Introduction to stochastic finance in continuous time

P.J.C. Spreij

this version: July 17, 2022

Preface

These lecture notes have been written for and during the course *Hedging en Derivaten* at the Universiteit van Amsterdam in Fall 2001. Students that took the course were assumed to have finished a course on Finance in Discrete Time and therefore to be familiar with the standard notions of Mathematical Finance. For those who missed that course I included a short summary of some essentials in Discrete Time Finance.

Students were not supposed to have experience in measure theory, but it turned out that they had some knowledge of it and followed a course in Measure Theory parallel to the course *Hedging en Derivaten*. Therefore, the later sections use more measure theoretic concepts than the first ones. The Appendix contains some notions from Probability theory that are essential for the course.

The course basically starts with showing the first steps towards continuous time models by invoking the central limit theorem for a sequence of discrete time models. This motivates the use of (log)normal models. Since Brownian motion is so prominently around in continuous time models it is extensively introduced. Other basic topics include the study of the heat equation and equations that can be derived from it, since these are instrumental in pricing financial derivatives. The use of Itô calculus has been postponed until it was unavoidable, the general definition of self financing strategies. Nevertheless, we also introduce a definition for self financing Markovian portfolios, that is based on a limit argument and discrete time analogues. A rigorous treatment of Itô integrals and the Itô rule was beyond the scope of the course, but we have provided the reader with some of the basic notions and results and gave here and there some heuristic arguments or a partial proof, when full proofs would have been technically too demanding.

Finally, these lectures notes will be updated and adapted for a next course. I already found some errors and omissions, many of them thanks to Ge Hong and Ramon van den Akker who carefully went through the first version of the manuscript. In 2016 a substantial revision has taken place, after I have taught the course *Financiële Wiskunde* at the Radboud University Nijmegen. Since then, there have been quite a few updates, a major one being the inclusion of a section on interest rate models in 2020. Suggestions for improvement are always welcome.

Amsterdam, May 2020 — Peter Spreij

Contents

1 From discrete to continuous time

In this section we briefly recall a number of fundamental issues and problems in the theory of pricing of derivatives in a finite security market where time is discrete. Because the emphasis is on concepts we consider a market that consists of only one risky asset (the stock) and one riskless asset (the bond). Later on we will study approximations when the number of trading times becomes large.

1.1 A summary of discrete time results

The basic setting is as follows. We consider a financial market where two kinds of products are traded, risky and non-risky assets. Trading takes place at time instants $t \in \{0, \ldots, N\}$. Traded are a bond (a non-risky asset) with corresponding prices at time t equal to B_t and a stock (a risky asset) with prices at time t equal to S_t. The typical example of a bond is that of a bank account. For the stock we could think of shares of a company at the stock exchange or of the exchange rate of the Euro against the US dollar. We adopt the normalization that $B_0 = 1$. Furthermore the bond price is assumed to grow exponentially with rate r, a deterministic real number, meaning that we have for all t that $B_t = (1 + r)^t$.

The price of the stock is assumed to be a random process–recall that the stock is the risky asset–, implying that each time instant t one is uncertain about the behavior of the stock price at future instants. Given that one knows the price at a time t, at time $t + 1$ and future times more than one price of the stock is possible, but we don't know exactly which one. In this section we will later on mainly treat the Cox-Ross-Rubinstein (CRR) model, which says that from a given price at a certain time, only two values of the price at the next time instant are possible, and with certain assigned values only.

In the sequel we will often work with *discounted* prices and values. These are denoted by a bar on the variable under consideration. More precisely, if Y_t denotes the value or price of some financial product (not necessarily the bond or the stock) at time t, then by $\bar{Y}_t$ we denote its discounted value or price and it is defined by

$$\bar{Y}_t = \frac{Y_t}{B_t}.$$

Note that we divide Y_t by the bond price B_t at the same time t. In the jargon of Mathematical Finance we say that we choose the bond price B as a *numéraire*. For example, we will often encounter the discounted price process $\bar{S}$ of the risky asset. Observe that $\bar{B}_t = 1$ for all t.

