathcal{H}] = \mathbb{E}[X|\mathcal{G}]$. In particular, $\mathbb{E}X$ is a version of $\mathbb{E}[X|\mathcal{H}]$ if $\sigma(X)$ and $\mathcal{H}$ are independent, in this case $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}X$.*

(iv) *Let X be a $\mathcal{G}$-measurable random variable and let the random variable Y be independent of $\mathcal{G}$. Assume that $h \in \mathcal{B}(\mathbb{R}^2)$ is such that $h(X,Y) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Put $\hat{h}(x) = \mathbb{E}[h(x,Y)]$. Then $\hat{h}$ is a Borel function and $\hat{h}(X)$ is a version of $\mathbb{E}[h(X,Y)|\mathcal{G}]$.*

(v) *If $c : \mathbb{R} \to \mathbb{R}$ is a convex function and $\mathbb{E}|c(X)| < \infty$, then $c(\hat{X}) \leq C$, a.s., where C is any version of $\mathbb{E}[c(X)|\mathcal{G}]$. We often write $c(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[c(X)|\mathcal{G}]$ (Jensen's inequality for conditional expectations).*

(vi) $||\hat{X}||_p \leq ||X||_p$, *for every* $p \geq 1$.

A *stochastic process*, or simply a *process*, (in discrete time) is nothing else but a sequence of random variables defined on some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The time set is often taken as $\{0, 1, 2, \ldots\}$ in which case we have e.g. a sequence of random variables $X_0, X_1, X_2, \ldots$. Such a sequence as a whole is often denoted by X. So we have $X = (X_n)_{n\geq 0}$. Unless otherwise stated, all process below have their values in $\mathbb{R}$, while the extension to $\mathbb{R}^d$-valued processes should be clear.

We shall need a sequence of sub-σ-algebras of $\mathcal{F}$, that form a *filtration* $\mathbb{F}$. This means that $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$, where each $\mathcal{F}_n$ is a σ-algebra satisfying $\mathcal{F}_n \subset \mathcal{F}$, and moreover $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, for all $n \geq 0$. The sequence $\mathbb{F}$ is thus increasing. Recall that in general a union of σ-algebras is not a σ-algebra itself. We define $\mathcal{F}_\infty := \sigma(\cup_{n=0}^\infty \mathcal{F}_n)$. Obviously $\mathcal{F}_n \subset \mathcal{F}_\infty$ for all n. If X is a stochastic process, then one defines $\mathcal{F}_n^X := \sigma(X_0, \ldots, X_n)$. It is clear that $\mathbb{F}^X := (\mathcal{F}_n^X)_{n\geq 0}$ is a filtration.

Given a filtration $\mathbb{F}$, we shall often consider $\mathbb{F}$-adapted processes. A process Y is $\mathbb{F}$-adapted (or adapted to $\mathbb{F}$, or just adapted), if for all n the random variable Y_n is $\mathcal{F}_n$-measurable ($Y_n \in \mathcal{F}_n$). If $\mathbb{F} = \mathbb{F}^X$ for some process X, then another process Y is $\mathbb{F}^X$-adapted, iff for all n, there exists, in view of Proposition B.10, a Borel function $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ such that $Y_n = f_n(X_0, \ldots, X_n)$. Obviously X is adapted to $\mathbb{F}^X$.

A filtration can be interpreted as an information flow, where each $\mathcal{F}_n$ represents the available information up to time n. For $\mathbb{F} = \mathbb{F}^X$, the information comes from the process X and the information at time n is presented by events in terms of $X_0, \ldots, X_n$.

Given a filtration $\mathbb{F}$, a process $Y = (Y_n)_{n\geq 1}$ is called $\mathbb{F}$-predictable (or just predictable) if $Y_n \in \mathcal{F}_{n-1}$, $n \geq 1$. A predictable process Y may be interpreted as a *strategy*, the action at time n depends on the information available at time $n-1$. In a trivial sense, one can 'perfectly' predict Y_n at time $n-1$.

Definition B.35 A stochastic process $M = (M_n)_{n\geq 0}$ is called a *martingale* (or $\mathbb{F}$-martingale), if it is adapted to a filtration $\mathbb{F}$, if $M_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ for all $n \geq 0$ and if

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n \text{ a.s.} \tag{B.12}$$

Equation (B.12), valid for all $n \geq 0$ is called the *martingale property* of M.

The definition of martingales has been given in terms of 'one-step-ahead' conditional expectations. If we change (B.12) in the sense that we replace on the left hand side $\mathbb{E}[M_{n+1}|\mathcal{F}_n]$ with $\mathbb{E}[M_m|\mathcal{F}_n]$, $m \geq n+1$ arbitrary, we obtain an equivalent definition.

