

STOCHASTIC FINANCIAL MODELS

SEBASTIAN ANDRES

CONTENTS

0. Motivation	2
1. Utility and mean variance	4
1.1. Reservation and marginal prices	6
1.2. Mean-variance analysis and the efficient frontier	8
1.3. The Capital Asset Pricing Model (CAPM)	12
1.4. Equilibrium pricing	14
2. Conditional expectations and martingales	14
2.1. Conditional expectations	14
2.2. Martingales	20
2.3. Martingale convergence	23
2.4. Stopping times and optional stopping	25
3. Arbitrage theory in discrete time	26
3.1. Single period model	27
3.2. Multi-period model	30
3.3. European contingent claims	33
4. The Cox-Ross-Rubinstein binomial model	37
5. Dynamic programming	45
6. Brownian motion	45
6.1. Definition and basic properties	46
6.2. The Reflection principle	50
6.3. Change of measure: The Cameron-Martin theorem	52
6.4. Martingales associated with Brownian motion	55
7. The Black-Scholes model	56
7.1. Black-Scholes via change of measure	57
7.2. The Black-Scholes Model as limit of the Binomial Model	58
7.3. Black-Scholes pricing formula for European Calls and Puts	62
7.4. Pricing exotic options	64
7.5. The Black-Scholes PDE	65
7.6. Numerical schemes	67
References	70

These notes are designed for the lecture course 'Stochastic Financial Models' in Part II of the mathematical tripos at the University of Cambridge. The course material covers roughly the first five chapters of [10]. Here we will mainly follow the

Date: March 14, 2019.

lecture notes by Friz and Rogers [8]. In fact, Section 1 below is almost identical to [8, Section 1], but is still included here for completeness. The selection of the material in Sections 2, 6 and 7 is also quite similar to [8] but the presentation deviates more significantly. Section 2.1 on conditional expectations is mainly based on [13] and the sections about martingales follows largely the book of Rogers and Williams [12], see also [2, Section 4]. Sections 3 and 4 are taken from certain parts of [7, Chapter 5], where much further material can be found. Finally, students who are interested in the more financial aspects of the topic are referred to [9].

0. MOTIVATION

An investor needs a certain quantity of a share (or currency, good, ...), however not right now at time $t = 0$ but at a later time $t = 1$. The price of the share $S(\omega)$ at $t = 1$ is random and uncertain, but already now at time $t = 0$ one has to make some calculations with it, which leads to the presence of risk. (Example: 500 USD are roughly 370 GBP today, but in one year?) A possible solution for the investor would be to purchase a financial derivative such as

- *Forward contract:* The owner of a forward contract has the right and the obligation to buy a share at time $t = 1$ for a delivery price K specified at time $t = 0$. Thus, the owner of the forward contract gains the difference between the actual market price $S(\omega)$ and the delivery price K if $S(\omega)$ is larger than K . If $S(\omega) < K$, the owner loses the amount $K - S(\omega)$ to the issuer of the forward contract. Hence, a forward contract corresponds to the random payoff

$$H(\omega) = S(\omega) - K.$$

- *Call option:* The owner of a call option has the right but not the obligation to buy a share at time $t = 1$ for a the strike price K specified at time $t = 0$. Thus, if $S(\omega) > K$ at time $t = 1$ the owner of the call gains again $S(\omega) - K$, but if $S(\omega) \leq K$ the owner buys the share from the market, and the call becomes worthless in this case. Hence, at time $t = 1$ the random payoff of the call option is given by

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

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

A classical approach to this problem is to regard the random payoff $H(\omega)$ as a 'lottery' modelled as a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with some 'objective' probability measure $\mathbb{P}$. Then the fair price is given the expected discounted payoff $\mathbb{E}[\frac{H}{1+r}]$, where $r \geq 0$ is the interest rate for both fund and loans from $t = 0$ to $t = 1$. Here we implicitly assume that both interest rates are the same, which seems reasonable for large investors. The economic reason for working with discounted prices is that one should distinguish between payments at time $t = 0$

and ones at time $t = 1$. Usually, people tend to prefer a certain amount today over the same amount paid at a later time, and this preference is reflected in the interest rate r paid by the riskless bond (riskless asset, bank account). An investment of the amount $1/(1 + r)$ at time zero in the bond results in value 1 at time $t = 1$.

Less classical approaches also take a subjective assessment of the risk by the involved agents (in this case buyer and seller of the derivative) into account (cf. Section 1 below).

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

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

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

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

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

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

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

$$H(\omega) = V(\omega), \quad \forall \omega \in \Omega.$$

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

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

which implies

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

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

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

we have an arbitrage. These considerations lead us to the questions we will mainly address in this lecture course.

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

1. UTILITY AND MEAN VARIANCE

A market is the interaction of agents trading goods and services, and the actions and choices of the individual agents are shaped by preferences over different *contingent claims*. A contingent claim is simply a well-specified random payment, mathematically, a random variable. We shall suppose that agents' preferences are expressed by an *expected utility representation*, that is

$$Y \text{ is preferred to } X \iff \mathbb{E}[U(X)] \leq \mathbb{E}[U(Y)],$$

where $U : \mathbb{R} \rightarrow [-\infty, \infty)$ is a non-decreasing *utility function*, so for any amount $x \in \mathbb{R}$ the value $U(x)$ represents the 'utility' of x for the respective agent. The agents may have different preferences, so every agent chooses a utility function individually and the choices may differ from one agent to another. We will assume that U is concave.

Definition 1.1. A function $U : \mathbb{R} \rightarrow [-\infty, \infty)$ is said to be *concave* if for all $p \in [0, 1]$,

$$pU(x) + (1-p)U(y) \leq U(px + (1-p)y), \quad \forall x, y \in \mathbb{R}.$$

Set $\mathcal{D}(U) := \{x : U(x) > -\infty\}$.

Remark 1.2. (i) If U is concave then $-U$ is convex.

(ii) If U is concave, then by Jensen's inequality,

$$\mathbb{E}[U(X)] \leq U(\mathbb{E}[X]),$$

so if an agent is offered the choice of a contingent claim X and a certain payment of $\mathbb{E}[X]$ we will prefer the latter. This property is called *risk-aversion*. Similarly if U is linear then the agent is *risk-neutral* and if U is convex he is *risk-friendly*.

(iii) $U(x) = -\infty$ means that the outcome x is unacceptable.

Example 1.3 (Examples for utility functions). (i) The function

$$U(x) = -\exp(-\gamma x)$$

with parameter $\gamma > 0$ is called the *constant absolute risk aversion (CARA)* utility.

(ii) The function

$$U(x) = \begin{cases} \frac{x^{1-R}}{1-R} & \text{if } x \geq 0, \\ -\infty & \text{if } x < 0, \end{cases}$$

with parameter $R \in (0, \infty) \setminus \{1\}$ is called the *constant relative risk aversion (CRRA)* utility.

(iii) The function

$$U(x) = \begin{cases} \log x & \text{if } x > 0, \\ -\infty & \text{if } x \leq 0, \end{cases}$$

is the *logarithmic* utility. It is often regarded as the CRRA utility with parameter $R = 1$.

(iv) $U(x) = \min(x, \alpha x)$ for $\alpha \in [0, 1)$.

(v) $U(x) = -\frac{1}{2}x^2 + ax$ for $a \geq 0$ is concave but not increasing.

(vi) If U_1 and U_2 are utilities then $\alpha_1 U_1 + \alpha_2 U_2$ is again a utility for any $\alpha_1, \alpha_2 > 0$.

(vii) If $\{U_\lambda, \lambda \in \Lambda\}$ is a family of utilities, then

$$U(x) = \inf_{\lambda \in \Lambda} U_\lambda(x)$$

is again a utility.

The certainly easiest criterion to check the concavity of a function is to verify the non-positivity of its second derivative in the case it exists. We will now briefly derive a more general characterisation.

Proposition 1.4. *A function $U : \mathbb{R} \rightarrow [-\infty, \infty)$ is concave if and only if for all $x_1, y_1, x_2, y_2 \in \mathcal{D}(U)$ such that $x_1 < y_1 \leq x_2 < y_2$ we have*

$$\frac{U(y_1) - U(x_1)}{y_1 - x_1} \geq \frac{U(y_2) - U(x_2)}{y_2 - x_2}. \quad (1.1)$$

Proof. First, suppose that U is concave. It is enough to show (1.1) in the case $y_1 = x_2$, that is

$$\frac{U(z) - U(x)}{z - x} \geq \frac{U(y) - U(z)}{y - z}, \quad \forall x < z < y,$$

which is equivalent to

$$U(z) \geq \frac{z - x}{y - x} U(y) + \frac{y - z}{y - x} U(x). \quad (1.2)$$

Setting $p := (z - x)/(y - x) \in [0, 1]$ we observe that $z = py + (1 - p)x$ and (1.2) holds by the concavity of U . Conversely, if (1.1) holds, then (1.2) also holds and implies concavity. $\square$

By taking limits we immediately get the following statement.

Corollary 1.5. (i) *Let U be concave. For any $z \in \text{int}\mathcal{D}(U)$ the left- and right-hand derivatives*

$$U'_-(z) := \lim_{x \uparrow z} \frac{U(z) - U(x)}{z - x}, \quad U'_+(z) := \lim_{y \downarrow z} \frac{U(y) - U(z)}{y - z}$$

exist. Both U'_- and U'_+ are decreasing functions and satisfy $U'_- \geq U'_+$.

(ii) *If $U \in C^2(\mathbb{R})$, then $U''(x) \leq 0$ for all $x \in \mathbb{R}$ if and only if U is concave.*

From now on, unless stated otherwise, we will make the following

Assumption. All utility functions are strictly increasing and strictly concave.

Now consider an agent with wealth w and utility function $U \in C^2(\mathbb{R})$ who is contemplating whether or not to accept a contingent claim X . He will do so provided

$$\mathbb{E}[U(w + X)] > U(w).$$

If we suppose that X is small so that we may perform a Taylor expansion, this condition is approximately the same as the condition

$$U(w) + U'(w) \mathbb{E}[X] + \frac{1}{2} U''(w) \mathbb{E}[X^2] > U(w). \quad (1.3)$$

Since $U'(w) > 0$ (U is strictly increasing) and $U''(w) < 0$ (U is strictly concave) the benefits of a positive mean $\mathbb{E}[X]$ are offset by the disadvantage of positive variance; the balance is just right (to this order of approximation) when

$$\frac{2 \mathbb{E}[X]}{\mathbb{E}[X^2]} = -\frac{U''(w)}{U'(w)},$$

where the right-hand side is the so-called *Arrow-Pratt coefficient of absolute risk aversion*. If we consider instead the effect of the proposed gamble to be multiplicative rather than additive, the decision for the agent will be to accept if

$$\mathbb{E}[U(w(1 + X))] > U(w).$$

Assuming that $w > 0$, a similar argument shows that to this order of approximation the agent should accept when

$$\frac{2 \mathbb{E}[X]}{\mathbb{E}[X^2]} \geq -\frac{w U''(w)}{U'(w)},$$

where the right-hand side is the so-called *Arrow-Pratt coefficient of relative risk aversion*. This explains the names of the CARA and CRRA utilities, for which the Arrow-Pratt coefficients are constant γ and R , respectively.

1.1. Reservation and marginal prices. Although the derivations above are not rigorous, they do build our intuition. Developing this intuitive theme a bit further, let us consider an agent with utility U who is able to choose any contingent claim X from an admissible set $\mathcal{A}$; he will naturally choose X to achieve

$$\sup_{X \in \mathcal{A}} \mathbb{E}[U(X)].$$

We shall suppose that the supremum is achieved at some $X^* \in \mathcal{A}$. In the special case where $\mathcal{A}$ is an affine space¹ taking the form $\mathcal{A} = X + \mathcal{V}$ for some vector space $\mathcal{V}$, we have therefore that for all $\xi \in \mathcal{V}$ and all $t \in \mathbb{R}$,

$$\mathbb{E}[U(X^*)] \geq \mathbb{E}[U(X^* + t\xi)],$$

¹That is, for any $X_1, X_2 \in \mathcal{A}$ and $t \in \mathbb{R}$, $tX_1 + (1 - t)X_2 \in \mathcal{A}$. Equivalently, there exists a vector space $\mathcal{V}$ such that for any $X \in \mathcal{A}$ we have $\mathcal{A} = X + \mathcal{V}$.

and formally differentiating the right hand side with respect to t gives

$$\mathbb{E} [U'(X^*)\xi] = 0, \quad \forall \xi \in \mathcal{V}. \quad (1.4)$$

Suppose now that the agent considers whether to buy a contingent claim Y for price π . To fix our ideas, let us suppose that $Y \geq 0$, though this is not essential. For any π for which

$$\sup_{X \in \mathcal{A}} \mathbb{E} [U(X + Y - \pi)] \geq \mathbb{E} [U(X^*)],$$

he would be willing to buy Y ; the largest such π , denoted by $\pi_b(Y)$, is called the (*reservation*) *bid price*. Similarly, the (*reservation*) *ask price* $\pi_a(Y)$ is the smallest value of π such that

$$\sup_{X \in \mathcal{A}} \mathbb{E} [U(X - Y + \pi)] \geq \mathbb{E} [U(X^*)],$$

Obviously, $\pi_b(Y) = -\pi_a(-Y)$. Moreover,

$$\pi_a(Y) \geq \pi_b(Y), \quad \text{'ask above, bid below'}. \quad (1.5)$$

Proof of (1.5). Let X' and X'' be the optimal choices from $\mathcal{A}$ when selling Y for $\pi_a(Y)$ and when buying Y for $\pi_b(Y)$, respectively. Then,

$$\mathbb{E} [U(X' - Y + \pi_a(Y))] = \mathbb{E} [U(X^*)] = \mathbb{E} [U(X'' + Y - \pi_b(Y))].$$

Further, since U is concave and $\frac{1}{2}(X' + X'') \in \mathcal{A}$ since $\mathcal{A}$ is convex, we get

$$\begin{aligned} & \sup_{X \in \mathcal{A}} \mathbb{E} [U(X + \frac{1}{2}(\pi_a(Y) - \pi_b(Y)))] \\ & \geq \mathbb{E} [U(\frac{1}{2}(X' + X'') + \frac{1}{2}(\pi_a(Y) - \pi_b(Y)))] \\ & \geq \frac{1}{2} \left(\mathbb{E} [U(X' - Y + \pi_a(Y))] + \mathbb{E} [U(X'' + Y - \pi_b(Y))] \right) \\ & = \mathbb{E} [U(X^*)] = \sup_{X \in \mathcal{A}} \mathbb{E} [U(X)], \end{aligned}$$

and (1.5) follows since U is strictly increasing. $\square$

Further, one can show that for $0 < \alpha < \beta$,

$$\frac{\pi_b(\beta Y)}{\beta} \leq \frac{\pi_b(\alpha Y)}{\alpha}, \quad \frac{\pi_a(\beta Y)}{\beta} \geq \frac{\pi_a(\alpha Y)}{\alpha}.$$

In particular, the mapping

$$f_Y : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R} : \quad t \mapsto \frac{\pi_a(tY)}{t}$$

is increasing and therefore limits at zero from either side exist. For $t \neq 0$ let now X_t^* be defined via

$$\sup_{X \in \mathcal{A}} \mathbb{E} [U(X - tY + \pi_a(tY))] = \mathbb{E} [U(X_t^* - tY + \pi_a(tY))].$$

Then, a (non-justified) Taylor expansion gives

$$\begin{aligned}
\mathbb{E}[U(X^*)] &= \mathbb{E}[U(X_t^* - tY + \pi_a(tY))] \\
&= \mathbb{E}[U(X^* + (X_t^* - X^*) - tY + tf_Y(t))] \\
&= \mathbb{E}\left[U(X^*) + U'(X^*)\{(X_t^* - X^*) - tY + tf_Y(t)\}\right] + o(t) \\
&= \mathbb{E}\left[U(X^*) + U'(X^*)\{-tY + tf_Y(t)\}\right] + o(t).
\end{aligned}$$

In the last step we used (1.4) and the fact that $X_t^* - X^* \in \mathcal{V}$. Further, for the remainder term to be in $o(t)$ we actually also require that $|X_t^* - X^*| \in o(t^{1/2})$ $\mathbb{P}$ -a.s., for instance. Nevertheless, this non-rigorous computation implies

$$\lim_{t \rightarrow 0} \frac{\pi_a(tY)}{t} = \frac{\mathbb{E}[U'(X^*)Y]}{\mathbb{E}[U'(X^*)]}. \quad (1.6)$$

This expression is the agents marginal price for Y , that is, the price per unit at which he would be prepared to buy or sell an infinitesimal amount of Y . Notice that the marginal price is linear in the contingent claim, in contrast to the bid and ask prices. If prices had been derived from some economic equilibrium, and the contingent claim Y was one which was marketed, then the market price of Y would have to equal the marginal price of Y given by (1.6), and this would have to hold for every agent. This is not to say that for every agent the marginal utility of optimal wealth would have to be the same; in general they are not. But the prices obtained by each agent from their marginal utility of optimal wealth via (1.6) would have to agree on all marketed contingent claims.

This heuristic discussion provides us with firm guidance for our intuition, and the form of the prices frequently fits (1.6). Although there are many steps where the analysis could fail, where we assume that suprema are attained, or that we can differentiate under the expectation, the most common reason for the above analysis to fail is that $\mathcal{A}$ is not an affine space! For a mathematically more rigorous discussion we refer to [5].

1.2. Mean-variance analysis and the efficient frontier. In the discussion below (1.3) it has already been indicated that a risk-averse agent with expected utility preferences will tend to accept contingent claims with large mean and small variance. In other words, given a choice of contingent claims, all with the same mean, the agent should take the one with smallest variance. This is the main idea of *mean-variance analysis*.

Consider a single-period model with d assets in which an agent may invest; the non-random prices of the assets at time $t = 0$ are denoted by $S_0 = (S_0^1, \dots, S_0^d)^T$ and their random values at time $t = 1$ by $S_1 = (S_1^1, \dots, S_1^d)^T$. Further, let

$$\mu := \mathbb{E}[S_1] \in \mathbb{R}^d, \quad V := (\text{cov}(S_1^i, S_1^j))_{i,j=1,\dots,d} = \mathbb{E}[(S_1 - \mathbb{E}[S_1])(S_1 - \mathbb{E}[S_1])^T]$$

denote mean vector and covariance matrix of S_1 , respectively. Suppose that at time $t = 0$ the agent chooses to hold θ^j units of asset j for $j = 1, \dots, d$ and write

$\theta = (\theta^1, \dots, \theta^d)^T$. Then at time $t = 1$ his portfolio is worth

$$w_1 = \theta \cdot S_1 = \sum_{i=1}^d \theta^i S_1^i$$

with

$$\mathbb{E}[w_1] = \theta \cdot \mu, \quad \text{var}(w_1) = \theta \cdot V \theta.$$

(Here and below $x \cdot y = x^T y$ with $x, y \in \mathbb{R}^d$ denotes the canonical scalar product in $\mathbb{R}^d$.) If the agent now requires to choose θ to give a predetermined mean value $\mathbb{E}[w_1] = m$ and to have minimal variance, then his optimisation problem is to find

$$\min_{\theta} \frac{1}{2} \theta \cdot V \theta \quad \text{subject to} \quad \theta \cdot \mu = m, \quad \theta \cdot S_0 = w_0. \quad (1.7)$$

The second constraint is the budget constraint, that the cost at time $t = 0$ of the chosen portfolio must equal the agents wealth w_0 at time $t = 0$. To solve this, we introduce the Lagrangian

$$L = \frac{1}{2} \theta \cdot V \theta + \lambda_1 (m - \theta \cdot \mu) + \lambda_2 (w_0 - \theta \cdot S_0).$$

Assuming V is regular, this is minimised by choosing

$$\theta = V^{-1}(\lambda_1 \mu + \lambda_2 S_0).$$

We still need to determine the multipliers λ_1 and λ_2 using the constraints in (1.7). By the choice of θ ,

$$\begin{pmatrix} m \\ w_0 \end{pmatrix} = \begin{pmatrix} \mu^T \\ S_0^T \end{pmatrix} \theta = \begin{pmatrix} \mu^T \\ S_0^T \end{pmatrix} V^{-1} \begin{pmatrix} \mu & S_0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mu \cdot V^{-1} \mu & \mu \cdot V^{-1} S_0 \\ \mu \cdot V^{-1} S_0 & S_0 \cdot V^{-1} S_0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

Recall that the inverse of a general regular 2×2 matrix is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Hence, provided μ is not a multiple of S_0 , we get

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} S_0 \cdot V^{-1} S_0 & -\mu \cdot V^{-1} S_0 \\ -\mu \cdot V^{-1} S_0 & \mu \cdot V^{-1} \mu \end{pmatrix} \begin{pmatrix} m \\ w_0 \end{pmatrix},$$

where $\Delta := (\mu \cdot V^{-1} \mu)(S_0 \cdot V^{-1} S_0) - (\mu \cdot V^{-1} S_0)^2$. For the variance of w_1 we obtain

$$\begin{aligned} \theta \cdot V \theta &= (\lambda_1 \quad \lambda_2) \begin{pmatrix} \mu^T \\ S_0^T \end{pmatrix} V^{-1} \begin{pmatrix} \mu & S_0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = (\lambda_1 \quad \lambda_2) \begin{pmatrix} m \\ w_0 \end{pmatrix} \\ &= \frac{1}{\Delta} (m \quad w_0) \begin{pmatrix} S_0 \cdot V^{-1} S_0 & -\mu \cdot V^{-1} S_0 \\ -\mu \cdot V^{-1} S_0 & \mu \cdot V^{-1} \mu \end{pmatrix} \begin{pmatrix} m \\ w_0 \end{pmatrix} \\ &= \frac{1}{\Delta} (m^2 S_0 \cdot V^{-1} S_0 - 2mw_0 S_0 \cdot V^{-1} \mu + w_0^2 \mu \cdot V^{-1} \mu), \end{aligned}$$

which is *quadratic* in the required mean m . This variance is minimised (over m) to the value $w_0^2 / (S_0 \cdot V^{-1} S_0)$ and the minimum is attained for $m = w_0 (S_0 \cdot V^{-1} \mu) / (S_0 \cdot V^{-1} S_0)$, which corresponds to the portfolio $\theta = (w_0 / S_0 \cdot V^{-1} S_0) V^{-1} S_0$.