Next we consider *portfolios*. Formally, a portfolio (in our context) is a sequence of (random) real pairs (x_t, y_t) with the interpretation that x_t is the amount of stock that an investor holds at time t and y_t the number of bonds. The *value* at t of such a portfolio is denoted V_t and is given by

$$V_t = x_t S_t + y_t B_t.$$

Note that the discounted value process of the portfolio is $\bar{V}_t = x_t \bar{S}_t + y_t$ and observe that we allow x_t and y_t to be any real number. A negative value of y_t then corresponds to borrowing money from the bank and a negative value of x_t

to short selling of the stock. Furthermore, we will not allow to have x_t and y_t to depend on future values S_u $(u \geq t)$, the investor is not clairvoyant. So for all t the values of x_t and y_t only depend on the stock through S_0 to S_{t-1}. In more sophisticated terms, the process $S = (S_0, \ldots, S_T)$ is adapted to a certain filtration, and the processes (sequences) $x = (x_0, \ldots, x_T)$ and $y = 0, \ldots, y_T)$ are predictable. See Section A.7 for this terminology.

An important concept is that of a *self-financing* portfolio. Intuitively speaking, a portfolio is called self financing if and only if an initial investment is made and any reallocation of the portfolio is made without infusion or withdrawal of money, so it is done in a budget neutral way. To be precise, we adopt the following definition.

Definition 1.1 A portfolio is called self-financing if for all $t \geq 0$ we have

$$x_t S_t + y_t B_t = x_{t+1} S_t + y_{t+1} B_t. \tag{1.1}$$

We can give this definition an equivalent expression after having introduced the difference operator Δ. For any process Y the process ΔY is defined by $\Delta Y_t = Y_t - Y_{t-1}$ for $t \geq 1$ and $\Delta Y_0 = Y_0$. Now we have the following statement. A portfolio is self-financing iff

$$\Delta V_t = x_t \Delta S_t + y_t \Delta B_t \text{ for all } t \geq 0. \tag{1.2}$$

In terms of discounted prices and values Equation (1.2) takes the following simpler form.

$$\Delta \bar{V}_t = x_t \Delta \bar{S}_t \text{ for all } t \geq 0. \tag{1.3}$$

The proof of Equations (1.2) and (1.3) both being equivalent to (1.1) is left as Exercise 1.3.

A desirable property of a financial market is that it is *free of arbitrage*, meaning –in a sense to be specified below– that it is impossible to make money out of nothing. Formally, we call a portfolio an *arbitrage opportunity* (over the discrete time interval $\{0, \ldots, N\}$) if it is self-financing, its value V_0 at time zero is equal to zero and its value at N is always nonnegative, whereas V_N strictly positive is possible. So, with an arbitrage portfolio it is impossible to loose money, whereas making a profit is a possibility. Often probabilities are attributed to the stock price movements. If these are modeled by a probability measure $\mathbb{P}$, then we have the following formal definition in probabilistic terms.

Definition 1.2 We say that a portfolio with value process $V = \{V_t : t = 0, \ldots, N\}$ is an *arbitrage opportunity* if $\mathbb{P}(V_0 = 0) = 1$, $\mathbb{P}(V_N \geq 0) = 1$ and $\mathbb{P}(V_N > 0) > 0$. A market is *arbitrage free* if no arbitrage possibilities exist.

One of the main issues in Mathematical Finance is the pricing (or valuation) of contingent claims, also called derivatives. Contingent claims are financial products that are defined in terms of *underlying* products. It is therefore reasonable to think that the price of a contingent claim should in some way depend on the price of these underlying products. Indeed, under the no arbitrage condition, we will see that this is the case for the CRR model. However, in general absence of arbitrage is not sufficient to determine one single price of the claim that is consistent with this condition.

From now on we will work with the Cox-Ross-Rubinstein (CRR) market. To make precise what this means, we introduce the process Z which is defined for $t \geq 1$ by

$$Z_t = \frac{S_t}{S_{t-1}}. \tag{1.4}$$

In the CRR model it is assumed that for all t the ratio Z_t takes on one of only two values u and d, where $u > d$. We assume also that S_0 is a fixed positive number s. Since $S_t = S_0 \prod_{k=1}^t Z_k$, we get that S_t takes its values in the set $\{sd^t, sud^{t-1}, \ldots, su^t\}$. Along with the process Z we introduce the *(cumulative) return process* $R = (R_1, \ldots, R_N)$. It is defined by the equations

$$\Delta R_t = \frac{\Delta S_t}{S_{t-1}}, \text{ for } t \in \{1, \ldots, N\}. \tag{1.5}$$

Trivially, we get from (1.5) the following equivalent relations

$$\begin{aligned} \Delta S_t &= S_{t-1} \Delta R_t, \ t \geq 1, \\ S_n &= S_0 \prod_{t=1}^n (1 + \Delta R_t), \ n \geq 1. \end{aligned} \tag{1.6}$$

All these relations are summarized in the *notation*

$$S = S_0 \, \mathcal{E}(R), \tag{1.7}$$

with $\mathcal{E}(R)_n = \prod_{t=1}^n (1 + \Delta R_t)$, also called the Doléans exponential of R.