If X is any process, we define the process ΔX by

$$\Delta X_n = X_n - X_{n-1},\ n \geq 1.$$

It trivially follows that $X_n = X_0 + \sum_{k=1}^n \Delta X_k$. Sometimes it is convenient to adopt the convention $\Delta X_0 = X_0$, from which we then obtain $X_n = \sum_{k=0}^n \Delta X_k$. The martingale property of an adapted integrable process X can then be formulated as $\mathbb{E}[\Delta X_{n+1}|\mathcal{F}_n] = 0$ a.s. for $n \geq 0$. An equivalent formulation is $\mathbb{E}[\Delta X_{n+1}\mathbf{1}_F] = 0$ for all $F \in \mathcal{F}_n$ for $n \geq 0$.

Let $S_n = \sum_{k=1}^n Y_k \Delta X_k$, $n \geq 1$, $S_0 = 0$. When X is a martingale, one speaks of a *martingale transform*, the terminology is justified is justified by the following proposition.

Proposition B.36 *Let X be an adapted process and Y a predictable process. Assume that the X_n are in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ as well as the $Y_n \Delta X_n$. Then S is a martingale if X is a martingale.*

B.5 Change of measure

The following can also be presented in a much more general way for measures on a measurable space (S, Σ), but we only need the probabilistic setting.

Definition B.37 Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on a measurable space $(\Omega, \mathcal{F})$. We say that $\mathbb{Q}$ is *absolutely continuous* w.r.t. $\mathbb{P}$ (notation $\mathbb{Q} \ll \mathbb{P}$), if $\mathbb{Q}(E) = 0$ for every $E \in \mathcal{F}$ with $\mathbb{P}(E) = 0$. The measures $\mathbb{P}$ and $\mathbb{Q}$ are called *equivalent* (notation $\mathbb{P} \sim \mathbb{Q}$) if $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$.

The next theorem is known as the Radon-Nikodym theorem for absolutely continuous probability measures.

Theorem B.38 *Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on $\mathcal{F}$. There is equivalence between $\mathbb{Q} \ll \mathbb{P}$ and the existence of a random variable $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{Q}(E) = \mathbb{E}(\mathbf{1}_E Z)$ for all $E \in \mathcal{F}$. Moreover, in this case Z is unique in the sense that any other Z' with this property is such that $\mathbb{P}(Z = Z') = 1$. The random variable Z is called the Radon-Nikodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ and is often written as $Z = \frac{d\mathbb{Q}}{d\mathbb{P}}$.*

To distinguish between expectations w.r.t. $\mathbb{P}$ and $\mathbb{Q}$, these are often denoted by $\mathbb{E}_{\mathbb{P}}$ (instead of $\mathbb{E}$) and $\mathbb{E}_{\mathbb{Q}}$ respectively. Note that for $\mathbb{Q} \ll \mathbb{P}$ one has $\mathbb{E}_{\mathbb{P}} Z = 1$. If $\mathbb{Q} \ll \mathbb{P}$, one can additionally show that $\mathbb{P}(Z > 0) = 1$ is equivalent to $\mathbb{P} \sim \mathbb{Q}$.

For $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{Q})$ one has

$$\mathbb{E}_{\mathbb{Q}} X = \mathbb{E}_{\mathbb{P}}[XZ]. \tag{B.13}$$

For *conditional* expectations, the corresponding result is different.

Proposition B.39 *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$ and assume that $\mathbb{Q} \ll \mathbb{P}$ with $Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let X be a random variable. Then $\mathbb{E}_{\mathbb{Q}}|X| < \infty$ iff $\mathbb{E}_{\mathbb{P}}|X|Z < \infty$ and in either case we have*

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}] = \frac{\mathbb{E}_{\mathbb{P}}[XZ|\mathcal{G}]}{\mathbb{E}_{\mathbb{P}}[Z|\mathcal{G}]} \text{ a.s. w.r.t. } \mathbb{Q}. \tag{B.14}$$

If $\mathbb{Q} \sim \mathbb{P}$, then this equality is also valid $\mathbb{P}$-a.s.

Put $\tilde{Z} = \frac{Z}{\mathbb{E}_{\mathbb{P}}[Z|\mathcal{G}]}$, a kind of conditionally normalized version of Z as $\mathbb{E}_{\mathbb{P}}[\tilde{Z}|\mathcal{G}] = 1$. Observe that (B.14) can then be written as $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}] = \mathbb{E}_{\mathbb{P}}[X\tilde{Z}|\mathcal{G}]$, much like formula (B.13) for unconditional expectations.

Index