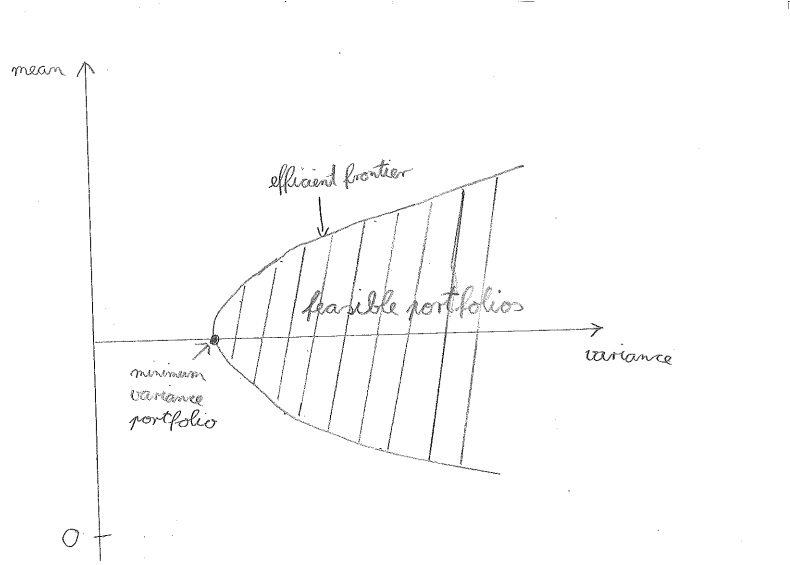


FIGURE 1. Mean-variance efficient frontier

We can display the conclusions of this analysis graphically by a parabola (see Figure 1). For any chosen value of the mean m , corresponding to a given level, values of the portfolio variance corresponding to points to the left of the parabola are not achievable, whereas points on and to the right of the parabola are. The parabola is called the *mean-variance efficient frontier*.

We noted already earlier that while it is natural to think that from among all available contingent claims with given mean we should choose the one with the smallest variance, this is not in general correct. Indeed, assuming our agent has expected-utility preferences, if two contingent claims are considered equally desirable if they have the same mean and the same variance, then the utility must be a function only of the mean and the variance. But in this case one can check (see exercises) that the utility must be quadratic, which is disqualified because it is not increasing.

Despite this, the kind of mean-variance analysis set forth above, and graphs efficient frontiers are ubiquitous in the practice of portfolio management. There are two reasons for this:

- (i) This analysis is just about as sophisticated as you can expect to put across to the mathematically untrained.
- (ii) In one very special situation, when S_1 is Gaussian, the mean-variance analysis in effect amounts to the correct expected-utility maximisation. We study this situation right now, assuming that agents have CARA utilities.

Example 1.6 (CARA, S_1 Gaussian, no riskless asset). Suppose that S_1 is $\mathcal{N}(\mu, V)$ -distributed, that is normal distributed with mean vector $\mu \in \mathbb{R}^d$ and regular covariance matrix V . Further, suppose that the agent has CARA utility, and so he aims to

maximise

$$\mathbb{E} \left[-\exp(-\gamma w_1) \right] \quad \text{with} \quad w_1 = \theta \cdot S_1 = \sum_{j=1}^d \theta^j S_1^j.$$

Recall that since $S_1 \sim \mathcal{N}(\mu, V)$ we have $\mathbb{E} [\exp(x \cdot S_1)] = \exp(x \cdot \mu + \frac{1}{2} x \cdot V x)$ for every $x \in \mathbb{R}^d$. So the agents objective is to minimise

$$\mathbb{E} [\exp(-\gamma \theta \cdot S_1)] = \exp(-\gamma \theta \cdot \mu + \frac{1}{2} \gamma^2 \theta \cdot V \theta),$$

again under a budget constraint $\theta \cdot S_0 = w_0$. This leads to the optimisation problem to minimise

$$-\gamma \theta \cdot \mu + \frac{1}{2} \gamma^2 \theta \cdot V \theta \quad \text{subject to} \quad \theta \cdot S_0 = w_0.$$

Using the Lagrangian method, we convert this problem into the unconstrained minimisation of

$$-\gamma \theta \cdot \mu + \frac{1}{2} \gamma^2 \theta \cdot V \theta + \gamma \lambda (w_0 - \theta \cdot S_0).$$

Differentiating with respect to θ gives

$$\gamma V \theta = \mu + \lambda S_0,$$

which is solved by taking

$$\theta = \gamma^{-1} V^{-1} (\mu + \lambda S_0),$$

and from the budget constraint $\theta \cdot S_0 = w_0$ we get

$$\lambda = \frac{\gamma w_0 - S_0 \cdot V^{-1} \mu}{S_0 \cdot V^{-1} S_0}.$$

Hence, the optimal θ has the explicit form

$$\theta = \gamma^{-1} V^{-1} \mu + \frac{\gamma w_0 - S_0 \cdot V^{-1} \mu}{\gamma S_0 \cdot V^{-1} S_0} V^{-1} S_0. \quad (1.8)$$

Notice that this optimal portfolio is a weighted average of two portfolios, the minimum-variance portfolio $V^{-1} S_0$, which minimises the variance of $w_1 = \theta \cdot S_1$ subject to the initial budget constraint $\theta \cdot S_0 = w_0$ (see above), and the diversified portfolio $V^{-1} \mu$. This is an example of a mutual fund theorem.

Example 1.7 (CARA, S_1 Gaussian, with a riskless asset). Consider exactly the situation of the previous example, but add one more asset, denoted S^0 , whose return is riskless (bond), that is at time $t = 0$ it has initial value $S_0^0 > 0$ and at has value

$$S_1^0 = S_0^0 (1 + r),$$

where r is the *riskless interest rate*. Let $\bar{S} = (S^0, S^1, \dots, S^d)^T$ denote the enlarged vector of assets, with corresponding mean vector $\bar{\mu} = (S_0^0 (1+r), \mu)^T$ and covariance matrix $\bar{V}$, where all entries of $\bar{V}$ in the zeroth row and the zeroth column will be

zero. Again we are aiming to maximise the expected CARA utility. So, similarly as before we need to minimise

$$-\gamma \bar{\theta} \cdot \bar{\mu} + \frac{1}{2} \gamma^2 \bar{\theta} \cdot \bar{V} \bar{\theta} + \gamma \lambda (w_0 - \bar{\theta} \cdot \bar{S}_0),$$

where $\bar{\theta} = (\theta^0, \theta^1, \dots, \theta^d)^T$ and θ^0 is the number of units of bonds the agent is holding. Then, differentiating with respect to $\bar{\theta}$ gives the condition

$$\gamma \bar{V} \bar{\theta} = \bar{\mu} + \lambda \bar{S}_0.$$

However, $\bar{V}$ is not invertible, but the equation in the top row yields

$$\lambda = -\frac{S_0^0(1+r)}{S_0^0} = -(1+r).$$

Solving the remaining equations as before gives $\theta = \gamma^{-1} V^{-1}(\mu + \lambda S_0)$, and we obtain

$$\theta = \gamma^{-1} \theta_M := \gamma^{-1} V^{-1}(\mu - (1+r)S_0). \quad (1.9)$$

Remark 1.8. (i) Once again, the optimal portfolio (1.9) is a weighted average of the minimum-variance portfolio and the diversified portfolio, though this time the weights are in fixed proportions. The portfolio θ_M is referred to as the *market portfolio* (think of a major share index), for reasons we shall explain shortly.

(ii) Notice that in contrast to the solution (1.8) to the previous example, the optimal θ does not depend on w_0 , the initial wealth of the agent. How can this be reconciled with the initial budget constraint? Very simply: the agent takes up the portfolio (1.9) in the risky assets, and his holding θ^0 of the riskless asset adjusts to pay for it.

(iii) Looking at (1.9), we see that the more risk-averse the agent is (that is, the larger γ), the less he invests in the risky assets - evidently sensible. If we took the simple special case where V were diagonal, we see that the position in asset j is

$$\theta^j = \frac{\mu^j - (1+r)S_0^j}{\gamma V_{jj}}$$

proportional to the excess mean return $R_j := \mu^j - (1+r)S_0^j$ of asset j , that is, the average amount by which investing in asset j improves upon investing the same initial amount S_0^j in the riskless asset. We also see that the higher the variance of asset j , the less we are prepared to invest in it, again evidently sensible.

1.3. The Capital Asset Pricing Model (CAPM). In the situation of Example 1.7 let us assume that all covariances, variances, mean rates of return of stocks and so on are known to all agents, who are supposed to be all risk-averse rational investors using the same mean-variance approach portfolio selection. Then each agent will have a portfolio on the same efficient frontier, and hence has a portfolio that is a mixture of the risk-free asset and a unique efficient fund of risky assets, namely the market portfolio $\theta_M = V^{-1}(\mu - (1+r)S_0)$ derived in Example 1.7.

Consider now an agent with initial wealth $w_0 = \theta_M \cdot S_0$ at time $t = 0$, who is investing in the market portfolio. Then, the value of his portfolio at time $t = 1$ is random with mean

$$\mu_M = \theta_M \cdot \mu.$$

On the other hand, investing w_0 in the bond S^0 would give him a certain return of $(1 + r)\theta_M \cdot S_0$, so the mean excess return of the market portfolio is given by

$$R_M := \mu_M - (1 + r)\theta_M \cdot S_0 = \theta_M \cdot V\theta_M.$$

Now for each asset i we define the *beta* of that asset by

$$\beta_i := \frac{\text{cov}(S_1^i, \theta_M \cdot S_1)}{\text{var}(\theta_M \cdot S_1)}, \quad i = 1, \dots, d.$$

The quantity β_i serves as an important measure of risk for individual assets that is different from $\text{var}(S_1^i)$. More precisely, for any asset i , $\text{var}(S_1^i)$ describes the risk associated with its own fluctuations around its mean, also called *unsystematic risk*, also known as *specific risk* or *diversifiable risk*. Unsystematic risk can be reduced through diversification. On the other hand, β_i measures the uncertainty inherent to the entire market or entire market segment, also known as *non-diversifiable risk*, *market risk* or *systematic risk*.

We observe that β_i can be rewritten as

$$\beta_i = \frac{(V\theta_M)^i}{\theta_M \cdot V\theta_M} = \frac{(\mu - (1 + r)S_0)^i}{\theta_M \cdot V\theta_M},$$

so we obtain for the mean excess return R_i of asset i that

$$R_i := \mu^i - (1 + r)S_0^i = \beta_i \theta_M \cdot V\theta_M = \beta_i R_M,$$

so the mean excess return of asset i equals β_i times the mean excess return of the market portfolio,

$$R_i = \beta_i R_M. \tag{1.10}$$

Is this a profound result, or merely a tautologous reworking of the definition of β_i ? It is both; the profundity lies in the fact that (1.10) expresses a relation between on the one hand the mean rates of return of individual assets and of the market portfolio, and on the other, the variances and covariances of asset returns, which could all be estimated very easily from market data, thereby providing a test of the CAPM analysis. It is rare to find a verifiable prediction from economic theory; sadly, it turns out in practice to be very hard to make reliable estimates of rates of return (see [8, Section 1.3] for more details).

1.4. Equilibrium pricing. We end this section with a short discussion about equilibrium pricing. So far we have been looking at a market with d risky assets whose values S_1 at time $t = 1$ are Gaussian random variables, and whose values S_0 at time $t = 0$ are given constants; but where did those constants come from? How were they determined? An economist would answer these questions by saying that the prices at time 0 are *equilibrium prices*, determined by the agents in the market and their interaction. The central idea of equilibrium analysis is that we now *adjust the prices until the market is cleared*, that is, the supply and demand are matched

Suppose there is unit net supply of asset i , for each $i = 1, \dots, d$, and zero net supply of the riskless asset; the (equilibrium) prices must be such that the total demand of all agents for each risky asset is 1, and for riskless asset the total demand is 0. Without loss of generality, we assume that $S_0^0 = 1$. Further, suppose that there are K agents in the market, agent k having CARA utility with coefficient of absolute risk aversion γ_k , and that agent k enters the market. According to (1.9) will hold a portfolio $\theta_k \in \mathbb{R}^d$ in the risky assets given by

$$\theta_k = \gamma_k^{-1} \theta_M.$$

Thus, the total holdings of all agents in the market will be

$$\sum_{k=1}^K \theta_k = \Gamma^{-1} \theta_M = \Gamma^{-1} V^{-1} (\mu - (1+r)S_0) \quad (1.11)$$

where $\Gamma = (\sum_k \gamma_k^{-1})^{-1}$. Now market-clearing at time $t = 0$ requires that

$$\sum_{k=1}^K \theta_k = \mathbf{1}. \quad (1.12)$$

where we write $\mathbf{1} := (1, \dots, 1)^\top \in \mathbb{R}^d$. By combining (1.11) and (1.12) we see that the market-clearing prices for the risky assets must be

$$S_0 = \frac{(\mu - \Gamma V \mathbf{1})}{1+r}.$$

We still need to check that the total demand for the riskless asset is zero, which we leave as an exercise.

2. CONDITIONAL EXPECTATIONS AND MARTINGALES

2.1. Conditional expectations. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, its expected value $\mathbb{E}[X]$, provided it exists, serves as a prediction for the random outcome of X . From now on we will occasionally also write $\mathbb{E}[X] = \int X \, d\mathbb{P}$ and $\mathbb{E}[X \mathbf{1}_A] = \int_A X \, d\mathbb{P}$ for $A \in \mathcal{F}$.

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

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

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

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

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

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

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

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

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

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

$$\mathbb{P}[A | \text{Dow Jones tomorrow} = x], \quad x = 0, 1, \dots$$

as a function of x .

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

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

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

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

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

Definition 2.3. The random variable

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[X_0 \mathbb{1}_A]. \quad (2.2)$$

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

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

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

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

(ii) If $\mathcal{F}_0 = \sigma(Z)$ for any random variable Z , then $\mathbb{E}[X | Z] := \mathbb{E}[X | \sigma(Z)]$ is $\sigma(Z)$ -measurable by condition (i). In particular, by the so-called factorisation lemma (see e.g. [1]) it is of the form $f(Z)$ for some function f . It is then common to define

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

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

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

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

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

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

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

The existence follows rather easily from the following important result in measure theory.

Theorem 2.11 (Radon-Nikodym, 1930). *Let μ be a measure and ν be a probability measure on $(\Omega, \mathcal{F})$. Then the following are equivalent.*

- (i) *μ is absolutely continuous with respect to ν (notation: $\mu \ll \nu$), that is for every $A \in \mathcal{F}$ we have $\nu(A) = 0 \Rightarrow \mu(A) = 0$.*

(ii) There exists an $\mathcal{F}$ -measurable function $\varphi \geq 0$ such that

$$\mu(A) = \int_A \varphi \, d\nu, \quad \forall A \in \mathcal{F}.$$

The function φ is called density or Radon-Nikodym derivative and is denoted by $\frac{d\mu}{d\nu}$. Moreover, φ is unique up to ν -null sets.

Proof. See, for instance, [1, Chapter 17]. $\square$

Proof of Theorem 2.10. (i) Define $\mu(A) = \int_A X \, d\mathbb{P}$, $A \in \mathcal{F}$. Then μ is a measure and $\mu \ll \mathbb{P}$ on $\mathcal{F}$. In particular, $\mu \ll \mathbb{P}$ also on $\mathcal{F}_0$. We apply Theorem 2.11 (on the space $(\Omega, \mathcal{F}_0)$) to obtain that there exists an $\mathcal{F}_0$ -measurable function

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

In other words, $\int_{A_0} X \, d\mathbb{P} = \int_{A_0} X_0 \, d\mathbb{P}$ or $\mathbb{E}[X \mathbb{1}_{A_0}] = \mathbb{E}[X_0 \mathbb{1}_{A_0}]$ for all $A_0 \in \mathcal{F}_0$.

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

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

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

Thus,

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

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

2.1.3. Properties of conditional expectations.

Proposition 2.12. *The conditional expectation has the following properties.*

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

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

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

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

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

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

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

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

Proof. (i) follows directly from the definition of the conditional expectation. Statements (ii)-(vi) all follow from the corresponding properties of the expected value, and (vii) can be shown similarly as for the usual expected value. $\square$

Proposition 2.13. *Let $Y_0 \geq 0$ be $\mathcal{F}_0$ -measurable. Then,*

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

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

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

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

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

For general Y_0 the statement follows by linearity and approximation. $\square$

Proposition 2.14 ('Projectivity' or 'Tower property' of conditional expectations). *Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$ be σ -algebras. Then,*

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

Proof. Let $A \in \mathcal{F}_0$. Then, clearly $A \in \mathcal{F}_1$ and therefore

$$\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_1] \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[\mathbb{E}[X | \mathcal{F}_1] | \mathcal{F}_0] \mathbb{1}_A].$$

$\square$

Proposition 2.15. *Let X be independent of $\mathcal{F}_0$ ². Then,*

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

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

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

$\square$

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

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

Proposition 2.16. *Let X and Y be real-valued random variables with densities f_X and f_Y . Assume that (X, Y) admits a joint density f_{XY} . Then the conditional distribution of X given Y is a random distribution with density*

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

and the conditional expectation of X given Y is

$$\mathbb{E}[X | Y] = \int_{\mathbb{R}} x f_{X|Y}(x) dx.$$

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

Proposition 2.17. *Let $F : \mathbb{R}^2 \rightarrow [0, \infty)$ be measurable, X be independent of $\mathcal{F}_0$ and Y be $\mathcal{F}_0$ -measurable. Then*

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

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

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

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

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

For general F the statement now follows from a monotone class argument. $\square$

2.2. Martingales. In the subsequent chapters we will consider risky assets in a multi-period model. Their values at time $t \in \{0, \dots, T\}$ will be modelled by a stochastic process $(S_t)_{t=0, \dots, T}$, which is a collection of random variables.

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

First we need to introduce the notion of a filtration.

Definition 2.18. A filtration is $(\mathcal{F}_t)_{t \in I}$ is an increasing family of σ -algebras, that is $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for all $s, t \in I$ with $s < t$.