Many of the ideas that have been introduced above can already be illustrated within a *single-period* CRR market, that is a market with $N = 1$. We have the following result.

Proposition 1.3 *The CRR market with $N = 1$ is free of arbitrage iff $d < 1 + r < u$.*

Proof Exercise 1.1. □

Let us now consider how to price contingent claims in a single-period CRR market. A contingent claim X is now by definition a financial product of the form $X = f(S_1)$, defined in terms of some function f. The prime example of such a claim is the *European call option* with exercise price K for which we have $f(x) = (x - K)^+$ and hence $X = (S_1 - K)^+$ [for any real number u, one abbreviates $\max\{u, 0\}$ as u^+]. This is explained below.

In general, the holder of an option has the right, but not the obligation to exercise it. In the case of the call option, the holder has the right to purchase the stock at time $t = 1$ for the price K. She is willing to do this at time $t = 1$ if the true market price S_1 is then greater than K, after which she sells the stock agains the market price, incurring a profit $S_1 - K$. If it happens that $S_1 \leq K$, she does nothing. Combining the two cases the profit she makes can be written as $(S_1 - K)^+$.

The principal question is here: how much is one willing to pay for such a claim at $t = 0$. Note that the future (at $t = 1$) value of the claim is uncertain, due to the different values that S_1 may assume. The solution to this question is obtained by comparing the claim to a portfolio that gives at time $t = 1$ exactly the same payoff as the claim, no matter how the market will evolve. And since these values are the same at $t = 1$, the fair price of the claim at $t = 0$ should be consistent with the no arbitrage condition, the same as the value of the portfolio at $t = 0$. This principle is called *the law of one price*.

The portfolio that has the above mentioned property (exactly the same pay off as the claim) is called *hedging* against the contingent claim. Since we keep the portfolio constant over time in this case, the problem boils down to the finding of real (nonrandom) numbers x and y such that

$$xS_1 + yB_1 = f(S_1),$$

whatever the value of S_1. Since only the two values su and sd are possible, we are faced with the following system of equations

$$\begin{aligned} xsu + y(1+r) &= f(su) \\ xsd + y(1+r) &= f(sd), \end{aligned}$$

whose unique solution (if $u \neq d$) is

$$\begin{aligned} x &= \frac{f(su) - f(sd)}{s(u-d)} \\ y &= \frac{uf(sd) - df(su)}{(1+r)(u-d)}. \end{aligned}$$

With the thus found values we compute the value of the portfolio at $t = 0$ to get

$$\bar{V}_0 = V_0 = \frac{1}{1+r}\left(\frac{1+r-d}{u-d}f(su) + \frac{u-(1+r)}{u-d}f(sd)\right).$$

Hence, by investing an initial capital V_0 needed to purchase the portfolio with the just found quantities x and y we find its value at time $t = 1$ always coinciding with the claim $f(S_1)$. By the *law of one price* and excluding arbitrage, the value of the claim at time $t = 0$ has to be equal to V_0.

Under the no arbitrage condition of Proposition 1.3 the numbers

$$q_u := \frac{1+r-d}{u-d} \text{ and } q_d := \frac{u-(1+r)}{u-d} \tag{1.8}$$

are in $(0, 1)$ and sum to 1, so we can interpret them as *probabilities*. Let us introduce a probability measure $\mathbb{Q}$ on the outcome space of the return Z_1 by putting $\mathbb{Q}(Z_1 = u) = q_u$ and $\mathbb{Q}(Z_1 = d) = q_d$. Then we can write

$$V_0 = \mathbb{E}_{\mathbb{Q}}\,\frac{1}{1+r}f(sZ_1) = \mathbb{E}_{\mathbb{Q}}\,\frac{1}{1+r}f(S_1),$$

or, in discounted terms, $V_0 = \mathbb{E}_{\mathbb{Q}}\,\bar{f}(S_1)$. We conclude that in this example the fair price of a contingent is the mathematical expectation of the discounted value of the claim under a suitably chosen probability measure. This measure, $\mathbb{Q}$, is called the *risk neutral measure.* It has another interesting property.

$$\mathbb{E}_{\mathbb{Q}} S_1 = q_u su + q_d sd = s(1 + r),$$

or, again in discounted terms, $\mathbb{E}_{\mathbb{Q}} \bar{S}_1 = s$. We see that under the risk neutral measure the expectation of the discounted stock price is the same as its initial price. In more sophisticated terms we say that the discounted price process is a *martingale under* $\mathbb{Q}$. If one wishes, one can associate with the price process S (with $t = 0, 1$) any other probability measure $\mathbb{P}$. The fact that S_1 allows two possible outcomes is then reflected by imposing $\mathbb{P}(S_1 = su) > 0$ and $\mathbb{P}(S_1 = sd) > 0$. Since also the corresponding probabilities under $\mathbb{Q}$ are positive, this means that the two probability measures $\mathbb{Q}$ and $\mathbb{P}$ are equivalent in the measure theoretic sense, notation $\mathbb{P} \sim \mathbb{Q}$. Therefore $\mathbb{Q}$ is also called the *equivalent martingale measure.* Note that we have in fact also shown that $\mathbb{Q}$ is the unique measure that makes $\bar{S}$ a martingale.

Since it is now understood how to price contingent claims in a simple one-period market, we proceed by outlining the approach in a multi-period CRR market, so with a time horizon $N > 1$.

We first consider the case $N = 2$ and a *simple* contingent claim X, i.e. we have $X = f(S_2)$, for some real valued function f, defined on the state space $\mathcal{S}_2 = \{su^2, sud, sd^2\}$. Like in the one-period model we try to find a hedging strategy, that is a portfolio that replicates the value of the claim X, no matter how the market evolves. This is in general impossible if one follows a *buy and hold* strategy, like in the single-period model. Recalling the procedure that we followed there, this would in the present case amount to solving a system of three (linear) equations with two unknowns, and this is in general impossible. However, we now allow self-financing strategies! So we are allowed to rebalance our portfolio as long as we respect the budget neutral condition (1.1). We thus have more freedom, namely x_2 and $x_1 = x_0$ together with y_2 and $y_1 = y_0$. Combined with the budget constraint this results in as many equations as variables and so there is some hope for a unique solution. This argument can be made precise and holds for an arbitrary horizon, not only $N = 2$.

Let's see how it works. We consider a general (composite) claim of the type $X = F(s, Z_1, \ldots, Z_N)$. By relabeling, X is also a function of $S_0 = s$ and $S_1, \ldots, S_N$. It is our purpose to find a dynamic portfolio that is such that at $t = N$ its value equals the pay off of the claim. So at time N we should have identically

$$x_N S_N + y_N B_N = X.$$

Let us suppose that we know all Z_i for $i = 1, \ldots, N-1$, and thus in particular S_{N-1}. Then S_N can assume only the two values $S_{N-1}u$ and $S_{N-1}d$, depending on the value of Z_N. Whichever of these two values S_N assumes, x_N and y_N must be the same, hence we have the two equations

$$\begin{aligned} x_N S_{N-1} u + y_N B_N &= F(s, Z_1, \ldots, Z_{N-1}, u) \\ x_N S_{N-1} d + y_N B_N &= F(s, Z_1, \ldots, Z_{N-1}, d). \end{aligned}$$

These equations are like the ones we met before in the one-period case. The solution is

$$x_N = \frac{F(s, Z_1, \ldots, Z_{N-1}, u) - F(s, Z_1, \ldots, Z_{N-1}, d)}{S_{N-1}(u-d)} \tag{1.9}$$

$$y_N = \frac{1}{B_N} \frac{uF(s, Z_1, \ldots, Z_{N-1}, d) - dF(s, Z_1, \ldots, Z_{N-1}, u)}{u-d}. \tag{1.10}$$