Typical choices for the index set I are $[0, \infty)$, $[0, T]$, $\mathbb{N}$ or $\{0, \dots, T\}$. In almost all situations the index t represents time. Then the σ -algebra $\mathcal{F}_t$ contains all the events that are observable up to time t , so $\mathcal{F}_t$ models the information available at time t . A stochastic process $S = (S_t)_{t=0, 1, \dots}$ naturally induces a filtration defined via

$\mathcal{F}_t = \sigma(S_0, \dots, S_t)$. If S models asset prices as in example mentioned above then $\mathcal{F}_t$ represents the information about all prices up to time t .

Recall that for any random variable Y the σ -algebra $\sigma(Y)$ is the smallest σ -algebra such that $\{Y \leq c\} \in \sigma(Y)$ for all $c \in \mathbb{R}$. Similarly, $\sigma(S_0, \dots, S_t)$ is the smallest σ -algebra such that $\{S_0 \leq c_0, \dots, S_T \leq c_T\} \in \sigma(S_0, \dots, S_t)$ for all $c_0, \dots, c_T \in \mathbb{R}$.

Example 2.19. Consider the simple symmetric random walk S on $\mathbb{Z}$ started from $S_0 := 0$, that is $S_t = \sum_{k=1}^t Z_k$, $t \geq 1$, where $(Z_k)_{k \geq 1}$ are i.i.d. random variable with $\mathbb{P}[Z_k = 1] = \mathbb{P}[Z_k = -1] = \frac{1}{2}$. Then, $\mathcal{F}_t = \sigma(S_1, \dots, S_t) = \sigma(Z_1, \dots, Z_t)$, $t \geq 1$, defines a filtration and

$$\{S_1 \leq 0, S_3 \geq 2\} \in \mathcal{F}_3 \quad \text{but} \quad \{S_4 > 0\} \notin \mathcal{F}_3.$$

Since $S_0 = 0$ is deterministic, $\sigma(S_0, \dots, S_t)$ and we could have started the filtration with the trivial σ -algebra $\mathcal{F}_0 = \sigma(S_0) = \{\emptyset, \Omega\}$.

Definition 2.20. A stochastic process $Y = (Y_t)_{t \in I}$ is said to be *adapted* to a filtration $(\mathcal{F}_t)_{t \in I}$ if Y_t is $\mathcal{F}_t$ -measurable for all $t \in I$.

Now we define martingales.

Definition 2.21. Let $(\mathcal{F}_t)_{t=0, \dots, T}$ be a filtration on $(\Omega, \mathcal{F})$. A stochastic process $M = (M_t)_{t=0, \dots, T}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0, \dots, T}, \mathbb{P})$ is a *martingale* (or $\mathbb{P}$ -martingale) if and only if the following hold.

- (a) M is adapted, that is M_t is $\mathcal{F}_t$ -measurable for every t .
- (b) $M_t \in \mathcal{L}^1(\Omega, \mathbb{P})$, i.e. $\mathbb{E}[|M_t|] < \infty$ for every t .
- (c) The *martingale property* holds, i.e. for all $0 \leq s \leq t \leq T$,

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s, \quad \mathbb{P}\text{-a.s.}$$

If (a) and (b) hold, but instead of (c), it holds $\mathbb{E}[M_t | \mathcal{F}_s] \geq M_s$, respectively $\mathbb{E}[M_t | \mathcal{F}_s] \leq M_s$, then the process M is called a *sub-martingale*, respectively a *super-martingale*.

Remark 2.22. (i) Similarly one defines martingales $(M_t)_{t \in I}$ if the index set I is $[0, \infty)$, $[0, T]$ or $\mathbb{N}$.

(ii) If M is a martingale it holds $\mathbb{E}[M_t] = \mathbb{E}[M_s]$ for all $0 \leq s \leq t$, for a sub-martingale we have $\mathbb{E}[M_t] \geq \mathbb{E}[M_s]$, finally, for a super-martingale $\mathbb{E}[M_t] \leq \mathbb{E}[M_s]$.

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

$$\mathbb{E}[M_t - M_s | \mathcal{F}_s] = 0, \quad \mathbb{P}\text{-a.s.}, \quad \forall 0 \leq s \leq t \leq T,$$

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

(iv) In discrete time, that is $I = \{0, \dots, T\}$ (or $I = \mathbb{N}$), (c) is equivalent to

$$\mathbb{E}[M_{t+1} | \mathcal{F}_t] = M_t, \quad \mathbb{P}\text{-a.s.} \quad \forall 0 \leq t \leq T-1. \quad (2.4)$$

Warning: In continuous time, i.e. if I is $[0, \infty)$ or $[0, T]$, (2.4) is not sufficient for (c) to hold.

(v) If $I = \{0, \dots, T\}$, a martingale $M = (M_t)_{t=0, \dots, T}$ is determined by M_T via $M_t = \mathbb{E}[M_T | \mathcal{F}_t]$. Conversely, every $F \in \mathcal{L}^1(\Omega, \mathcal{F}_T, \mathbb{P})$ defines a martingale via

$$M_t := \mathbb{E}[F | \mathcal{F}_t], \quad t = 0, \dots, T.$$

Example 2.23. Let $Z_1, \dots, Z_T$ be independent random variables with $Z_k \in \mathcal{L}^1$ and $\mathbb{E}[Z_k] = 0$ for all $k = 1, \dots, T$. Set

$$M_0 := 0, \quad M_t := \sum_{k=1}^t Z_k, \quad t = 1, \dots, T,$$

and $\mathcal{F}_t := \sigma(M_0, \dots, M_t)$, $t = 0, \dots, T$. Obviously, M is adapted to $(\mathcal{F}_t)_t$ and $M_t \in \mathcal{L}^1$ for all $t = 0, \dots, T$. Further, for $t = 0, \dots, T-1$,

$$\mathbb{E}[M_{t+1} | \mathcal{F}_t] = \mathbb{E}[M_t + Z_{t+1} | \mathcal{F}_t] = M_t + \mathbb{E}[Z_{t+1} | \mathcal{F}_t] = M_t + \mathbb{E}[Z_{t+1}] = M_t,$$

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

Definition 2.24. A stochastic process $(C_n)_{n \geq 1}$ is called *previsible*³ with respect to a filtration $(\mathcal{F}_n)_{n \geq 0}$, if, for all $n \in \mathbb{N}$, C_n is $\mathcal{F}_{n-1}$ -measurable.

Proposition 2.25. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$ be a filtered probability space.

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

$$Y_n := \sum_{k=1}^n C_k (M_k - M_{k-1}), \quad Y_0 := 0,$$

is a martingale.

(ii) If M is a sub-martingale (or supermartingale) and $(C_n)_{n \geq 1}$ is a bounded previsible and non-negative, then Y is a sub-martingale (or super-martingale, respectively).

Proof. (i) Since C is bounded, $Y_n \in \mathcal{L}^1$ for all n . For all $k \leq n$ the random variables C_k , M_{k-1} and M_k are all $\mathcal{F}_n$ -measurable, so Y_n is $\mathcal{F}_n$ -measurable, which means that Y is adapted. Finally, for $n \geq 1$,

$$\begin{aligned} \mathbb{E}[Y_n - Y_{n-1} | \mathcal{F}_{n-1}] &= \mathbb{E}[C_n(M_n - M_{n-1}) | \mathcal{F}_{n-1}] \\ &= C_n \mathbb{E}[M_n - M_{n-1} | \mathcal{F}_{n-1}] = 0. \end{aligned} \quad (2.5)$$

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

(ii) Since C_n is now assumed to be non-negative, we have in the last step of (2.5) that $C_n \mathbb{E}[M_n - M_{n-1} | \mathcal{F}_{n-1}]$ is non-negative if M is a sub-martingale and non-positive if M is a super-martingale. $\square$

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

Remark 2.26. (i) Sometimes the process $(C_n)_{n \geq 1}$ represents a gambling strategy. If M models the price process of a share, then Y_n represents the wealth at time n .

(ii) Y is a discrete time version of the stochastic integral $\int C \, dM$.

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

Definition 2.27. Let $a < b \in \mathbb{R}$. We say that an *upcrossing* of $[a, b]$ occurs between times s and t , if

- (i) $X_s < a, X_t > b$,
- (ii) for all r such that $s < r < t$, $X_r \in [a, b]$.

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

$$C_1 := \mathbb{1}_{\{X_0 < a\}}, \quad C_n := \mathbb{1}_{\{C_{n-1}=1\}} \mathbb{1}_{\{X_{n-1} \leq b\}} + \mathbb{1}_{\{C_{n-1}=0\}} \mathbb{1}_{\{X_{n-1} < a\}}, \quad n \geq 2. \quad (2.6)$$

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

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

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

$$W_N \geq (b - a) U_N(X, [a, b]) - |a - X_N| \mathbb{1}_{\{X_N < a\}}, \quad (2.7)$$

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

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

Theorem 2.28 (Doob's upcrossing lemma). *Let X be a super-martingale. Then for any $a < b \in \mathbb{R}$,*

$$\mathbb{E} [U_N(X, [a, b])] \leq \frac{\mathbb{E} [(X_N - a)^-]}{b - a}. \quad (2.8)$$

Proof. The process $(C_n)_{n \geq 1}$ defined in (2.6) is obviously bounded, non-negative and previsible, so by Proposition 2.25 (ii) the wealth process $(W_n)_{n \geq 0}$ is a super-martingale with $W_0 = 0$. Therefore $\mathbb{E}[W_N] \leq 0$ and taking expectation in (2.7) gives (2.8). $\square$

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

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

Corollary 2.29. *Let $(X_n)_{n \geq 0}$ be an $\mathcal{L}^1$ -bounded super-martingale, i.e. $\sup_n \mathbb{E}[|X_n|] < \infty$. Then*

$$\mathbb{E}[U_\infty(X, [a, b])] \leq \frac{a + \sup_n \mathbb{E}[|X_n|]}{b - a} < \infty. \quad (2.9)$$

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

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

Proof. This follows directly from Theorem 2.28 and the monotone convergence theorem since $\sup_n \mathbb{E}[(X_n - a)^-] \leq a + \sup_n \mathbb{E}[|X_n|]$. $\square$

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

Theorem 2.30 (Doob's super-martingale convergence theorem). *Let $(X_n)_{n \geq 0}$ be an $\mathcal{L}^1$ -bounded super-martingale. Then, $\mathbb{P}$ -a.s., $X_\infty := \lim_{n \rightarrow \infty} X_n$ exists and is a finite random variable.*

Proof. Define

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

But

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

Therefore, by Corollary 2.29, $\mathbb{P}[\Lambda_{a, b}] = 0$, and thus also

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

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

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

So X_∞ is almost surely finite. $\square$

Remark 2.31. Doobs convergence theorem implies that positive super-martingale always converge a.s. This is because the super-martingale property ensures in this case that $\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \leq \mathbb{E}[X_0]$, so the uniform boundedness in $\mathcal{L}^1$ is always guaranteed.

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

Definition 2.32. A map $\tau : \Omega \rightarrow \{0, 1, \dots\} \cup \{\infty\}$ is called a *stopping time* (with respect to a filtration $(\mathcal{F}_n)_{n \geq 0}$) if $\{\tau \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$ or, equivalently, $\{\tau = n\} \in \mathcal{F}_n$ for all $n \geq 0$.

Example 2.33. The most important examples of stopping times are hitting times. Let $(X_n)_{n \geq 0}$ be an adapted process, and let $B \in \mathcal{B}(\mathbb{R})$. Define

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

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

Definition 2.34. Let $(X_n)_{n \geq 0}$ be a stochastic process and τ be a stopping time. We define the *stopped process* X^τ via

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

Proposition 2.35. Let $(X_n)_{n \geq 0}$ be a (sub-)martingale and τ be a stopping time. Then the stopped process X^τ is a (sub-)martingale.

Proof. Exercise! □

Theorem 2.36 (Doob's Optional stopping theorem). Let $(X_n)_{n \geq 0}$ be a martingale and τ be a stopping time. Then, $X_\tau \in \mathcal{L}^1$ and

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

if one of the following conditions holds.

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

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

Proof. By Proposition 2.35 the stopped process $X_n^\tau = X_{\tau \wedge n}$ is a martingale. In particular, its expected value is constant in n , so that

$$\mathbb{E}[X_{\tau \wedge n}] = \mathbb{E}[X_n^\tau] = \mathbb{E}[X_0^\tau] = \mathbb{E}[X_0]. \quad (2.10)$$

Consider the case (a). By assumption τ is bounded, so there exists $N \in \mathbb{N}$ such that $\tau(\omega) \leq N$ for all $\omega \in \Omega$. Then, choosing $n = N$ in (2.10) gives the claim.

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

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

which together with (2.10) implies the result.

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

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

and by assumption $\mathbb{E}[K\tau] < \infty$. So again by the dominated convergence theorem we can pass to the limit in (2.10). $\square$

Remark 2.37. (i) By similar arguments Theorem 2.36 extends immediately to super- resp. submartingales in which case the conclusion reads

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

(ii) Theorem 2.36 may look strange and contradict the 'no strategy' idea. Take a simple random walk $(S_n)_{n \geq 0}$ on $\mathbb{Z}$ (i.e. a series of fair games), and define a stopping time $\tau = \inf\{n : S_n = 10\}$. Then clearly $\mathbb{E}[S_\tau] = 10 \neq \mathbb{E}[S_0] = 0$! So we conclude, using (c), that $\mathbb{E}[\tau] = +\infty$. In fact, the 'sure' gain if we achieve our goal is offset by the fact that on average, it takes infinitely long to reach it (of course, most games will end quickly, but chances are that some may take very very long!).

3. ARBITRAGE THEORY IN DISCRETE TIME

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

Definition 3.1. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$.

(i) We say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$ if

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

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

(ii) We say that $\mathbb{Q}$ is *equivalent* to $\mathbb{P}$ if $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, that is

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

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

Theorem 3.2. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$.

- (i) *Radon-Nikodym theorem:* $\mathbb{Q} \ll \mathbb{P}$ if and only if there exists an $\mathcal{F}$ -measurable function $f \geq 0$ with $\mathbb{E}_{\mathbb{P}}[f] = 1$ such that

$$\mathbb{Q}[A] = \int_A f \, d\mathbb{P} = \mathbb{E}_{\mathbb{P}}[f \mathbb{1}_A], \quad \forall A \in \mathcal{F}.$$

The function f is called density or Radon-Nikodym derivative and is often denoted by $f = \frac{d\mathbb{Q}}{d\mathbb{P}}$.

- (ii) $\mathbb{Q} \approx \mathbb{P}$ if and only if $\frac{d\mathbb{Q}}{d\mathbb{P}} > 0$ $\mathbb{P}$ -a.s. In this case we have

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

- (iii) If $\mathbb{Q} \ll \mathbb{P}$ and $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{Q})$ then

$$\mathbb{E}_{\mathbb{Q}}[Y] = \mathbb{E}_{\mathbb{P}}[fY].$$

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

Definition 3.3. We say that a portfolio $\bar{\theta} \in \mathbb{R}^{d+1}$ is an *arbitrage opportunity* if $\bar{\theta} \cdot \bar{S}_0 \leq 0$ but $\bar{\theta} \cdot \bar{S}_1 \geq 0$ $\mathbb{P}$ -a.s. and $\mathbb{P}[\bar{\theta} \cdot \bar{S}_1 > 0] > 0$.

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

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

Theorem 3.4. Assume that $S_t^0 = 1$ for all $t \in \{0, 1\}$. Then the following are equivalent.

- (i) *There is no arbitrage.*
(ii) *There exists a probability measure $\mathbb{Q} \approx \mathbb{P}$ such that*

$$\mathbb{E}_{\mathbb{Q}}[S_1] = S_0.$$

In this case we may take a density of the form

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{1}{Z_{\theta}} \exp \left(-\theta(S_1 - S_0) - \frac{1}{2}|S_1 - S_0|^2 \right)$$

for some $\theta \in \mathbb{R}^d$, where $Z_{\theta} := \mathbb{E} \left[\exp(-\theta(S_1 - S_0) - \frac{1}{2}|S_1 - S_0|^2) \right]$ is a normalising constant.

Proof. Write $Y := S_1 - S_0$ for abbreviation. Without loss of generality we may assume that $\mathbb{P}[\theta \cdot Y = 0] < 1$ for all $\theta \in \mathbb{R}^d \setminus \{0\}$. Otherwise we would have linear dependencies in the model and some assets would be redundant.

(ii) $\Rightarrow$ (i): Assume that $\bar{\theta}$ is an arbitrage. By condition (ii) we have $\mathbb{E}_{\mathbb{Q}}[\bar{\theta} \cdot S_1] = \bar{\theta} \cdot S_0$, and since $S^0 \equiv 1$ this yields $\mathbb{E}_{\mathbb{Q}}[\bar{\theta} \cdot \bar{S}_1] = \bar{\theta} \cdot \bar{S}_0$. Then, by the fact that $\bar{\theta}$ is an arbitrage we get

$$0 \leq \mathbb{E}_{\mathbb{Q}}[\bar{\theta} \cdot \bar{S}_1] = \bar{\theta} \cdot \bar{S}_0 \leq 0,$$

so $\mathbb{E}_{\mathbb{Q}}[\bar{\theta} \cdot \bar{S}_1] = 0$. Further, $\bar{\theta} \cdot \bar{S}_1 \geq 0$ $\mathbb{P}$ -a.s., and, since $\mathbb{Q} \approx \mathbb{P}$, we also have $\bar{\theta} \cdot \bar{S}_1 \geq 0$ $\mathbb{Q}$ -a.s. Hence, $\bar{\theta} \cdot \bar{S}_1 = 0$ $\mathbb{Q}$ -a.s. and therefore also $\bar{\theta} \cdot \bar{S}_1 = 0$ $\mathbb{P}$ -a.s., so $\bar{\theta}$ is not an arbitrage.

(i) $\Rightarrow$ (ii): Consider the function

$$\varphi : \mathbb{R}^d \rightarrow [0, \infty) \quad \theta \mapsto \frac{\mathbb{E}[\exp(-\theta \cdot Y - \frac{1}{2}|Y|^2)]}{\mathbb{E}[\exp(-\frac{1}{2}|Y|^2)]},$$

which is continuous, convex and differentiable.

Let us first consider the case that the infimum $\inf_{\theta} \varphi(\theta)$ is attained at some θ^* . Then we get by differentiating

$$\mathbb{E}[\exp(-\theta^* \cdot Y - \frac{1}{2}|Y|^2) Y] = 0,$$

and, defining $\mathbb{Q}$ via $\mathbb{Q}[A] = \int_A \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P}$ with $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{1}{Z_{\theta^*}} \exp(-\theta^* \cdot Y - \frac{1}{2}|Y|^2)$, this can be rewritten as $\mathbb{E}_{\mathbb{Q}}[Y] = 0$, so we obtain (ii).

It suffices to show now that if there is no arbitrage in the model then the infimum $\inf_{\theta} \varphi(\theta)$ is attained. Let us assume that $\inf_{\theta} \varphi(\theta)$ is *not* attained. Consider the sets

$$F_{\alpha} := \{\theta \in \mathbb{R}^d : |\theta| = 1, \varphi(\alpha\theta) \leq 1\}, \quad \alpha \geq 0,$$

which are compact subsets of $\mathbb{R}^d$. Since φ is convex and $\varphi(0) = 1$ we have $F_{\beta} \subseteq F_{\alpha}$ for $\alpha \leq \beta$. By the Finite Intersection Property⁴ either the intersection $\bigcap_{\alpha \geq 0} F_{\alpha}$ is non-empty, or for some $\alpha > 0$, $F_{\alpha} = \emptyset$. By assumption $\inf_{\theta} \varphi(\theta)$ is not attained, so $\inf_{\theta} \varphi(\theta) < \varphi(0) = 1$ and there exists a sequence $(a_k)_{k \geq 0}$ in $\mathbb{R}^d$ such that $\varphi(a_k) \downarrow \inf_{\theta} \varphi(\theta)$ as $k \uparrow \infty$. Then (a_k) cannot be bounded, because otherwise a subsequence would converge to a point where the infimum is attained. Thus, there exists a sequence of points tending to infinity where φ is less than 1. In particular, $F_{\alpha} = \emptyset$ for some $\alpha > 0$ cannot hold. Therefore, $\bigcap_{\alpha \geq 0} F_{\alpha} \neq \emptyset$, so there exists $a \in \mathbb{R}^d$ with $|a| = 1$ such that

$$\varphi(ta) = \frac{\mathbb{E}[\exp(-ta \cdot Y - \frac{1}{2}|Y|^2)]}{\mathbb{E}[\exp(-\frac{1}{2}|Y|^2)]} \leq 1, \quad \text{for all } t \geq 0.$$

⁴We say that a collection of subsets $\mathcal{A}$ of a Hausdorff space $\mathcal{X}$ (in our context $\mathcal{X}$ is the unit sphere $S^{d-1} \subset \mathbb{R}^d$) has the *Finite Intersection Property* if for any finite selection $\{A_1, \dots, A_N\} \subset \mathcal{A}$ we have $\bigcap_{i=1}^N A_i \neq \emptyset$.