To get the value of the hedge portfolio at time $N-1$ we use the self-financing property to write $V_{N-1} = x_{N-1}S_{N-1} + y_{N-1}B_{N-1}$ as $x_N S_{N-1} + y_N B_{N-1}$. Inserting the expressions for x_N and y_N results by direct calculation in

$$V_{N-1} = \frac{B_{N-1}}{B_N} \left(F(s, Z_1, \ldots, Z_{N-1}, u)q_u + F(s, Z_1, \ldots, Z_{N-1}, d)q_d \right), \tag{1.11}$$

with $q_u = \frac{1+r-d}{u-d}$ and $q_d = \frac{u-(1+r)}{u-d}$. Hence the value of the portfolio at time $N-1$ is known (given the past price movements), whereas the value at time N always coincides with X. We conclude, from the law of one price, that the value of the claim X at time $N-1$ has to coincide with V_{N-1}.

If we fix the values of $Z_1, \ldots, Z_{N-1}$ at $z_1, \ldots, z_{N-1}$, then we can write the right hand side of (1.11) as

$$V_{N-1} =: v_{N-1}(z_1, \ldots, z_{N-1}) = \frac{1}{1+r} \mathbb{E}_{\mathbb{Q}} F(s, z_1, \ldots, z_{N-1}, Z_N),$$

where as before we have $\mathbb{Q}(Z_N = u) = q_u$.

Now we are going to find the values of x_{N-1} and y_{N-1} such that the value of the portfolio at time $N-1$ coincides with V_{N-1}, given the price movements up to time $N-2$, the Z_i with $i \leq N-2$, no matter which value Z_{N-1} assumes. One could say that we treat V_{N-1} as a claim at time $N-1$, since we observe from Equation (1.11) that V_{N-1} depends on the $Z_1, \ldots, Z_{N-1}$. Mimicking the previous step, we consider the two cases for the stock price movement Z_{N-1} separately. We get from Equation (1.11) the two relations

$$\begin{aligned}
&x_{N-1}S_{N-2}u + y_{N-1}B_{N-1} = \\
&\frac{B_{N-1}}{B_N} \left(F(s, Z_1, \ldots, Z_{N-2}, u, u)q_u + F(s, Z_1, \ldots, Z_{N-2}, u, d)q_d \right) \\
&x_{N-1}S_{N-2}d + y_{N-1}B_{N-1} = \\
&\frac{B_{N-1}}{B_N} \left(F(s, Z_1, \ldots, Z_{N-2}, d, u)q_u + F(s, Z_1, \ldots, Z_{N-1}, d, d)q_d \right),
\end{aligned}$$

which can analogously be solved as before. We skip the explicit expression and move straight on to the value of the portfolio at time $N-2$. It becomes (you do the computation yourself as Exercise 1.4)

$$\begin{aligned}
V_{N-2} = \frac{1}{(1+r)^2} \big(&F(s, Z_1, \ldots, Z_{N-2}, u, u)q_u^2 \\
&+ F(s, Z_1, \ldots, Z_{N-2}, u, d)q_u q_d \\
&+ F(s, Z_1, \ldots, Z_{N-2}, d, u)q_u q_d \\
&+ F(s, Z_1, \ldots, Z_{N-2}, d, d)q_d^2 \big).
\end{aligned} \tag{1.12}$$

Let us fix the values of $Z_1, \ldots, Z_{N-2}$ at $z_1, \ldots, z_{N-2}$. Then we can write

$$V_{N-2} =: v_{N-2}(z_1, \ldots, z_{N-2})$$

$$= \frac{1}{(1+r)^2}\mathbb{E}_{\mathbb{Q}}F(s, z_1, \ldots, z_{N-2}, Z_{N-1}, Z_N), \tag{1.13}$$

where $\mathbb{Q}$ has the property that Z_N and Z_{N-1} are independent and identically distributed. An explicit formula for the pricing formula is rather involved for general claims, but from (1.13) it becomes clear how the compact form should look like. At any time n, given that the values of $Z_1, \ldots, Z_n$ are $z_1, \ldots, z_n$, we have

$$\begin{aligned} V_n &=: v_n(z_1, \ldots, z_n) \\ &= \frac{1}{(1+r)^{N-n}}\mathbb{E}_{\mathbb{Q}}F(s, z_1, \ldots, z_n, Z_{n+1}, \ldots, Z_N). \end{aligned} \tag{1.14}$$

In this construction V_n, the value of the portfolio at time n has to equal the value of the claim X at time n, since at time N the values of the claim and of the portfolio coincide.