Theorem: $\mathcal{X}$ is compact if and only if every collection of closed subsets satisfying the Finite Intersection Property has non-empty intersection.

But this can only happen if

$$\mathbb{P}[a \cdot Y < 0] = 0,$$

and since we assumed that $\mathbb{P}[a \cdot Y = 0] < 1$ this implies

$$\mathbb{P}[a \cdot (S_1 - S_0) > 0] > 0.$$

Now choose the strategy $\bar{a} = (-a \cdot S_0, a) \in \mathbb{R} \times \mathbb{R}^d$, that is take a portfolio consisting of $a \in \mathbb{R}^d$ units in the risky assets $S^1, \dots, S^d$ and $-a \cdot S_0$ units in the riskless asset S^0 . Then at time $t = 0$ this is worth $\bar{a} \cdot \bar{S}_0 = 0$ (recall that $S_0 \equiv 1$) and at time $t = 1$ its value is

$$\bar{a} \cdot \bar{S}_1 = -a \cdot S_0 + a \cdot S_1 = a \cdot (S_1 - S_0) \begin{cases} \geq 0 & \mathbb{P}\text{-a.s.} \\ > 0 & \text{with positive } \mathbb{P}\text{-probability.} \end{cases}$$

Thus, $\bar{a}$ is an arbitrage opportunity, which contradicts condition (i). $\square$

The assumption that S^0 is identically 1 is restrictive, asymmetric and unnecessary; the notion of arbitrage for any vector $\bar{S} \in \mathbb{R}^{d+1}$ of assets does not require this, and in fact we can deduce a far more flexible form of the above result.

Corollary 3.5 (Fundamental Theorem of Asset Pricing (FTAP)). *Assume that $S_t^0 > 0$ for all $t \in \{0, 1\}$. Then the following are equivalent.*

- (i) *There is no arbitrage.*
- (ii) *There exists a probability measure $\mathbb{Q} \approx \mathbb{P}$ such that*

$$\mathbb{E}_{\mathbb{Q}} \left[\frac{S_1}{S_1^0} \right] = \frac{S_0}{S_0^0}.$$

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

Proof. Note that $\bar{\theta}$ is an arbitrage for $\bar{S}$ if and only if it is an arbitrage for $\tilde{S}$ defined by $\tilde{S}_t^i := S_t^i / S_t^0$ for $i \in \{0, 1, \dots, d\}$, $t \in \{0, 1\}$. The result follows by applying Theorem 3.4 to $\tilde{S}$. $\square$

Remark 3.6. (i) The strictly positive asset S^0 above is referred to as a *numéraire*. We have often considered a situation where there is a single riskless asset (referred to variously as the money-market account, the bond, the bank account, ...) in the market, and it is very common to use this asset as numéraire. It turns out that this will serve for our present applications, but there are occasions when it is advantageous to use other numéraires. Note that the Fundamental Theorem of Asset Pricing does not require the existence of a riskless asset, that is S_1^0 could also be random as long as it is strictly positive.

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

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

Definition 3.7. A trading strategy is an $\mathbb{R}^{d+1}$ -valued, previsible process $\bar{\theta} = (\theta^0, \theta) = (\theta_t^0, \dots, \theta_t^d)_{t=1, \dots, T}$, i.e. $\bar{\theta}_t$ is $\mathcal{F}_{t-1}$ -measurable for all $t = 1, \dots, T$.

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

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

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

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

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

Definition 3.8. A trading strategy $\bar{\theta}$ is called *self-financing* if

$$\bar{\theta}_t \cdot \bar{S}_t = \bar{\theta}_{t+1} \cdot \bar{S}_t, \quad \forall t = 1, \dots, T - 1. \quad (3.1)$$

Intuitively, (3.1) means that the value of the portfolio at any time t equals the amount invested at time t . It follows that the accumulated gains and losses resulting from the price fluctuations are the only source of variations of the portfolio:

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

and summing up yields

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

Here, the constant $\bar{\theta}_1 \cdot \bar{S}_0$ can be interpreted as the initial investment for the purchase of the portfolio $\bar{\theta}_1$, while the second term may be regarded as a discrete stochastic integral (cf. Proposition 2.25).

Example 3.9. Often it is assumed that asset 0 plays the role of a *locally riskless bond*. In this case, one takes $S_0^0 \equiv 1$ and one lets S_t^0 evolve according to spot rate $r_t \geq 0$. At time t , an investment x made at time $t - 1$ yields the payoff $x(1 + r_t)$. Thus, a unit investment made at time $t = 0$ produces the value $S_t^0 = \prod_{k=1}^t (1 + r_k)$

at time t . In order for an investment in S^0 to be 'locally riskless' the spot rate r_t has to be known beforehand at time $t - 1$. In other words, the process $(r_t)_{t=1,\dots,T}$ and therefore also $(S_t^0)_{t=1,\dots,T}$ need to be previsible.

Without assuming previsibility as in the previous example, we assume from now on that

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

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

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

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

Definition 3.10. The (discounted) *value process* $V = (V_t)_{t=0,\dots,T}$ of a trading strategy $\bar{\theta}$ is given by

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

Proposition 3.11. For a trading strategy $\bar{\theta}$ the following are equivalent.

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

Proof. By dividing both sides of (3.1) by S_t^0 we easily see that (i) and (ii) are equivalent. Moreover, (ii) holds if and only if

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

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

Remark 3.12. If $\bar{\theta}$ is a self-financing trading strategy, then $(\bar{\theta}_{t+1} - \bar{\theta}_t) \cdot \bar{X}_t = 0$ for all $t = 1, \dots, T - 1$. In particular, the numéraire component satisfies

$$\theta_{t+1}^0 - \theta_t^0 = -(\theta_{t+1} - \theta_t) \cdot X_t, \quad t = 1, \dots, T - 1, \quad (3.2)$$

and

$$\theta_1^0 = V_0 - \theta_1 \cdot X_0. \quad (3.3)$$

Thus, the entire process θ^0 is determined by the initial investment V_0 and the d -dimensional process θ . Consequently, if a value $V_0 \in \mathbb{R}$ and any d -dimensional previsible process θ are given, we can define the process θ^0 via (3.2) and (3.3) to obtain a self-financing strategy $\bar{\theta} := (\theta^0, \theta)$ with initial capital V_0 , and this construction is unique.

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

Definition 3.13. A self-financing strategy $\bar{\theta}$ is called an *arbitrage opportunity* if its value process V satisfies

$$V_0 \leq 0, \quad V_T \geq 0 \quad \mathbb{P}\text{-a.s.}, \quad \text{and} \quad \mathbb{P}[V_T > 0] > 0.$$

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

Definition 3.14. A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ is called an *equivalent martingale measure* if $\mathbb{Q} \approx \mathbb{P}$ and the discounted price process X is a d -dimensional martingale. The set of all equivalent martingale measures is denoted by $\mathcal{P}$.

Proposition 3.15. Let $\mathbb{Q} \in \mathcal{P}$ and $\bar{\theta}$ be a self-financing strategy with value process V satisfying $V_T \geq 0$ $\mathbb{P}$ -a.s. Then V is a $\mathbb{Q}$ -martingale and $\mathbb{E}_{\mathbb{Q}}[V_T] = V_0$.

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

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

so that

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

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

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

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

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

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

For any $c > 0$ let θ_t^c be defined via $\theta_t^c := \mathbb{1}_{\{|\theta_t| \leq c\}} \theta_t$. Then $\mathbb{E}_{\mathbb{Q}}[V_{t-1} \mathbb{1}_{\{|\theta_t| \leq c\}} | \mathcal{F}_{t-1}]$ is well defined since

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

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

$$V_{t-1} \mathbb{1}_{\{|\theta_t| \leq c\}} = \mathbb{E}_{\mathbb{Q}}[V_{t-1} \mathbb{1}_{\{|\theta_t| \leq c\}} | \mathcal{F}_{t-1}] \geq -\mathbb{E}_{\mathbb{Q}}[\theta_t^c \cdot (X_t - X_{t-1}) | \mathcal{F}_{t-1}] = 0.$$

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

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

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

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

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

Step 4. Since we have assumed that $\mathcal{F}_0$ is $\mathbb{P}$ -trivial, i.e. $\mathbb{P}[A] \in \{0, 1\}$ for all $A \in \mathcal{F}_0$, and $\mathbb{Q} \approx \mathbb{P}$, clearly $\mathcal{F}_0$ is also $\mathbb{Q}$ -trivial. Hence, by Proposition 2.12 and Step 3,

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

Moreover, we use repeatedly Step 3 to obtain

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

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

Theorem 3.16 (Fundamental Theorem of Asset Pricing, FTAP). *Assumed that S^0 is an a.s. strictly positive numéraire, i.e. $S_t^0 > 0$ $\mathbb{P}$ -a.s. for all $t = 0, \dots, T$. Then, the following are equivalent.*

- (i) *There is no arbitrage.*
- (ii) *$\mathcal{P} \neq \emptyset$, that is there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the discounted value process X defined by*

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

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

Proof. (ii) $\Rightarrow$ (i): Let $\mathbb{Q} \in \mathcal{P}$ and $\bar{\theta}$ be a self-financing strategy with a value process V satisfying $V_0 \leq 0$ and $V_T \geq 0$ $\mathbb{P}$ -a.s. Then, by Proposition 3.15,

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

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

(i) $\Rightarrow$ (ii): A nice proof in our current discrete time setting, which is based on an application of the Hahn-Banach separation theorem, can be found in [7, Theorem 5.17, Section 1.6]. In continuous time the proof is even much more complicated, see [6]. $\square$

3.3. European contingent claims.

Definition 3.17. A non-negative random variable C on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *European contingent claim* or *European option*. A European contingent claim C is called a *derivative* of the underlying assets $S^0, S^1, \dots, S^d$ if C is measurable with respect to the σ -algebra generated by the price process $(\bar{S}_t)_{t=0, \dots, T}$.

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

Example 3.18. (i) The owner of a *European call option* has the right, but not the obligation, to buy a unit of an asset, say asset i , at time T for a strike price K . The corresponding contingent is given by

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

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

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

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

$$S_{\text{av}}^i := \frac{1}{|\mathbb{T}|} \sum_{i \in \mathbb{T}} S_t^i$$

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

- Average price call: $(S_{\text{av}}^i - K)^+$,
- Average price put: $(K - S_{\text{av}}^i)^+$,
- Average strike call: $(S_T^i - S_{\text{av}}^i)^+$,
- Average strike put: $(S_{\text{av}}^i - S_T^i)^+$.

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

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

$$C_{\text{uo}}^{\text{call}} = \begin{cases} (S_T^i - K)^+ & \text{if } \max_{0 \leq t \leq T} S_t^i < B, \\ 0 & \text{else.} \end{cases}$$

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

$$C_{\text{di}}^{\text{put}} = \begin{cases} (K - S_T^i)^+ & \text{if } \min_{0 \leq t \leq T} S_t^i < B, \\ 0 & \text{else.} \end{cases}$$

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

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

$$S_T^i - \min_{0 \leq t \leq T} S_t^i$$

and a *lookback put*

$$\max_{0 \leq t \leq T} S_t^i - S_T^i.$$

Definition 3.19. A European contingent claim C is called *attainable* (*replicable*, *redundant*) if there exists a self-financing strategy $\bar{\theta}$ whose terminal portfolio coincides with C , i.e.

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

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

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

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

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

$$H = \frac{\theta_T \cdot \bar{S}_T}{S_T^0} = \theta_T \cdot \bar{X}_T = V_T = V_0 + \sum_{t=1}^T \theta_t \cdot (X_t - X_{t-1}), \quad (3.4)$$

where V denotes the value process of the replicating strategy $\bar{\theta} = (\theta^0, \theta)$ (cf. Proposition 3.11). In this case, we will also say that the discounted claim H is attainable with replicating strategy $\bar{\theta}$.

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

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

Theorem 3.20. *Let H be an attainable discounted claim. Then*

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

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

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

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

Proof. From (3.4) we see that $V_T = H \geq 0$. Then, by Proposition 3.15 the value process V is a $\mathbb{Q}$ -martingale for any $\mathbb{Q} \in \mathcal{P}$, so $V_t = \mathbb{E}_{\mathbb{Q}}[V_T | \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[H | \mathcal{F}_t]$. $\square$

Remark 3.21. The last result has two remarkable implications. First, note that $\mathbb{E}_{\mathbb{Q}}[H | \mathcal{F}_t]$ is independent of the replicating strategy, so all replicating strategies have the same value process. Further, for any $t = 0, \dots, T$, since $V_t = \bar{\theta}_t \cdot \bar{X}_t$ is independent of the equivalent martingale measure $\mathbb{Q}$, V_t is a version of $\mathbb{E}_{\mathbb{Q}}[H | \mathcal{F}_t]$ for all $\mathbb{Q} \in \mathcal{P}$.

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

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

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

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

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

Remark 3.22. Theorem 3.20 provides not only the price of an attainable claim (i.e. its value at time $t = 0$), but also its value at any time $t \in \{0, \dots, T\}$, which is given by $V_t = \mathbb{E}_{\mathbb{Q}}[H | \mathcal{F}_t]$, which equals the value of a replicating strategy at time t .

We have now discussed pricing market models that are complete in the following sense.

Definition 3.23. An arbitrage-free market model is called *complete* if every European contingent claim is attainable.

Theorem 3.24 (Second Fundamental Theorem of Asset Pricing). *An arbitrage-free market model is complete if and only if there exists exactly one equivalent martingale measure, i.e. $|\mathcal{P}| = 1$.*

Proof. See [7, Section 5.4]. □

In the next section we will study the prototype of a complete market model, the Cox-Ross-Rubinstein binomial model. However, in discrete time only a very limited class of models turn out to be complete. For incomplete models pricing is more difficult (see e.g. [7, Theorem 5.30]).

4. THE COX-ROSS-RUBINSTEIN BINOMIAL MODEL

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

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

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

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

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

$$R_t = \frac{S_t^1 - S_{t-1}^1}{S_{t-1}^1}$$

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

We now construct the model on the sample space

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

Denote by

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

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

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

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

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

where the initial value $S_0^1 > 0$ is a given constant. The discounted value process is given by

$$X_t = \frac{S_t^1}{S_t^0} = S_0^1 \prod_{k=1}^t \frac{1 + R_k}{1 + r}.$$

As a filtration we take

$$\mathcal{F}_t := \sigma(S_0^1, \dots, S_t^1) = \sigma(X_0, \dots, X_t), \quad t = 0, \dots, T.$$

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

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

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

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

or, in other words,

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

Definition 4.1. This model is called *binomial model* or *CRR model* (for Cox, Ross, Rubinstein).

Theorem 4.2. *The CRR model is arbitrage-free if and only if $a < r < b$. In this case, there exists a unique equivalent martingale measure $\mathbb{Q}$, i.e. $\mathcal{P} = \{\mathbb{Q}\}$, and $\mathbb{Q}$ is characterised by the fact that the random variables $R_1, \dots, R_T$ are independent under $\mathbb{Q}$ with common distribution*

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

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

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

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

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

This can be rewritten as

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

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

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

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

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

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

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

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

Proposition 4.3. *The value process*

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

of a replicating strategy for H is of the form

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

where the function v_t is given by

$$v_t(x_0, \dots, x_t) = \mathbb{E}_{\mathbb{Q}} \left[h \left(x_0, \dots, x_t, x_t \cdot \frac{S_1^1}{S_0^1}, \dots, x_t \cdot \frac{S_{T-t}^1}{S_0^1} \right) \right]. \quad (4.1)$$

Proof. Since the equivalent martingale measure is unique, the model is complete and the every contingent claim is attainable, and by Theorem 3.20 the value process of any replicating strategy is given by $V_t = \mathbb{E}_{\mathbb{Q}}[H | \mathcal{F}_t]$. Thus,

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

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

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

The claim follows now from Proposition 2.17. □

Since the value process V is characterised by the recursion

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

we obtain the following recursion formula for the function v_t in Proposition 4.3,

$$v_T(x_0, \dots, x_T) = h(x_0, \dots, x_T),$$

and for $t < T$,

$$\begin{aligned} v_t(x_0, \dots, x_t) &= p^* v_{t+1}(x_0, x_1, \dots, x_t, x_t(1+b)) + (1-p^*) v_{t+1}(x_0, x_1, \dots, x_t, x_t(1+a)). \end{aligned} \quad (4.2)$$

Indeed, for $t < T$,

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

Example 4.4. Suppose that $H = h(S_T^1)$ only depends on the terminal value S_T^1 of the stock price, then V_t depends only on the value S_t^1 of the stock at time t , i.e. $V_t(\omega) = v_t(S_t^1(\omega))$ and the formula (4.1) reduces to

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

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

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

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

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

Proposition 4.5. *The hedging strategy is given by*

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

where

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

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

Proof. By Proposition 3.11 we have that for each $\omega = (y_1, \dots, y_T) \in \{-1, 1\}^T$ any self-financing strategy $\bar{\theta}$ must satisfy

$$\theta_t(\omega) \cdot (X_t(\omega) - X_{t-1}(\omega)) = V_t(\omega) - V_{t-1}(\omega). \quad (4.3)$$

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

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

Plugging ω^+ and ω^- into (4.3) gives

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

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

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

and the claim follows. $\square$

The recursion formula (4.2) can be used for the numeric computation of the value process of a contingent claim. Nevertheless, for the value process of certain exotic options which depend on the maximum of the stock price, it is possible to get some quite explicit analytic formulas if we make the additional assumption that

$$(1+a)(1+b) = 1. \quad (4.4)$$

Further, as the price formulas only depend on the equivalent martingale measure $\mathbb{Q}$, of course, and not on the original measure $\mathbb{P}$, we may now introduce as an auxiliary probability measure the uniform distribution

$$\mathbb{P}[\{\omega\}] := \frac{1}{|\Omega|} = 2^{-T}, \quad \forall \omega \in \Omega = \{-1, 1\}^T.$$

We still denote $Y_t(\omega) := y_t$ for $\omega = (y_1, \dots, y_T)$ the projection onto the t -th coordinate of ω (also called coordinate process). Then, under the uniform distribution $\mathbb{P}$, we have that $Y_1, \dots, Y_T$ are i.i.d. random variables with $\mathbb{P}[Y_t = 1] = \mathbb{P}[Y_t = -1] = \frac{1}{2}$. Hence, the stochastic process Z defined by

$$Z_0 := 0, \quad Z_t := Y_1 + \dots + Y_t, \quad t = 1, \dots, T,$$

is a symmetric simple random walk on $\mathbb{Z}$. Further, due to (4.4) the price process of the risky asset can be written as

$$S_t^1(\omega) = S_0^1 \prod_{k=1}^t (1 + R_k(\omega)) = S_0^1 (1+b)^{Z_t(\omega)}.$$

Moreover, by path counting we have that

$$\mathbb{P}[Z_t = k] = \begin{cases} 2^{-t} \binom{t}{\frac{t+k}{2}} & \text{if } t+k \text{ is even,} \\ 0 & \text{else.} \end{cases} \quad (4.5)$$

The next result is the key to numerous explicit results on the distribution of the one-dimensional simple random walk Z . For its statement, it will be convenient to

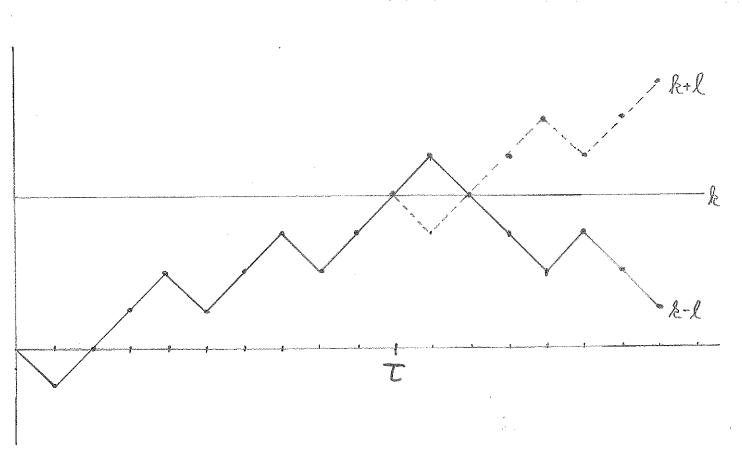


FIGURE 2. The reflection principle for the simple random walk

assume that Z is defined up to time $T+1$, which always can be achieved by enlarging the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 4.6 (Reflection principle for simple random walk). *For all $k \geq 1$ and $l \geq 0$,*

$$\mathbb{P} \left[\max_{0 \leq t \leq T} Z_t \geq k \text{ and } Z_T = k - l \right] = \mathbb{P} [Z_T = k + l] \quad (4.6)$$

and

$$\mathbb{P} \left[\max_{0 \leq t \leq T} Z_t = k \text{ and } Z_T = k - l \right] = \frac{2(k+l+1)}{T+1} \mathbb{P} [Z_{T+1} = 1 + k + l]. \quad (4.7)$$

Proof. The proof is based on the fact that the uniform distribution $\mathbb{P}$ is preserved under bijections between paths. Define

$$\tau(\omega) := \inf \{t \geq 0 \mid Z_t(\omega) = k\},$$

and

$$A_{k,l} := \{\omega \in \Omega \mid \tau(\omega) \leq T, Z_T(\omega) = k - l\}.$$

For $\omega = (y_1, \dots, y_T) \in A_{k,l}$ set

$$\phi(\omega) := (y_1, \dots, y_{\tau(\omega)}, -y_{\tau(\omega)+1}, \dots, -y_T).$$

Intuitively, for $\omega \in A_{k,l}$ the two trajectories $(Z_t(\omega))_{t=0, \dots, T}$ and $(Z_t(\phi(\omega)))_{t=0, \dots, T}$ coincide up to time $\tau(\omega)$, but from then on the latter path is obtained by reflecting the original one on the horizontal axis at level k (see Figure 2). In particular, note that

$$\phi : A_{k,l} \rightarrow \{Z_T = k + l, \tau(\omega) \leq T\} \stackrel{l \geq 0}{=} \{Z_T = k + l\}$$

is a bijection. Therefore,

$$\mathbb{P}[A_{k,l}] = \frac{|A_{k,l}|}{|\Omega|} = \frac{|\phi(A_{k,l})|}{|\Omega|} = \mathbb{P} [Z_T = k + l],$$

so we obtain (4.6).

In order to show (4.7) we first observe that it trivial as both sides become zero if $T + k + l$ is odd. Otherwise, we set $j = (T + k + l)/2$ and use (4.6) and (4.5) to obtain that

$$\begin{aligned}
& \mathbb{P} \left[\max_{0 \leq t \leq T} Z_t = k, Z_T = k - l \right] \\
&= \mathbb{P} \left[\max_{0 \leq t \leq T} Z_t \geq k, Z_T = k - l \right] - \mathbb{P} \left[\max_{0 \leq t \leq T} Z_t \geq k + 1, Z_T = k - l \right] \\
&= \mathbb{P} [Z_T = k + l] - \mathbb{P} [Z_T = k + l + 2] \\
&= 2^{-T} \binom{T}{j} - 2^{-T} \binom{T}{j+1} = 2^{-T} \binom{T+1}{j+1} \frac{2j+1-T}{T+1} \\
&= \frac{2(k+l+1)}{T+1} \mathbb{P} [Z_{T+1} = 1 + k + l],
\end{aligned}$$

which is the claim. $\square$

Next we observe that the density of the equivalent martingale measure $\mathbb{Q}$ with respect to $\mathbb{P}$ is a function of the terminal value of the random walk Z . This will be the key for applying the reflection principle to the problem of pricing exotic options involving the maximum of the stock price.

Lemma 4.7. *The density of $\mathbb{Q}$ with respect to $\mathbb{P}$ is given by*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = 2^T (p^*)^{\frac{T+Z_T}{2}} (1-p^*)^{\frac{T-Z_T}{2}}.$$

Remark 4.8. The proof of Lemma 4.7 will be based on the following fact. Let Ω be a countable set, $\mathcal{F}$ be the power set of Ω and P be any probability measure on $(\Omega, \mathcal{F})$ with $P[\{\omega\}] > 0$ for all $\omega \in \Omega$. Then every probability measure Q on $(\Omega, \mathcal{F})$ is absolutely continuous with respect to P and the density is given by

$$\frac{dQ}{dP}(\omega) = \phi(\omega) := \frac{Q[\{\omega\}]}{P[\{\omega\}]}, \quad \omega \in \Omega.$$

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

$$\begin{aligned}
Q[A] &= \sum_{\omega \in A} Q[\{\omega\}] = \sum_{\omega \in A} \frac{Q[\{\omega\}]}{P[\{\omega\}]} P[\{\omega\}] = \sum_{\omega \in A} \phi(\omega) P[\{\omega\}] \\
&= \mathbb{E}_P [\phi \mathbb{1}_A] = \int_A \phi dP.
\end{aligned}$$

Proof of Lemma 4.7. For each $\omega = (y_1, \dots, y_T) \in \Omega$, which contains exactly k components with $y_i = +1$,

$$\mathbb{Q}[\{\omega\}] = (p^*)^k (1-p^*)^{T-k}.$$

But for such an ω we have $Z_T(\omega) = k - (T - k) = 2k - T$, so $k = (Z_T(\omega) + T)/2$. Since $\mathbb{P}[\{\omega\}] = 2^{-T}$, the result follows from the previous remark. $\square$

Example 4.9 (Up-and-in call option). Consider an up-and-in call of the form

$$C_{\text{ui}}^{\text{call}} = \begin{cases} (S_T^1 - K)^+ & \text{if } \max_{0 \leq t \leq T} S_t^1 \geq B, \\ 0 & \text{else,} \end{cases}$$

where $B > S_0^1 \vee K$ denotes the barrier and $K > 0$ the strike price. We may assume without loss of generality that B lies within the range of possible asset prices, i.e. $B = S_0^1(1+b)^k$ for some $k \in \mathbb{N}$. Our aim is to calculate the arbitrage-free price

$$\pi(C_{\text{ui}}^{\text{call}}) = \frac{1}{(1+r)^T} \mathbb{E}_{\mathbb{Q}} [C_{\text{ui}}^{\text{call}}].$$

First recall that $S_T^1 = S_0^1(1+b)^{Z_T}$. Further, notice that Z_T can only take values $-T, 2-T, 4-T, \dots, T$. Thus,

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}} [C_{\text{ui}}^{\text{call}}] &= \mathbb{E}_{\mathbb{Q}} \left[(S_T^1 - K)^+ \mathbb{1}_{\{\max_{0 \leq t \leq T} S_t^1 \geq B\}} \right] \\ &= \sum_{l=0}^T (S_0^1(1+b)^{2l-T} - K)^+ \mathbb{Q} \left[\max_{0 \leq t \leq T} Z_t \geq k, Z_T = 2l - T \right] \\ &= \sum_{l=0}^T (S_0^1(1+b)^{2l-T} - K)^+ 2^T (p^*)^l (1-p^*)^{T-l} \mathbb{P} \left[\max_{0 \leq t \leq T} Z_t \geq k, Z_T = 2l - T \right], \end{aligned}$$

where we used Lemma 4.7 in the last step. Denote by l_k the largest integer l such that $2l - T \leq k$. Then, for $l \leq l_k$ trivially $2l - T = k - j$ for $j := k + T - 2l \geq 0$, and we may use the reflection principle in Theorem 4.6 and (4.5) to obtain

$$\begin{aligned} \mathbb{P} \left[\max_{0 \leq t \leq T} Z_t \geq k, Z_T = 2l - T \right] &= \mathbb{P} [Z_T = k + j] = \mathbb{P} [Z_T = 2(k - l) + T] \\ &= \begin{cases} 2^{-T} \binom{T}{T+k-l} & \text{if } l \geq k, \\ 0 & \text{else.} \end{cases} \end{aligned}$$

On the other hand, if $l > l_k$ then $2l - T > k$ and therefore

$$\mathbb{P} \left[\max_{0 \leq t \leq T} Z_t \geq k, Z_T = 2l - T \right] = \mathbb{P} [Z_T = 2l - T] = 2^{-T} \binom{T}{l}.$$

Hence, by combining the previous three equations we finally get

$$\begin{aligned} \pi(C_{\text{ui}}^{\text{call}}) &= \frac{1}{(1+r)^T} \left[\sum_{l=k}^{l_k} (S_0^1(1+b)^{2l-T} - K)^+ (p^*)^l (1-p^*)^{T-l} \binom{T}{T+k-l} \right. \\ &\quad \left. + \sum_{l=l_k+1}^T (S_0^1(1+b)^{2l-T} - K)^+ (p^*)^l (1-p^*)^{T-l} \binom{T}{l} \right]. \end{aligned}$$

Similarly, one can obtain formulas for barrier options with a lower stock price barrier such as down-and-out put options or down-and-in calls. Lookback options can be handled in this manner, too (see exercises).

5. DYNAMIC PROGRAMMING

See Section 5 in [8] or Section 1 in [14].

6. BROWNIAN MOTION

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

It might be thought that we can now operate entirely with our sophisticated Brownian motion model, and forget about the much simpler binomial model, but this is far from being the case. Usually, the pricing PDE which arises for a given derivative must be solved numerically, and this requires us to discretise the PDE in some way. To do this, we can either discretise the derivatives in the PDE using some finite-difference approximation and then solve the resulting system of linear equations, or we can approximate the Brownian motion process by some random walk, and then solve the pricing problem for that process by the usual dynamic-programming methodology. The first computes an approximation to the solution to the problem we wanted to solve; the second computes the exact solution to an approximating problem. These slightly different points of view are both valuable; for basic pricing problems, the PDE technology is faster and more accurate, but if the problem is more complicated, the second approach is more robust.

We finish these motivating remarks with a very short overview about the history of Brownian motion.

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

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

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

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

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

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

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

6.1. Definition and basic properties.

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

- (a) $B_0 = 0$, $\mathbb{P}$ -a.s.
- (b) For $\mathbb{P}$ -a.e. ω , the map $t \mapsto B_t(\omega)$ is continuous.
- (c) For any $n \in \mathbb{N}$ and any $0 = t_0 < t_1 < \dots < t_n$, the increments $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ are independent and each increment $B_{t_i} - B_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})$, so it is a Gaussian random variable with mean zero and variance $t_i - t_{i-1}$.

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

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

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

$$\begin{aligned} & \mathbb{P}[B_{t_1} \in A_1, \dots, B_{t_n} \in A_n] \\ &= \int_{A_1} \int_{A_2} \dots \int_{A_n} p_{t_1}(0, x_1) p_{t_2-t_1}(x_1, x_2) \dots p_{t_n-t_{n-1}}(x_{n-1}, x_n) dx_n \dots dx_1 \end{aligned}$$

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

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

$$\mathbb{P}_x[B_t \in dy] = p_t(x, y) dy.$$

(iii) Brownian motion is a martingale with respect to its natural filtration defined by $\mathcal{F}_t = \sigma(B_s, s \leq t)$. Indeed, for $0 \leq s \leq t$,

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

where we used that the increment $B_t - B_s$ is independent of $\mathcal{F}_s$.

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

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

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

Alternatively, we could describe Brownian motion as follows.

Lemma 6.3. *Brownian motion is the Gaussian process⁵ $(B_t)_{t \geq 0}$ with values in $\mathbb{R}$ such that*

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

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

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

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

To prove the converse, i.e. that any process with the properties given in the statement is a Brownian motion, we can just use the fact that the law of a Gaussian process is uniquely determined by its mean and covariance (see e.g. [2, Section 3]). Thus the process has the same law as Brownian motion and has only continuous paths (by (b)), so it is Brownian motion. $\square$

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

⁵A stochastic process $(X_t)_{t \geq 0}$ is called a *Gaussian process* if for any $n \in \mathbb{N}$ and any $0 < t_1 < \dots < t_n$ the vector $(X_{t_1}, \dots, X_{t_n})$ is normally distributed.

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

The question remains whether such a process actually exists.

Theorem 6.5. *Brownian motion exists.*

We will not give a formal proof here; some nice short constructions of Brownian motion can be found, for instance, in [2, Section 6] or [11, Section 7]. However, the maybe most natural approach would require a good amount of preparation, so we only sketch the main idea here.

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

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

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

$$(X^{(n)})_{t \geq 0} \xrightarrow[n \rightarrow \infty]{\Rightarrow} (B_t)_{t \geq 0}$$

where B is a Brownian motion (and $\Rightarrow$ denotes convergence in distribution). If we would just consider the sequence $X_1^{(n)}$, i.e. the case $t = 1$, this is exactly the Central Limit Theorem, since $B_1 \sim \mathcal{N}(0, 1)$. But Donsker's theorem is in fact a much stronger result as it provides such a convergence *simultaneously for all* $t \geq 0$. Therefore the theorem is also called *Functional Central Limit Theorem*. Equivalently, this result could be also formulated as follows. The rescaled random walk $X^{(n)}$ and the Brownian motion B may be regarded as random variables taking values in the *path space* $C([0, \infty), \mathbb{R})$. Then Donsker's theorem says that the distribution $\mathbb{P} \circ (X^{(n)})^{-1}$ of the entire path of the rescaled random walk, which is a measure on $C([0, \infty), \mathbb{R})$, converges weakly to the distribution of Brownian motion, which is also called the *Wiener measure*.

As a consequence, on an intuitive level Brownian motion looks locally like a random walk on a very large time scale, so the paths are very rough. The next proposition provides a more precise statement. We recall that a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is locally Hölder continuous of order α for $\alpha \in [0, 1]$ if, for every $L > 0$,

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

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

Proposition 6.6. *Let B be a Brownian motion. Then, almost surely,*

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

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

Proof. See e.g. [11, Theorem 7.7.2]. $\square$

Proposition 6.7. *Let B be a Brownian motion. Then each of the following processes are also Brownian motions.*

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

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

$$\text{cov}(B_s^3, B_t^3) = \mathbb{E}[B_s^3 B_t^3] = st \mathbb{E}[B_{1/s} B_{1/t}] = st \frac{1}{t} = s, \quad 0 < s \leq t.$$

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

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

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

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

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

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

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

Corollary 6.8 (Law of large numbers). *Let B be a Brownian motion. Then,*

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

Proof. Let B^3 be as in Proposition 6.7, then

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

□

Proposition 6.9. *Let B be a Brownian motion. Then,*

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

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

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

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

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

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

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

6.2. The Reflection principle. In Proposition 4.6 we have seen already the reflection principle for a simple random walk. As Brownian motion may be regarded as a pendant of the simple random walk in continuous time, as analogous statement also holds for Brownian motion, which we will discuss in this section.

We will need the fact that Brownian motion enjoys the so-called *strong Markov property*, i.e. the Markov property can also be applied on random times provided they are stopping times, so we can split expectations between past and future also at stopping times.

To be more specific, for any stopping time τ with respect to the filtration generated by a Brownian motion B on $(\Omega, \mathcal{F}, \mathbb{P})$ we define

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, \forall t \geq 0\},$$

which is the σ -algebra of events observable up time τ .

Theorem 6.10 (Strong Markov property). *Let $(B_t)_{t \geq 0}$ be an $(\mathcal{F}_t)_{t \geq 0}$ -Brownian motion and let τ be a stopping time such that $\tau < \infty$ $\mathbb{P}$ -a.s. Then $(B_{\tau+t})_{t \geq 0}$ is an $(\mathcal{F}_{\tau+t})_{t \geq 0}$ -Brownian motion. In particular, $(B_{\tau+t} - B_\tau)_{t \geq 0}$ is a Brownian motion independent of $\mathcal{F}_\tau$ (cf. Proposition 6.7 for the special case $\tau = T$ constant).*

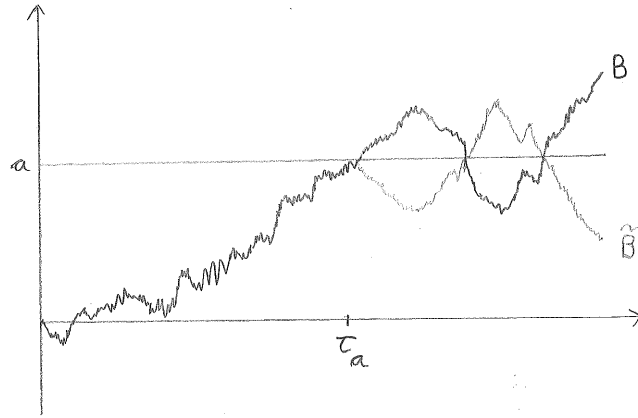


FIGURE 3. The reflection principle for Brownian motion

Proof. See [11, Theorem 7.5.1]. □

For any $a > 0$ let

$$\tau_a := \inf \{t \geq 0 : B_t = a\}$$

denote the first hitting time of a . Then τ_a is a stopping time which is almost surely finite by Proposition 6.9.

Theorem 6.11. Let $(B_t)_{t \geq 0}$ be a Brownian motion starting from 0 and let $a > 0$. Define

$$\tilde{B}_t := \begin{cases} B_t, & \text{if } t < \tau_a, \\ 2a - B_t, & \text{if } t \geq \tau_a. \end{cases}$$

Then $(\tilde{B}_t)_{t \geq 0}$ is also a Brownian motion starting from 0.

Proof. Set $Y_t := B_t$ for $t \leq \tau_a$ and $Z_t = B_{\tau_a+t} - B_{\tau_a} = B_{\tau_a+t} - a$ for $t \geq 0$. By the strong Markov property in Theorem 6.10 Z is a Brownian motion independent of Y . Of course, $-Z$ is also a Brownian motion (and also independent of Y) and therefore (Y, Z) and $(Y, -Z)$ have the same distribution.

For any time $t_0 > 0$ let $\Phi_{t_0} : C([0, t_0]) \times C([0, \infty)) \rightarrow C([0, \infty))$ be defined via

$$(y, z) \mapsto (\Phi_{t_0}(y, z))_{t \geq 0} \quad \text{with} \quad \Phi_{t_0}(y, z)_t := \begin{cases} y_t & \text{if } t \leq t_0, \\ y_{t_0} + z_{t-t_0} & \text{if } t > t_0. \end{cases}$$

Intuitively, the mapping Φ_{t_0} takes two continuous paths y and z and glues them together at time t_0 , provided both paths y and z are starting at zero.

In particular, notice that $\Phi_{\tau_a}(Y, Z)$ and $\Phi_{\tau_a}(Y, -Z)$ have the same distribution. But $\Phi_{\tau_a}(Y, Z) = B$ and $\Phi_{\tau_a}(Y, -Z) = \tilde{B}$ (see Figure 3). □

Corollary 6.12. *Let $(B_t)_{t \geq 0}$ be a Brownian motion. Then, for any $a, y \geq 0$ and $t \geq 0$,*

$$\mathbb{P} \left[\max_{0 \leq s \leq t} B_s \geq a, B_t \leq a - y \right] = \mathbb{P} [B_t \geq a + y].$$

Proof. Using Theorem 6.11 we have

$$\begin{aligned} \mathbb{P} \left[\max_{0 \leq s \leq t} B_s \geq a, B_t \leq a - y \right] &= \mathbb{P} \left[\max_{0 \leq s \leq t} \tilde{B}_s \geq a, \tilde{B}_t \leq a - y \right] \\ &= \mathbb{P} \left[\max_{0 \leq s \leq t} B_s \geq a, B_t \geq a + y \right] = \mathbb{P} [B_t \geq a + y]. \end{aligned}$$

Here we used in the second step that after τ_a the paths of $\tilde{B}$ are obtained from the paths of B by reflection at the horizontal axis at level a (see Figure 3). $\square$

As an immediate application of the reflection principle we compute the Laplace transform of τ_a .

Corollary 6.13. *For any $a > 0$ and $\lambda > 0$,*

$$\mathbb{E} [e^{-\lambda \tau_a}] = e^{-a\sqrt{2\lambda}}.$$

Proof. By using Corollary 6.12 and fact that $B_t \sim \mathcal{N}(0, t)$ for any $t > 0$ we have for any $a > 0$,

$$\begin{aligned} \mathbb{P}[\tau_a \leq t] &= \mathbb{P} \left[\max_{0 \leq s \leq t} B_s \geq a \right] \\ &= \mathbb{P} \left[\max_{0 \leq s \leq t} B_s \geq a, B_t \leq a \right] + \mathbb{P} \left[\max_{0 \leq s \leq t} B_s \geq a, B_t \geq a \right] \\ &= 2 \mathbb{P}[B_t \geq a] = 2 \left(1 - \Phi\left(\frac{a}{\sqrt{t}}\right) \right), \end{aligned}$$

where Φ denotes the distribution function of the standard normal distribution. Since the right hand side is differentiable in t , we see that the density of the random variable τ_a (with respect to Lebesgue measure) is given by

$$\mathbb{P}[\tau_a \in dt] = -\partial_t \left[2\Phi\left(\frac{a}{\sqrt{t}}\right) \right] = -2\Phi'\left(\frac{a}{\sqrt{t}}\right) \left(-\frac{a}{2\sqrt{t^3}} \right) = \frac{a}{\sqrt{2\pi t^3}} e^{-a^2/2t}.$$

Hence,

$$\mathbb{E} [e^{-\lambda \tau_a}] = \int_0^\infty e^{-\lambda t} \frac{a}{\sqrt{2\pi t^3}} e^{-a^2/2t} dt = e^{-a\sqrt{2\lambda}},$$

where the last step can be confirmed by a (tedious) direct computation, which we leave as an exercise. The statement can also be proved by using the optional stopping theorem (exercise). $\square$

6.3. Change of measure: The Cameron-Martin theorem. In this section it will be convenient to specify the underlying probability space, similarly as we did in Section 4. Let $\Omega = C([0, T])$ the *path space* of continuous function on $[0, T]$ and denote by B the *coordinate process*, that is $B_t(\omega) = \omega_t$ for $t \in [0, T]$ and $\omega \in \Omega$. We endow Ω with the σ -algebra $\mathcal{F} = \sigma(B_t, 0 \leq t \leq T)$, which can be shown to coincide with the Borel σ -algebra on $C([0, T])$ (with respect to the topology induced by the uniform convergence on $[0, T]$). Finally, let $\mathbb{P}$ be the probability measure on $(\Omega, \mathcal{F})$

under which the coordinate process B is a Brownian motion. This measure is known as *Wiener measure* and its existence follows along with the existence of Brownian motion from Donsker's invariance principle, for instance (cf. the discussion below Theorem 6.5).

Consider now a *Brownian motion with drift at speed c* , that is

$$B_t + ct, \quad 0 \leq t \leq T,$$

for any $c \in \mathbb{R}$. Recall that the transition density of the Brownian motion B is given by

$$p_t(x, y) := p_t(x - y) := \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y - x)^2}{2t}\right), \quad t > 0, x, y \in \mathbb{R}.$$

We now compute the finite dimensional distributions of the process $(B_t + ct)_{t \in [0, T]}$. For any $0 = t_0 < t_1 < \dots < t_n = T$ and any $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned} \mathbb{P}[B_{t_1} + ct_1 \in A_1, \dots, B_{t_n} + ct_n \in A_n] \\ &= \int_{A_1 - ct_1} \dots \int_{A_n - ct_n} p_{t_1}(x_1) p_{t_2 - t_1}(x_2 - x_1) \dots p_{t_n - t_{n-1}}(x_n - x_{n-1}) dx_n \dots dx_1 \\ &= \int_{A_1} \dots \int_{A_n} p_{t_1}(y_1 - ct_1) p_{t_2 - t_1}(y_2 - y_1 - c(t_2 - t_1)) \dots \\ &\quad \times p_{t_n - t_{n-1}}(y_n - y_{n-1} - c(t_n - t_{n-1})) dy_n \dots dy_1, \end{aligned}$$

where we used the substitution $y_i = x_i + ct_i$. Since

$$\begin{aligned} p_{t_i - t_{i-1}}(y_i - y_{i-1} - c(t_i - t_{i-1})) \\ &= p_{t_i - t_{i-1}}(y_i - y_{i-1}) \exp\left(c(y_i - y_{i-1}) - \frac{c^2}{2}(t_i - t_{i-1})\right) \end{aligned}$$

and $\sum_{i=1}^n t_i - t_{i-1} = t_n = T$ and $\sum_{i=1}^n y_i - y_{i-1} = y_n$ with $y_0 := 0$, this becomes

$$\begin{aligned} \mathbb{P}[B_{t_1} + ct_1 \in A_1, \dots, B_{t_n} + ct_n \in A_n] \\ &= \int_{A_1} \dots \int_{A_n} p_{t_1}(y_1) p_{t_2 - t_1}(y_2 - y_1) \dots \\ &\quad \times p_{t_n - t_{n-1}}(y_n - y_{n-1}) \exp\left(cy_n - \frac{c^2}{2}t_n\right) dy_n \dots dy_1 \\ &= \mathbb{E}\left[\mathbb{1}_{\{B_{t_1} \in A_1, \dots, B_{t_n} \in A_n\}} \exp\left(cB_{t_n} - \frac{c^2}{2}t_n\right)\right] \end{aligned}$$

with $t_n = T$. We have just shown that any *cylindrical functional* F , that is a functional $F : C([0, T]) \rightarrow \mathbb{R}$ of the form

$$F(\omega) = \begin{cases} 1 & \text{if } \omega_{t_1} \in A_1, \dots, \omega_{t_n} \in A_n, \\ 0 & \text{else,} \end{cases}$$

satisfies

$$\mathbb{E}[F(B_t + ct : 0 \leq t \leq T)] = \mathbb{E}\left[F(B_t : 0 \leq t \leq T) \exp\left(cB_T - \frac{c^2}{2}T\right)\right].$$

By linearity and approximation arguments this can be extended to all bounded and measurable $F : C([0, T]) \rightarrow \mathbb{R}$. Choosing $F = \mathbb{1}_A$ for any $A \in \mathcal{F}$ we get

$$\mathbb{P}^{(c)}[A] := \mathbb{P} \left[(B_t + ct : 0 \leq t \leq T) \in A \right] = \int_A \exp \left(cB_T - \frac{c^2}{2}T \right) d\mathbb{P}.$$

Thus, the measures $\mathbb{P}^{(c)}$ and $\mathbb{P}$ are equivalent with Radon-Nikodym density given by

$$\frac{d\mathbb{P}^{(c)}}{d\mathbb{P}} = \exp \left(cB_T - \frac{c^2}{2}T \right).$$

To summarize, under the Wiener measure $\mathbb{P}$ the paths in $C(0, T]$ have the distribution of a Brownian motion while under the measure $\mathbb{P}^{(c)}$ the paths in $C(0, T]$ have the distribution of a Brownian motion with drift c . We have arrived at

Theorem 6.14 (Cameron-Martin theorem). *For any $c \in \mathbb{R}$, $T > 0$ and any bounded and measurable $F : C([0, T]) \rightarrow \mathbb{R}$,*

$$\begin{aligned} \mathbb{E} \left[F(B_t + ct : 0 \leq t \leq T) \right] &= \mathbb{E} \left[F(B_t : 0 \leq t \leq T) \exp \left(cB_T - \frac{c^2}{2}T \right) \right] \\ &= \mathbb{E}_{\mathbb{P}^{(c)}} \left[F(B_t : 0 \leq t \leq T) \right]. \end{aligned}$$

For later use we establish the following consequence from the Cameron-Martin theorem and the reflection principle. We denote by $\mathbb{P}_x$ the probability measure on the path space $\Omega = C([0, T])$, under which the coordinate process B is a Brownian motion starting at x and by $\mathbb{E}_x$ the associated expectation operator.

Lemma 6.15. *For any $c \in \mathbb{R}$ and $T > 0$ define $\tilde{B}_t = B_t + ct$, $t \in [0, T]$. Then, for any $x, y > a$,*

$$\mathbb{P}_x \left[\tilde{B}_T \in dy, \min_{0 \leq s \leq T} \tilde{B}_s \geq a \right] = e^{c(y-x) - \frac{c^2}{2}T} (p_T(x, y) - p_T(x, 2a - y)) dy,$$

where $p_T(x, y)$ still denotes the transition probabilities as defined in (6.1).

Proof. Note that by the reflection principle

$$\mathbb{P}_x \left[B_T \geq y, \min_{0 \leq s \leq T} B_s < a \right] = \mathbb{P}_x \left[B_T \leq 2a - y \right],$$

so that

$$\mathbb{P}_x \left[B_T \geq y, \min_{0 \leq s \leq T} B_s \geq a \right] = \mathbb{P}_x \left[B_T \geq y \right] - \mathbb{P}_x \left[B_T \leq 2a - y \right].$$

Differentiating with respect to y gives

$$\mathbb{P}_x \left[B_T \in dy, \min_{0 \leq s \leq T} B_s \geq a \right] = (p_T(x, y) - p_T(x, 2a - y)) dy,$$

By the Cameron-Martin theorem we have for any bounded and measurable functional $F : C([0, T]) \rightarrow \mathbb{R}$,

$$\mathbb{E} \left[F(\tilde{B}_s : 0 \leq s \leq T) \right] = \mathbb{E} \left[F(B_s : 0 \leq s \leq T) \exp \left(c(B_T - B_0) - \frac{c^2}{2}T \right) \right].$$

Choosing now $F(\omega) = f(\omega_T) \mathbb{1}_{\{\min_{s \leq T} \omega_s \geq a\}}$ for any bounded, measurable $f : \mathbb{R} \rightarrow \mathbb{R}$, this implies

$$\begin{aligned} & \int_{\mathbb{R}} f(y) \mathbb{P}_x [\tilde{B}_T \in dy, \min_{0 \leq s \leq T} \tilde{B}_s \geq a] \\ &= \int_{\mathbb{R}} f(y) e^{c(y-x) - \frac{c^2}{2}T} \mathbb{P}_x [B_T \in dy, \min_{0 \leq s \leq T} B_s \geq a] \\ &= \int_{\mathbb{R}} f(y) e^{c(y-x) - \frac{c^2}{2}T} (p_T(x, y) - p_T(x, 2a - y)) dy, \end{aligned}$$

which is the claim. $\square$

6.4. Martingales associated with Brownian motion. Let B still be a Brownian motion. We have seen that the processes $(B_t^2 - t)_{t \geq 0}$ and $(\exp(\lambda B_t - \frac{\lambda^2}{2}t))_{t \geq 0}$, $\lambda \in \mathbb{R}$, are martingales (see exercises). Both processes are of the form $f(t, B_t)$ with $f(t, y) = y^2 - t$ and $f(t, y) = \exp(\lambda y - \frac{\lambda^2}{2}t)$, respectively. It is not a coincidence that both functions satisfy the partial differential equation

$$\mathcal{L}f = 0,$$

where

$$\mathcal{L}f := \partial_t f + \frac{1}{2} \partial_{yy} f.$$

In fact, the next result shows that both processes are just two examples of a much larger class of martingales.

Theorem 6.16. *Let $f \in C_b^{1,2}([0, \infty) \times \mathbb{R})$ and let $(B_t)_{t \geq 0}$ be a Brownian motion. Then, the process $(M_t)_{t \geq 0}$ defined by*

$$M_t := f(t, B_t) - f(0, B_0) - \int_0^t \mathcal{L}f(s, B_s) ds, \quad t \geq 0,$$

is a continuous martingale.

Proof. The following proof is taken from [11, Theorem 7.4.4]. It is straightforward to see that $(M_t)_{t \geq 0}$ is continuous, adapted and integrable. It remains to show, for $s, t \geq 0$, that

$$\mathbb{E} [M_{s+t} - M_s | \mathcal{F}_s] = 0, \quad \mathbb{P}\text{-a.s.}$$

Fix $s \geq 0$ and set

$$\tilde{f}(t, x) := f(s+t, x), \quad \tilde{B}_t := B_{s+t}, \quad \tilde{\mathcal{F}}_t = \mathcal{F}_{s+t}.$$

Then $\tilde{B}$ is an $(\tilde{\mathcal{F}}_t)_{t \geq 0}$ -Brownian motion starting at $\tilde{B}_0 = B_s$ and $M_{s+t} - M_s = \tilde{M}_t$, where

$$\begin{aligned} \tilde{M}_t &= f(s+t, B_{s+t}) - f(s, B_s) - \int_s^{s+t} \mathcal{L}f(r, B_r) dr \\ &= \tilde{f}(t, \tilde{B}_t) - \tilde{f}(0, \tilde{B}_0) - \int_0^t \mathcal{L}\tilde{f}(r, \tilde{B}_r) dr. \end{aligned}$$

We have to show $\mathbb{E}[\tilde{M}_t | \tilde{\mathcal{F}}_0] = 0$ almost surely. Since this is the same problem for all $s \geq 0$, it will suffice to show that $\mathbb{E}[M_t | \mathcal{F}_0] = 0$ almost surely. Now $\mathbb{E}[M_t | \mathcal{F}_0] = m(B_0)$ almost surely, where $m(x) = \mathbb{E}_x[M_t]$ and the subscript x specifies the case $B_0 = x$. So it will suffice to show that $\mathbb{E}_x[M_t] = 0$ for all $x \in \mathbb{R}$.

Now $\mathbb{E}_x[M_s] \rightarrow 0$ as $s \rightarrow 0$, so it will suffice to show that $\mathbb{E}_x[M_t - M_s] = 0$ for all $x \in \mathbb{R}$ and all $0 < s < t$. We compute

$$\begin{aligned} \mathbb{E}_x[M_t - M_s] &= \mathbb{E}_x \left[f(t, B_t) - f(s, B_s) - \int_s^t \mathcal{L}f(r, B_r) dr \right] \\ &= \mathbb{E}_x[f(t, B_t)] - \mathbb{E}_x[f(s, B_s)] - \mathbb{E}_x \left[\int_s^t (\partial_t + \frac{1}{2} \partial_{yy}) f(r, B_r) dr \right] \\ &= \mathbb{E}_x[f(t, B_t)] - \mathbb{E}_x[f(s, B_s)] - \int_s^t \int_{\mathbb{R}} p_r(x, y) \partial_t f(r, y) dy dr \\ &\quad - \frac{1}{2} \int_s^t \int_{\mathbb{R}} p_r(x, y) \partial_{yy} f(r, y) dy dr. \end{aligned}$$

Note that, for any $x \in \mathbb{R}$, $p_r(x, \cdot)$ satisfies the heat equation $\partial_t p = \frac{1}{2} \partial_{yy} p$. Thus, on integrating by parts with respect to time,

$$\begin{aligned} \int_s^t \int_{\mathbb{R}} p_r(x, y) \partial_t f(r, y) dy dr &= \int_{\mathbb{R}} p_r(x, y) f(r, y) \Big|_{r=s}^t - \int_s^t \partial_t p_r(x, y) f(r, y) dr dy \\ &= \mathbb{E}_x[f(t, B_t)] - \mathbb{E}_x[f(s, B_s)] - \frac{1}{2} \int_s^t \int_{\mathbb{R}} \partial_{yy} p_r(x, y) f(r, y) dy dr, \end{aligned}$$

and by integrating by parts twice in $\mathbb{R}$ we obtain

$$\frac{1}{2} \int_s^t \int_{\mathbb{R}} p_r(x, y) \partial_{yy} f(r, y) dy dr = \frac{1}{2} \int_s^t \int_{\mathbb{R}} \partial_{yy} p_r(x, y) f(r, y) dy dr.$$

By combining the last three equalities we get $\mathbb{E}_x[M_t - M_s] = 0$ as required. $\square$

Remark 6.17. (i) The conditions of boundedness on f and its derivatives can be relaxed, while taking care that $(M_t)_{t \geq 0}$ remains integrable and the integrations by parts remain valid. There is a natural alternative proof via Itô's formula once one has access to stochastic calculus.

(ii) The same proof also works in arbitrary dimensions, so for a d -dimensional Brownian motion $B = (B^1, \dots, B^d)$ and any suitable function f the process

$$f(t, B_t) - f(0, B_0) - \int_0^t (\partial_t + \frac{1}{2} \Delta) f(r, B_r) dr$$

is a continuous martingale.

7. THE BLACK-SCHOLES MODEL

In 1965 Paul Samuelson proposed the following market model in continuous time. There is a riskless bond

$$S_t^0 = e^{rt}, \quad 0 \leq t \leq T,$$

with interest rate $r \geq 0$ and one risky asset with price process given

$$S_t = S_0 \exp(\sigma B_t + \mu t), \quad 0 \leq t \leq T,$$

where B denotes a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, $\mu \in \mathbb{R}$ a drift, $\sigma > 0$ a volatility parameter and $S_0 > 0$ the initial price of the asset.

Fisher Black and Myron Scholes (1973) and Robert Merton (1973) added the crucial replication argument which leads to a complete pricing and hedging theory. Merton and Scholes received the 1997 Nobel Memorial Prize in Economic Sciences for their work (sadly, Black was ineligible for the prize because of his death in 1995).

7.1. Black-Scholes via change of measure. Our goal is to determine the price π_C at time $t = 0$ of a contingent claim C dependent on the entire path $(S_t)_{t \in [0, T]}$ with maturity at time $t = T$. In analogy to the general results obtained in Section 3 for the discrete time-setting, we suppose that

$$\pi_C = \mathbb{E}_{\mathbb{Q}} \left[\frac{C}{e^{rT}} \right],$$

where $\mathbb{Q}$ is an equivalent martingale measure, that is $\mathbb{Q} \approx \mathbb{P}$ and the discounted price process

$$X_t = e^{-rt} S_t = S_0 \exp(\sigma B_t + (\mu - r)t), \quad 0 \leq t \leq T,$$

is a $\mathbb{Q}$ -martingale with respect to the natural filtration $(\mathcal{F}_t)_{t \geq 0}$ generated by $(S_t)_{t \geq 0}$. How can we find such a measure $\mathbb{Q}$? First, recall that for a Brownian motion W the process $\exp(\lambda W_t - \frac{\lambda^2}{2}t)$ is a martingale for every $\lambda \in \mathbb{R}$ (see exercises). We define the measure $\mathbb{Q}$ by $d\mathbb{Q} := \exp(cB_T - \frac{c^2}{2}T) d\mathbb{P}$, which is short for

$$\mathbb{Q}[A] := \int_A \exp(cB_T - \frac{c^2}{2}T) d\mathbb{P} = \mathbb{E} \left[\exp(cB_T - \frac{c^2}{2}T) \mathbb{1}_A \right],$$

for some $c \in \mathbb{R}$ to be chosen later. In particular, $\mathbb{Q}$ is equivalent to $\mathbb{P}$ since the density $\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp(cB_T - \frac{c^2}{2}T) > 0$ $\mathbb{P}$ -a.s. (see Theorem 3.2 (ii)).

By the Cameron-Martin theorem we have for any bounded and measurable functional $F : C([0, T]) \rightarrow \mathbb{R}$,

$$\begin{aligned} \mathbb{E} \left[F(B_t : 0 \leq t \leq T) \right] &= \mathbb{E} \left[F(W_t : 0 \leq t \leq T) \exp(cB_T - \frac{c^2}{2}T) \right] \\ &= \mathbb{E}_{\mathbb{Q}} \left[F(W_t : 0 \leq t \leq T) \right], \end{aligned}$$

with $W_t := B_t - ct$, $t \in [0, T]$. (We apply here Theorem 6.14 on the functional $\tilde{F}(\omega) = F((\omega_t - ct)_{t \leq T})$.) In particular, W is a Brownian motion under the measure $\mathbb{Q}$. Now we choose c such that

$$\sigma c + \mu - r = -\frac{\sigma^2}{2} \quad \Longleftrightarrow \quad c = \frac{r - \frac{\sigma^2}{2} - \mu}{\sigma}.$$

Hence,

$$\begin{aligned} X_t &= e^{-rt} S_t = S_0 \exp(\sigma B_t + (\mu - r)t) = S_0 \exp(\sigma W_t + (\sigma c + \mu - r)t) \\ &= S_0 \exp(\sigma W_t - \frac{\sigma^2}{2}t), \end{aligned}$$

which is a $\mathbb{Q}$ -martingale. Thus, $\mathbb{Q}$ is an equivalent martingale measure, which can also be shown to be unique.

To summarize, under $\mathbb{Q}$ the price process $(S_t)_{t \in [0, T]}$ is of the form

$$S_t = S_0 \exp\left(\sigma W_t + \left(r - \frac{1}{2}\sigma^2\right)t\right),$$

where W is a $\mathbb{Q}$ -Brownian motion. Note that for pricing of a contingent claim only the behaviour of the price process under the equivalent martingale measure is relevant.

Consider now, as an example, a European option of the form $C = f(S_T)$ with expiry $T > 0$ for any bounded, continuous function $f : [0, \infty) \rightarrow [0, \infty)$. Then the Black-Scholes price $\pi_C = e^{-rT} \mathbb{E}_{\mathbb{Q}}[f(S_T)]$ of C is given by

$$\pi_C = e^{-rT} \mathbb{E}_{\mathbb{Q}}\left[f\left(S_0 \exp\left(\sigma W_T + \left(r - \frac{1}{2}\sigma^2\right)T\right)\right)\right]. \quad (7.1)$$

Since $W_T \sim \mathcal{N}(0, T)$, so $W_T = \sqrt{T}Y$ with $Y \sim \mathcal{N}(0, 1)$, it follows that $\pi_C = v(T, S_0)$, where

$$v(t, x) := e^{-rt} \int_{-\infty}^{\infty} f\left(x \exp\left(\sigma\sqrt{t}y + \left(r - \frac{1}{2}\sigma^2\right)t\right)\right) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \quad (7.2)$$

for $t \in [0, T]$, $x \in \mathbb{R}$. Of course, for this argument to work, the payoff function f does not have to be continuous or bounded. It suffices that $\mathbb{E}_{\mathbb{Q}}[f(S_T)] < \infty$, which is already guaranteed if f has polynomial volume growth, that is there exist $c > 0$ and $p \geq 0$ such that $f(x) \leq c(1+x)^p$ for all $x \geq 0$. In particular, we can use formula (7.1) to compute the price of a call option with $f(x) = (x - K)^+$.

7.2. The Black-Scholes Model as limit of the Binomial Model. The Black-Scholes model also arises as a natural limit of certain binomial models after a suitable scaling, meaning that the number of intermediate trading periods becomes large and their durations becomes small. This should not come as big surprise. In the CRR model the price process is a random walk and in the Black-Scholes model it is governed by a Brownian motion, which can be obtained as scaling limit of random walks.

Throughout this section, T will *not* denote the number of trading periods in a fixed discrete-time market model but rather a physical *date*. We divide the interval $[0, T]$ into $N \cdot T$ equidistant time steps $\frac{1}{N}, \frac{2}{N}, \dots, \frac{NT}{N}$. Then the i -th trading period corresponds to the 'real time interval' $(\frac{i-1}{N}, \frac{i}{N})$. Now consider a family of multi-period CRR-models, indexed by $N \in \mathbb{N}$, with parameters

$$r_N := \frac{r}{N}, \quad a_N := -\frac{\sigma}{\sqrt{N}}, \quad b_N := \frac{\sigma}{\sqrt{N}}, \quad p_N := \frac{1}{2} + \frac{1}{2} \frac{\mu}{\sigma\sqrt{N}},$$

where $r \geq 0$ is the instantaneous interest rate, $\mu \in \mathbb{R}$ a drift and $\sigma > 0$ a volatility parameter. We denote by $(S_{i,N}^0)_{i=0,\dots,NT}$ the riskless bond and by $(S_{i,N}^1)_{i=0,\dots,NT}$ the risky asset. The initial prices are assumed not to depend on N , i.e. $S_{0,N}^1 = S_0^1$ for some constant $S_0^1 > 0$.

The question is whether the prices of contingent claims in the approximating market models converge as N tends to infinity. It will turn out that they do converge towards the Black-Scholes prices derived in the last section. First, we note that

$$S_{NT,N}^0 = (1 + r_N)^{NT} = \left(1 + \frac{r}{N}\right)^{NT} \xrightarrow{N \rightarrow \infty} e^{rT}, \quad (7.3)$$

and

$$S_{NT,N}^1 = S_0^1 \prod_{i=1}^{NT} (1 + R_i^{(N)}) = S_0^1 \exp \left(\sum_{i=1}^{NT} \log(1 + R_i^{(N)}) \right),$$

where

$$R_i^{(N)} = \frac{S_{i,N}^1 - S_{i-1,N}^1}{S_{i-1,N}^1}$$

denotes the return in the i -th trading period. For every N , the returns $(R_i^{(N)})_{i=1,\dots,NT}$ are i.i.d. random variables with

$$R_i^{(N)} = \begin{cases} \frac{\sigma}{\sqrt{N}} & \text{with probability } \frac{1}{2} + \frac{1}{2} \frac{\mu}{\sigma\sqrt{N}} = p_N, \\ -\frac{\sigma}{\sqrt{N}} & \text{with probability } \frac{1}{2} - \frac{1}{2} \frac{\mu}{\sigma\sqrt{N}} = 1 - p_N. \end{cases}$$

Since $\log(1 + r) = r - \frac{1}{2}r^2 + O(r^3)$ for $r \rightarrow 0$, we have for sufficiently large N ,

$$\log \left(\frac{S_{NT,N}^1}{S_0^1} \right) = \sum_{i=1}^{NT} \left(R_i^{(N)} - \frac{1}{2} (R_i^{(N)})^2 \right) + \varepsilon_{N,T} \quad (7.4)$$

where $|\varepsilon_{N,T}| \leq c/\sqrt{N}$ for some constant $c > 0$ not depending on N . Further, we compute

$$\mathbb{E} [R_i^{(N)}] = \frac{1}{2} \left(1 + \frac{\mu}{\sigma\sqrt{N}} \right) \frac{\sigma}{\sqrt{N}} + \frac{1}{2} \left(1 - \frac{\mu}{\sigma\sqrt{N}} \right) \left(-\frac{\sigma}{\sqrt{N}} \right) = \frac{\mu}{N}$$

and

$$\mathbb{E} \left[(R_i^{(N)})^2 \right] = \frac{\sigma^2}{N}, \quad \text{var} [R_i^{(N)}] = \frac{\sigma^2}{N} - \frac{\mu^2}{N^2},$$

so that

$$\begin{aligned} \mathbb{E} \left[R_i^{(N)} - \frac{1}{2} (R_i^{(N)})^2 \right] &= \frac{\mu - \frac{1}{2}\sigma^2}{N}, \\ \text{var} \left(R_i^{(N)} - \frac{1}{2} (R_i^{(N)})^2 \right) &= \text{var} \left(R_i^{(N)} - \frac{1}{2} \frac{\sigma^2}{N} \right) = \text{var} (R_i^{(N)}) = \frac{\sigma^2}{N} - \frac{\mu^2}{N^2}. \end{aligned} \quad (7.5)$$

We are aiming to take the limit $N \rightarrow \infty$ in the right hand side of (7.4), which consists of a sum of the i.i.d. random variables $R_i^{(N)} - \frac{1}{2} (R_i^{(N)})^2$ and a negligible remainder term $\varepsilon_{N,T}$. The idea is to apply the central limit theorem, but this is in

fact problematic and we need to be careful here, because (7.5) shows that mean and variance of the $R_i^{(N)} - \frac{1}{2}(R_i^{(N)})^2$ do depend on N . Such a situation is not covered by the classical central limit theorem, but the following more general version will be applicable here.

Theorem 7.1. *Suppose that for each $N \in \mathbb{N}$ we are given N independent random variables $Y_1^{(N)}, \dots, Y_N^{(N)}$ on a probability space $(\Omega, \mathcal{A}, P)$, which satisfies the following conditions.*

- (i) *There are constants γ_N such that $\gamma_N \rightarrow 0$ and $|Y_i^{(N)}| \leq \gamma_N$ P -a.s.*
- (ii) *$\sum_{i=1}^N E[Y_i^{(N)}] \rightarrow m$ for some $m \in \mathbb{R}$ as $N \rightarrow \infty$.*
- (iii) *$\sum_{i=1}^N \text{var}[Y_i^{(N)}] \rightarrow \sigma^2$ for some $\sigma > 0$ as $N \rightarrow \infty$.*

Then the distributions of

$$Z_N := \sum_{i=1}^N Y_i^{(N)}, \quad N = 1, 2, \dots,$$

converge weakly to the normal distribution with mean m and variance σ^2 .

Proof. See, for instance, the corollary to Theorem 7.1.2 of [3]. □

We apply Theorem 7.1 on the random variables $Y_i^{(N)} = R_i^{(N)} - \frac{1}{2}(R_i^{(N)})^2$ in (7.4). Indeed, condition (i) is satisfied as $|R_i^{(N)}| \leq \sigma/\sqrt{N}$ and conditions (ii) and (iii) are immediate from (7.5) with $m = (\mu - \frac{1}{2}\sigma^2)T$ and $\sigma^2 = \sigma^2 T$. Thus, $\log\left(\frac{S_{NT,N}^1}{S_0^1}\right)$ converges in distribution towards a $\mathcal{N}\left((\mu - \frac{1}{2}\sigma^2)T, \sigma^2 T\right)$ -distributed random variable. This implies that⁶

$$S_{NT,N}^1 \xrightarrow{d} S_0^1 \exp\left(\sigma B_T + \left(\mu - \frac{1}{2}\sigma^2\right)T\right) \quad \text{as } N \rightarrow \infty,$$

where B_T is a $\mathcal{N}(0, T)$ -distributed random variable.

However, for each N the prices of contingent claims in the approximating model are expectations under the equivalent martingale measure $\mathbb{Q}_N$, which by Theorem 4.2 is specified by

$$p_N^* = \frac{r_N - a_N}{b_N - a_N} = \frac{\frac{r}{N} + \frac{\sigma}{\sqrt{N}}}{2\frac{\sigma}{\sqrt{N}}} = \frac{1}{2} + \frac{1}{2} \frac{r}{\sigma\sqrt{N}}.$$

We can simply replace p_N by p_N^* in the above argument (or even simpler replace μ by r) to obtain that under the measure $\mathbb{Q}_N$,

$$S_{NT,N}^1 \xrightarrow{d} S_0^1 \exp\left(\sigma W_T + \left(r - \frac{1}{2}\sigma^2\right)T\right) \quad \text{as } N \rightarrow \infty, \quad (7.6)$$

where W_T is a $\mathcal{N}(0, T)$ -distributed random variable.

⁶We use the fact that if a sequence of random variable (X_n) converges in distribution to a random variable X , then for any continuous function f the sequence $(f(X_n))$ converges in distribution to $f(X)$. In the present case $f(x) = S_0^1 \exp(x)$.

Let us now consider a derivative which is defined in terms of a function f of the risky asset's terminal value. In each approximating model, this corresponds to a contingent claim

$$C^{(N)} = f(S_{NT,N}^1).$$

Theorem 7.2. *Let f be bounded and continuous. Then the limit of the arbitrage-free prices of $C^{(N)} = f(S_{NT,N}^1)$ for $N \rightarrow \infty$ is given by the Black-Scholes price $v(T, S_0^1)$, where as before*

$$v(T, x) := e^{-rT} \int_{-\infty}^{\infty} f\left(x \exp\left(\sigma\sqrt{T}y + \left(r - \frac{1}{2}\sigma^2\right)T\right)\right) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \quad x \in \mathbb{R}.$$

Proof. Since f is bounded and continuous, (7.3) and (7.6) immediately imply that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{\mathbb{Q}_N} \left[\frac{C^{(N)}}{(1 + r_N)^{NT}} \right] = \mathbb{E} \left[e^{-rT} f\left(S_0^1 \exp\left(\sigma\sqrt{T}W + \left(r - \frac{1}{2}\sigma^2\right)T\right)\right) \right],$$

where W is a $\mathcal{N}(0, 1)$ -distributed random variable. $\square$

Remark 7.3. (i) Under suitable conditions, a much stronger version of the convergence in (7.6) holds in form of a functional central limit theorem. Let us consider each discrete-time model as a *continuous* process $\tilde{S}^{(N)} = (\tilde{S}_t^{(N)})_{0 \leq t \leq T}$ defined as $\tilde{S}_t^{(N)} = S_{i,N}^1$ at the date $t = \frac{i}{N}$, and by linear interpolation in between. Then the laws of the processes $\tilde{S}^{(N)}$, considered as $C([0, T])$ -valued random variables, converge weakly towards the process $(S_t)_{0 \leq t \leq T}$ given by

$$S_t = S_0 \exp\left(\sigma W_t + \left(r - \frac{1}{2}\sigma^2\right)t\right),$$

where W is a Brownian motion. So the limit coincide with the price process in the Black-Scholes model under the equivalent martingale measure.

(ii) Again the assumption that f needs to be bounded in Theorem 7.2 is quite restrictive as it excludes the call option, for instance. It turns out that the assumption can be relaxed and that Theorem 7.2 also holds for continuous payoff functions f for which there exist $c > 0$ and $q \in [0, 2)$ such that $f(x) \leq c(1 + x)^q$ for all $x \geq 0$, see [7, Proposition 5.59].

Hedging in the Black-Scholes model. In the Black-Scholes model consider an attainable contingent claim of the form $C = f(S_T)$ with replicating strategy (or *hedging strategy*) $\bar{\theta} = (\theta^0, \theta)$. Then the (discounted) value process of $\bar{\theta}$ is given by $V_t = v(T - t, S_t)$ with v defined in (7.2). Since the Black-Scholes model may be regarded as a limit of binomial models in the sense of (7.6), Theorem 7.2 and Remark 7.3, in view of the hedging strategy for the CRR model derived in Proposition 4.5, one can argue that the hedging strategy is given by

$$\theta_t(\omega) = \Delta(T - t, S_t(\omega)), \quad \theta_t^0(\omega) = v(T - t, S_t(\omega)) - \theta_t(\omega)e^{-rt}S_t,$$

where

$$\Delta(t, x) := \frac{\partial}{\partial x} v(t, x), \quad t \in [0, T], x \in \mathbb{R}.$$

In the financial language this is called '*Delta hedging*'.

7.3. Black-Scholes pricing formula for European Calls and Puts. We now derive an explicit formula for the Black-Scholes price of the European call option $C^{\text{call}} = (S_T - K)^+$. For that purpose we simply choose $f(x) = (x - K)^+$ in (7.1) and (7.2), so

$$v(T, x) = e^{-rT} \mathbb{E}_{\mathbb{Q}} \left[\left(x \exp \left(\sigma \sqrt{T} W + \left(r - \frac{1}{2} \sigma^2 \right) T \right) - K \right)^+ \right],$$

where W is $\mathcal{N}(0, 1)$ -distributed under $\mathbb{Q}$. Substituting $\tilde{K} = e^{-rT} K/x$ and $\tilde{\sigma} = \sigma \sqrt{T}$ we get

$$\begin{aligned} v(T, x) &= \frac{x}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(e^{\tilde{\sigma}y - \frac{1}{2}\tilde{\sigma}^2} - \tilde{K} \right)^+ e^{-y^2/2} dy \\ &= \frac{x}{\sqrt{2\pi}} \int_{\frac{\log \tilde{K} + \frac{1}{2}\tilde{\sigma}^2}{\tilde{\sigma}}}^{\infty} \left(e^{\tilde{\sigma}y - \frac{1}{2}\tilde{\sigma}^2} - \tilde{K} \right) e^{-y^2/2} dy \\ &= \frac{x}{\sqrt{2\pi}} \int_{-d_-}^{\infty} e^{-\frac{(y-\tilde{\sigma})^2}{2}} dy - x\tilde{K} (1 - \Phi(-d_-)) \\ &= x\Phi(d_+) - Ke^{-rT}\Phi(d_-), \end{aligned}$$

where Φ denotes the distribution function of the standard normal distribution,

$$d_- = d_-(T, x) := \frac{\log(\frac{x}{K}) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = -\frac{\log \tilde{K} + \frac{1}{2}\tilde{\sigma}^2}{\tilde{\sigma}}$$

and

$$d_+ = d_+(T, x) := d_-(T, x) + \sigma\sqrt{T} = \frac{\log(\frac{x}{K}) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}.$$

To summarize, the Black-Scholes price for a European call option with strike price K is given by $v(T, S_0)$ where

$$v(T, x) = x\Phi(d_+) - Ke^{-rT}\Phi(d_-) \quad (7.7)$$

with

$$d_{\pm} = d_{\pm}(T, x) := \frac{\log(\frac{x}{K}) + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}},$$

and for any $t \in [0, T]$ the value of the option at time t is given by $v(T - t, S_t)$.

Now we turn to pricing European put options $C^{\text{put}} = (K - S_T)^+$. Since we have computed already the price of the corresponding call option we can use the so-called *put-call parity*, which refers to the fact that

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

and the right-hand side equals the pay-off of a forward contract (cf. Section 0) with price $S_0 - e^{-rT}K$ (note that the contingent claim $C = S_T$ can be trivially replicated just by holding one unit of the risky asset which requires an initial investment S_0).

Hence, the price $\pi(C^{\text{put}})$ for C^{put} can be obtained from the price $\pi(C^{\text{call}})$ for the call C^{call} , namely

$$\begin{aligned}\pi(C^{\text{put}}) &= \pi(C^{\text{call}}) - (S_0 - e^{-rT}K) \\ &= S_0 \Phi(d_+) - Ke^{-rT} \Phi(d_-) - S_0 + e^{-rT}K \\ &= e^{-rT}K \Phi(-d_-) - S_0 \Phi(-d_+).\end{aligned}$$

We can also determine a hedging strategy $\bar{\theta}$ for the call option if we use the 'Delta hedging' discussed at the end of Section 7.2, which gives $\theta_t = \Delta(T - t, S_t)$. The *Delta* of the call option $C^{\text{call}} = (S_T - K)^+$ can be computed by differentiating the value function in (7.7) with respect to x ,

$$\Delta(t, x) := \frac{\partial}{\partial x} v(t, x) = \Phi(d_+(t, x)).$$

In particular, note that $\theta_t \in (0, 1)$ a.s. ('long in stock').

The *Gamma* of the call option is given by

$$\Gamma(t, x) := \frac{\partial}{\partial x} \Delta(t, x) = \frac{\partial^2}{\partial x^2} v(t, x) = \varphi(d_+(t, x)) \frac{1}{x\sigma\sqrt{t}},$$

where $\varphi = \Phi'$ denotes the density of the standard normal distribution. Large Gamma values occur in regions where the Delta changes rapidly, corresponding to the need for frequent readjustments of the Delta hedging portfolio. It follows that $x \mapsto v(t, x)$ is strictly convex.

Another important parameter is the *Theta*

$$\Theta(t, x) := \frac{\partial}{\partial t} v(t, x) = \frac{x\sigma}{2\sqrt{t}} \varphi(d_+(t, x)) + Kre^{-rt} \Phi(d_-(t, x)).$$

The fact that $\Theta > 0$ corresponds to the observation that arbitrage-free prices of European call options are typically increasing functions of the maturity. Note that the parameters Δ , Γ and Θ are related by the equation

$$rv(t, x) = rx \Delta(t, x) + \frac{1}{2} \sigma^2 x^2 \Gamma(t, x) - \Theta(t, x).$$

Thus, the function v solves the following partial differential equation, often called the *Black-Scholes equation*

$$rv = rx \frac{\partial v}{\partial x} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 v}{\partial x^2} - \frac{\partial v}{\partial t}.$$

Recall that the Black-Scholes price $v(T, S_0)$ was obtained as the expectation of the discounted payoff $e^{-rT}(S_T - K)^+$ under the equivalent martingale measure $\mathbb{Q}$. Thus, at first glance it may come as surprise that the *Rho* of the option

$$\varrho(t, x) := \frac{\partial}{\partial r} v(t, x) = Kte^{-rt} \Phi(d_-(t, x))$$

is strictly positive, i.e. the price is increasing in r . Note, however, that the martingale measure $\mathbb{Q}$ depends itself on the interest rate r .

The parameter σ is called the *volatility* of the model and may be regarded as a measure of the fluctuations in the stock price process. The price of a European call option is an increasing function of the volatility as the *Vega* of the option

$$\mathcal{V}(t, x) := \frac{\partial}{\partial \sigma} v(t, x) = x\sqrt{t} \varphi(d_+(x, t))$$

is strictly positive. The functions $\Delta, \Gamma, \Theta, \rho$ and $\mathcal{V}$ are usually called the *Greeks* (although 'vega' is not a letter in the Greek alphabet). We refer to [7, Section 5.6] for more details and some nice plots of the Greeks.

Remark 7.4 (Implied volatility). In practice, the prices for European call and put options are known as they can be directly observed in the market, but the volatility parameter σ is unknown. Since the Vega is strictly positive the function $\sigma \mapsto v(t, x)$ is injective, and by inverting this function one can deduce a value for σ from the observed market prices, the so-called *implied volatility*.

7.4. Pricing exotic options. Our next aim is to derive a formula for the Black-Scholes price of a more exotic contingent claim of the form

$$C = \phi(S_T, \inf_{0 \leq t \leq T} S_t).$$

Recall that under $\mathbb{Q}$ the price process is of the form

$$S_t = S_0 e^{\sigma W_t + (r - \frac{1}{2}\sigma^2)t} = S_0 e^{\sigma(W_t + ct)}$$

where $c := (r - \frac{1}{2}\sigma^2)/\sigma$ and W is a $\mathbb{Q}$ -Brownian motion. Thus, the Black-Scholes prices of C is given by

$$\begin{aligned} \pi(C) &= e^{-rT} \mathbb{E}_{\mathbb{Q}} \left[\phi(S_T, \inf_{0 \leq t \leq T} S_t) \right] \\ &= e^{-rT} \mathbb{E}_{\mathbb{Q}} \left[\phi \left(S_0 \exp(\sigma(W_T + cT)), S_0 \exp(\sigma \inf_{0 \leq t \leq T} (W_t + ct)) \right) \right] \\ &= e^{-rT} \mathbb{E}_{\mathbb{Q}} \left[\phi \left(S_0 e^{\sigma \tilde{W}_T}, S_0 e^{\sigma Y_T} \right) \right], \end{aligned}$$

where $\tilde{W}_t := W_t + ct$, $t \geq 0$, and $Y_T := \inf_{0 \leq t \leq T} \tilde{W}_t$. Recall that by Lemma 6.15,

$$\mathbb{Q}[\tilde{W}_T \in dx, Y_T \geq y] = e^{cx - \frac{c^2}{2}T} (p_T(0, x) - p_T(0, 2y - x)) dx,$$

for $y < 0$ and $x - y > 0$. By differentiating with respect to y we get that the joint density f of $(\tilde{W}_T, Y_T)$ under $\mathbb{Q}$ is given by

$$\begin{aligned} f(x, y) &= -\frac{\partial}{\partial y} \mathbb{Q}[\tilde{W}_T \in dx, Y_T \geq y] = \frac{\partial}{\partial y} [p_T(0, 2y - x)] e^{cx - \frac{c^2}{2}T} \\ &= \frac{\partial}{\partial y} \left[\frac{1}{\sqrt{2\pi T}} e^{-\frac{(2y-x)^2}{2T}} \right] e^{cx - \frac{c^2}{2}T} = \frac{2(x-2y)}{\sqrt{2\pi T^3}} e^{-\frac{(2y-x)^2}{2T}} e^{cx - \frac{c^2}{2}T}. \end{aligned}$$

Hence, we obtain for the Black-Scholes price of C ,

$$\pi(C) = e^{-rT} \int_{-\infty}^0 \int_y^{\infty} \phi(S_0 e^{\sigma x}, S_0 e^{\sigma y}) f(x, y) dx dy.$$

We can rewrite the formula by using the change of variables $(x, y) \mapsto \gamma(x, y) = (x - y, y) = (x', y')$ with $\det D\gamma = 1$. In these coordinates the density becomes

$$g(x', y') := f \circ \gamma^{-1}(x', y') = \frac{2(x' - y')}{\sqrt{2\pi T^3}} e^{-\frac{(x' - y')^2}{2T}} e^{c(x' + y') - \frac{c^2}{2}T}, \quad x' > 0, y' < 0,$$

and the formula for the Black-Scholes price reads

$$\pi(C) = e^{-rT} \int_0^\infty \int_{-\infty}^0 \phi(S_0 e^{\sigma(x' + y')}, S_0 e^{\sigma y'}) g(x', y') dy' dx'.$$

As an example we derive a price formula for a European down-and-out option of the form

$$C = h(S_T) \mathbb{1}_{\{\inf_{0 \leq t \leq T} S_t > B\}}.$$

Let $b \in \mathbb{R}$ be such that $B = S_0 e^{\sigma b}$. Then,

$$\inf_{0 \leq t \leq T} S_t = S_0 e^{\sigma Y_T} > B \iff Y_T > b,$$

and by using again Lemma 6.15 we obtain

$$\begin{aligned} \pi(C) &= e^{-rT} \mathbb{E}_{\mathbb{Q}} \left[h(S_T) \mathbb{1}_{\{\inf_{0 \leq t \leq T} S_t > B\}} \right] = e^{-rT} \int_b^\infty h(S_0 e^{\sigma x}) \mathbb{Q}[\tilde{W}_T \in dx, Y_T \geq b] \\ &= e^{-rT} \int_b^\infty h(S_0 e^{\sigma x}) e^{cx - \frac{c^2}{2}T} (p_T(0, x) - p_T(0, 2b - x)) dy. \end{aligned}$$

7.5. The Black-Scholes PDE. In Theorem 6.16 we have seen that for a Brownian motion $(B_t)_{t \geq 0}$ and any $f \in C_b^{1,2}([0, \infty) \times \mathbb{R})$ the process

$$f(t, B_t) - f(0, B_0) - \int_0^t \mathcal{L}f(s, B_s) ds, \quad t \geq 0,$$

is a continuous martingale, where

$$\mathcal{L}f := \partial_t f + \frac{1}{2} \partial_{xx}^2 f.$$

In particular,

$$\mathbb{E} \left[f(t, B_t) - f(0, B_0) - \int_0^t \mathcal{L}f(s, B_s) ds \right] = 0, \quad \forall t \geq 0.$$

In the context of the Black-Scholes model we are interested in the price process $(S_t)_{t \geq 0}$ under the equivalent martingale measure $\mathbb{Q}$ rather than a standard Brownian motion. Recall that

$$S_t = S_0 e^{\sigma W_t + (r - \frac{1}{2}\sigma^2)t}$$

where W is a $\mathbb{Q}$ -Brownian motion. Our aim is to derive a version of Theorem 6.16 for the price process, that is to construct an analogous class of martingales associated with $(S_t)_{t \geq 0}$. The idea is to use a change variables. More precisely, for any $u \in C_b^{1,2}([0, \infty) \times \mathbb{R})$ set

$$f(t, x) := u(t, s(t, x))$$

where

$$s(t, x) := S_0 e^{\sigma x + (r - \frac{1}{2}\sigma^2)t}$$

for fixed S_0 , σ and r . How does the differential operator $\mathcal{L}$ now transform to the new coordinates? By the chain-rule

$$\begin{aligned}\partial_t f &= \partial_t u + \partial_t s \partial_s u = \partial_t u + \left(r - \frac{1}{2}\sigma^2\right) s \partial_s u \\ \partial_x f &= \partial_x s \partial_s u = \sigma s \partial_s u \\ \partial_{xx} f &= \sigma \partial_x s \partial_s u + \sigma s \partial_{xs} \partial_s u = \sigma^2 s \partial_s u + \sigma^2 s^2 \partial_{ss} u,\end{aligned}$$

so that

$$\begin{aligned}\mathcal{L}f &= \partial_t f + \frac{1}{2}\partial_{xx} f \\ &= \partial_t u + \left(r - \frac{1}{2}\sigma^2\right) s \partial_s u + \frac{1}{2}(\sigma^2 s \partial_s u + \sigma^2 s^2 \partial_{ss} u) \\ &= \partial_t u + r s \partial_s u + \frac{1}{2}\sigma^2 s^2 \partial_{ss} u =: \mathcal{G}u.\end{aligned}$$

Thus, for all $t \geq 0$,

$$u(t, S_t) - f(0, S_0) - \int_0^t \mathcal{G}u(r, S_r) dr = f(t, W_t) - f(0, W_0) - \int_0^t \mathcal{L}f(r, W_r) dr,$$

where the right hand side is a continuous $\mathbb{Q}$ -martingale by Theorem 6.16. Therefore, also the left hand side is a continuous $\mathbb{Q}$ -martingale, and both sides have mean zero under $\mathbb{Q}$.

Now suppose that u solves the partial differential equation (PDE)

$$\begin{aligned}\mathcal{G}u &= 0 \quad \text{on } [0, T] \times \mathbb{R}_+ \\ u(T, \cdot) &= \phi.\end{aligned}\tag{7.8}$$

Then, the process $(u(t, S_t))_{t \in [0, T]}$ is a martingale. Further, recall that $(\mathcal{F}_t)_{t \geq 0}$ denotes the natural filtration generated by $(S_t)_{t \geq 0}$ and note that $(S_t)_{t \geq 0}$ is a Markov process (since it can be obtained as the image of the Markov process W under a one-to-one mapping). Hence, for $0 \leq t \leq T$,

$$\mathbb{E}_{\mathbb{Q}} [\phi(S_T) | S_t] = \mathbb{E}_{\mathbb{Q}} [\phi(S_T) | \mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}} [u(T, S_T) | \mathcal{F}_t] = u(t, S_t),$$

so for every $s \in \mathbb{R}_+$,

$$\mathbb{E}_{\mathbb{Q}} [\phi(S_T) | S_t = s] = u(t, s).$$

In particular, $u(t, s)$ is the undiscounted value of a European option $C = \phi(S_T)$ with maturity T conditioned on $S_t = s$ and $u(0, S_0) = \mathbb{E}_{\mathbb{Q}} [\phi(S_T)]$ is the undiscounted price at time $t = 0$.

Conversely, one can also show that if we define $u(t, s) := \mathbb{E}_{\mathbb{Q}} [\phi(S_T) | S_t = s]$, then u solves the PDE (7.8). However, the standard argument for such a result requires some more advanced stochastic calculus, in particular Itô's formula and some facts about martingales in continuous time.

Let now

$$V(t, s) := e^{-r(T-t)} u(t, s)$$

be the discounted value at time t . Then, since $\mathcal{G}u = 0$,

$$\begin{aligned}\partial_t V &= rV + e^{-r(T-t)} \partial_t u = rV + e^{-r(T-t)} \left(-rs \partial_s u - \frac{1}{2} \sigma^2 s^2 \partial_{ss} u \right) \\ &= rV - rs \partial_s V - \frac{1}{2} \sigma^2 s^2 \partial_{ss} V,\end{aligned}$$

that is V solves the *Black-Scholes-PDE*

$$\partial_t V + rs \partial_s V + \frac{1}{2} \sigma^2 s^2 \partial_{ss} V - rV = 0 \quad (7.9)$$

with terminal condition $V(T, \cdot) = \phi$.

Remark 7.5. In the special case of a European call option, i.e. $\phi(x) = (x - K)^+$ we have seen already at the end of Section 7.3 that the value function $v(T - t, x)$ of the call given in (7.7) satisfies the Black-Scholes PDE.

Similarly, if we consider a Black-Scholes model with dividends at rate $D > 0$, i.e.

$$\tilde{S}_t = S_0 e^{\sigma W_t + (r - D - \frac{1}{2} \sigma^2) t}, \quad t \geq 0,$$

then the undiscounted value of an option $C = \phi(\tilde{S}_T)$ at time t conditional on $\tilde{S}_t = s$ is given by $\tilde{u}(t, s)$, where $\tilde{u}$ solves the PDE

$$\begin{aligned}\tilde{\mathcal{G}}u &= 0 \quad \text{on } [0, T] \times \mathbb{R}_+ \\ \tilde{u}(T, \cdot) &= \phi.\end{aligned}$$

with

$$\tilde{\mathcal{G}}\tilde{u} = \partial_t \tilde{u} + (r - D)s \partial_s \tilde{u} + \frac{1}{2} \sigma^2 s^2 \partial_{ss} \tilde{u}$$

For the discounted value function we get $\tilde{V}(t, s) = e^{-r(T-t)} \tilde{u}(t, s)$, which satisfies the PDE

$$\partial_t \tilde{V} + (r - D)s \partial_s \tilde{V} + \frac{1}{2} \sigma^2 s^2 \partial_{ss} \tilde{V} - r\tilde{V} = 0.$$

Note that a naive replacement of r by $r - D$ in (7.9) does not give the correct equation. This is because the term $-rV$ comes from discounting at riskfree rate r and has nothing to do with dividends.

To summarise, we have seen that in the Black-Scholes model the value process of an option $C = \phi(S_T)$ solves the Black-Scholes PDE (7.9) with terminal condition $V(T, \cdot) = \phi$. In order to find the Black-Scholes price for such an option, i.e. the value function at time 0, one could try to solve the PDE, for which some numerical methods are available.

7.6. Numerical schemes. From the discussion in Section 7.5 it should come not as a surprise that the Black-Scholes PDE (7.9) can be reduced to a heat equation of the form

$$\begin{aligned}\partial_t u + \partial_{xx} u &= 0, \quad \text{on } [0, T] \times [0, L] \\ u(0, \cdot) &= g\end{aligned} \quad (7.10)$$

(see exercises). This comforts us in discussing numerical PDE schemes in the simplest possible setting of the standard heat equation (7.10) with additional initial/boundary data

$$u(\cdot, 0) = a, \quad u(\cdot, L) = b,$$

which are required for the implementation of numerical schemes.⁷ We are given the grid

$$\{(ik, jh), i = 1, \dots, N_t, j = 1, \dots, N_x\} \subset [0, T] \times [0, L]$$

with $h = L/N_x$ and $k = T/N_t$. We seek for approximations $U_j^i \approx u(ik, jh)$. The simplest approach is to replace the derivatives ∂_t and ∂_{xx} by finite differences. There are several methods to implement that.

1) *FTCS-method (Forward-difference-in-Time and Central-difference-in-Space).*

$$\frac{U_j^{i+1} - U_j^i}{k} = \frac{U_{j+1}^i - 2U_j^i + U_{j-1}^i}{h^2},$$

which readily rewrites as

$$U_j^{i+1} = \nu U_{j+1}^i + (1 - 2\nu)U_j^i + \nu U_{j-1}^i$$

with $\nu := k/h^2$. Note that for $j = 0$ and $j = N_x$ we need to use the boundary information

$$U_0^{i+1} = a((i+1)k), \quad U_{N_x}^{i+1} = b((i+1)k).$$

2) *BTCS-method (Backward-difference-in-Time and Central-difference-in-Space).*

$$\frac{U_j^i - U_j^{i-1}}{k} = \frac{U_{j+1}^i - 2U_j^i + U_{j-1}^i}{h^2},$$

which after replacing i by $i+1$ rewrites as

$$U_j^{i+1} - \nu(U_{j+1}^{i+1} - 2U_j^{i+1} + U_{j-1}^{i+1}) = U_j^i.$$

In fact, here we need to solve a system of linear equations to get to the $(i+1)$ -st time level.

⁷In the Black-Scholes model the value process of a European call option has boundary data $V(t, 0) = 0$ and $V(t, s) \sim s$ as $s \rightarrow \infty$ for all $t \in [0, T]$. This suggests to approximate the Black-Scholes PDE on $[0, T] \times [0, L]$ for large L with boundary data $V(\cdot, 0) = 0$ and $V(\cdot, L) = L$.

3) *Crank-Nicolson method*. Here the idea is to use a 'half-central' approximation in space, that is to replace the second discrete derivative (in space) on the right hand side in FTCS or BTCS by the average of the second derivatives in time level i and $i + 1$.

$$\frac{U_j^{i+1} - U_j^i}{k} = \frac{1}{2} \left(\frac{U_{j+1}^i - 2U_j^i + U_{j-1}^i}{h^2} + \frac{U_{j+1}^{i+1} - 2U_j^{i+1} + U_{j-1}^{i+1}}{h^2} \right),$$

and rearranging leads to

$$2(1 + \nu) U_j^{i+1} = \nu U_{j+1}^{i+1} + \nu U_{j-1}^{i+1} + \nu U_{j+1}^i + 2(1 - \nu) U_j^i + \nu U_{j-1}^i.$$

Remark 7.6. For $i = 1, \dots, N_t$, $U^i := (U_j^i)_{j=1, \dots, N_x-1} \in \mathbb{R}^{N_x-1}$ can be written as follows.

1) In FTCS,

$$U^{i+1} = F U^i + p^i$$

for suitable $F \in \mathbb{R}^{(N_x-1) \times (N_x-1)}$ and $p^i \in \mathbb{R}^{N_x-1}$.

2) In BTCS,

$$B U^{i+1} = U^i + q^i$$

for suitable $B \in \mathbb{R}^{(N_x-1) \times (N_x-1)}$ and $q^i \in \mathbb{R}^{N_x-1}$.

3) In Crank-Nicolson,

$$(I + B) U^{i+1} = (F + I) U^i + p^i + q^i,$$

where I denotes the identity matrix.

Thus, in a sense the Crank-Nicolson method may be regarded as the 'sum' of FTCS and BTCS. We leave the proof as an exercise.

To compare this three schemes we need a notion of 'local accuracy'. It is defined by sticking the exact solution into the difference formula under consideration, so U_j^i is replaced by $u(ik, jh)$. For instance, for FTCS we get by Taylor expansion that

$$\begin{aligned} R_j^i &:= \frac{u((i+1)k, jh) - u(ik, jh)}{k} - \frac{u(ik, (j+1)h) - 2u(ik, jh) + u(ik, (j-1)h)}{h^2} \\ &= \partial_t u(ik, jh) + \frac{1}{2} \partial_{tt} u(ik, jh) k + O(k^2) \\ &\quad - \frac{1}{h^2} \left[\partial_x u(ik, jh) h + \frac{1}{2} \partial_{xx} u(ik, jh) h^2 + \frac{1}{6} \partial_{xxx} u(ik, jh) h^3 + O(h^4) \right. \\ &\quad \left. - \partial_x u(ik, jh) h + \frac{1}{2} \partial_{xx} u(ik, jh) h^2 - \frac{1}{6} \partial_{xxx} u(ik, jh) h^3 + O(h^4) \right] \\ &= \frac{1}{2} \partial_{tt} u(ik, jh) k + O(k^2) + O(h^2) = O(k) + O(h^2), \end{aligned}$$

where we also used that $\partial_t u = \partial_{xx} u$. A similar computation for BTCS gives local accuracy

$$-\frac{1}{2} \partial_{tt} u(ik, jh) k + O(k^2) + O(h^2) = O(k) + O(h^2).$$

This suggests that simple averaging of FCTS and BTCS schemes leads to a scheme with local accuracy of order $O(k^2) + O(h^2)$. As discussed in Remark 7.6 above, such an 'average' of FCTS and BTCS corresponds exactly to the Crank-Nicolson method.

REFERENCES

- [1] H. Bauer. *Measure and integration theory*, volume 26 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2001. Translated from the German by Robert B. Burckel.
- [2] A. Bovier. Stochastic Processes. *Lecture Notes, University of Bonn*, 2013.
- [3] K. L. Chung. *A course in probability theory*. Academic Press, New York-London, second edition, 1974. Probability and Mathematical Statistics, Vol. 21.
- [4] J. Cox, S. Ross, and M. Rubinstein. Option pricing: a simplified approach. *J. Financial Econom.*, 7:229–263, 1979.
- [5] M. H. Davis. Option Pricing in Incomplete Markets. Available at http://www.imperial.ac.uk/~mdavis/docs/option_pricing_in_incomplete_markets.pdf.
- [6] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520, 1994.
- [7] H. Föllmer and A. Schied. *Stochastic finance*. Walter de Gruyter & Co., Berlin, extended edition, 2011. An introduction in discrete time.
- [8] P. Friz and L. Rogers. Stochastic Financial Models. *Lecture notes*.
- [9] J. Hull. *Options, Futures and Other Derivatives*. Pearson, 2017.
- [10] D. Kennedy. *Stochastic financial models*. Chapman & Hall/CRC Financial Mathematics Series. CRC Press, Boca Raton, FL, 2010.
- [11] J. Norris. Advanced Probability. *Lecture notes, University of Cambridge*, 2016.
- [12] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester, second edition, 1994. Foundations.
- [13] A. Schied. Wahrscheinlichkeitstheorie II. *Lecture notes, TU Berlin*, 2005.
- [14] R. Weber. Optimization and Control. *Lecture notes*.

UNIVERSITY OF CAMBRIDGE

Current address: Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB

E-mail address: s.andres@statslab.cam.ac.uk