In Equation (1.14) $\mathbb{E}_{\mathbb{Q}}$ denotes expectation taken under the probability measure $\mathbb{Q}$ that is such that $\mathbb{Q}(Z_n = u) = q_u$, $\mathbb{Q}(Z_n = d) = q_d$ for $n = 1, \ldots, N$ and, moreover, it makes the Z_k independent random variables with identical distributions. We note the following important property of $\mathbb{Q}$: It is the *unique* probability measure that makes $\bar{S}$ martingale. The martingale property here means that for all $n \geq 1$ we have

$$\mathbb{E}_{\mathbb{Q}}[\bar{S}_n | Z_1, \ldots, Z_{n-1}] = \bar{S}_{n-1}.$$

Equivalently we have

$$\mathbb{E}_{\mathbb{Q}}[Z_n | Z_1, \ldots, Z_{n-1}] = 1 + r.$$

Let us show the asserted uniqueness of $\mathbb{Q}$. Computing the last conditional expectation as $u\mathbb{Q}(Z_n = u|Z_1, \ldots, Z_{n-1}) + d\mathbb{Q}(Z_n = d|Z_1, \ldots, Z_{n-1}) = 1 + r$, we see that the conditional distribution of Z_n given $Z_1, \ldots, Z_{n-1}$ is determined by $\mathbb{Q}(Z_n = u|Z_1, \ldots, Z_{n-1}) = \frac{1+r-d}{u-d}$, and that this conditional probability doesn't depend on the conditioning random variables, and is therefore equal to the unconditional probability. Hence it follows that the distribution under $\mathbb{Q}$ of the vector $(Z_1, \ldots, Z_n)$ is the product of its marginals. So, under $\mathbb{Q}$, the Z_i are independent random variables.

Knowing this and using properties of conditional expectation (see the Appendix, Propositions A.10 or A.12), we can write (1.14) in the following equivalent form (with $X = F(s, Z_1, \ldots, Z_N)$).

$$V_n = \frac{1}{(1+r)^{N-n}}\mathbb{E}_{\mathbb{Q}}[F(s, Z_1, \ldots, Z_N)|Z_1, \ldots, Z_n] \tag{1.15}$$

$$= \frac{1}{(1+r)^{N-n}}\mathbb{E}_{\mathbb{Q}}[X|S_1, \ldots, S_n], \tag{1.16}$$

where we used in the last equation that conditioning on $Z_1, \ldots, Z_n$ is equivalent to conditioning on $S_1, \ldots, S_n$. Switching to the discounted value process $\bar{V}$ we get from expression (1.16) that

$$\bar{V}_n = \mathbb{E}_{\mathbb{Q}}[\bar{X}|S_1, \ldots, S_n]. \tag{1.17}$$

It is not hard to see (a rather straightforward computation yields the result) that

$$\bar{V}_n = \mathbb{E}_{\mathbb{Q}}[\bar{V}_{n+1}|S_1,\dots,S_n], \text{ for } n = 0,\dots,N-1. \tag{1.18}$$

Both equations (1.17) and (1.18) express the fact that also the discounted value process $\bar{V}$ is a martingale under the measure $\mathbb{Q}$.

Computing explicit expressions for the x_n and y_n is even more cumbersome than finding the values V_n. Important is however that these exist (under the proviso that $u > d$) for any contingent claim X, that is X is of the form $F(s, Z_1, \dots, Z_N)$. It follows that every contingent claim in the CRR market can be *hedged*. The sequence $(x_n, y_n), n = 1, \dots, N$ is called a *hedging portfolio* or a *hedging strategy*. A market in which every claim can be hedged is said to be *complete*. This result is important enough to state as a proposition.

Proposition 1.4 *The multi-period CRR market (with $u > 1 + r > d$) is complete.*

Remark 1.5 In order to compute hedging strategies in the CRR market as above, it is sufficient to have $u \neq d$ (and then w.l.o.g. $u > d$). But in absence of the condition $u > 1 + r > d$ the expressions for the V_n loose their meaning as conditional expectations, since q_u or q_d may then become negative and can therefore not be interpreted as probabilities anymore. For this reason it is common to impose $u > 1 + r > d$.

Not only is the CRR market complete, we also have

Proposition 1.6 *The multi-period Cox-Ross-Rubinstein market is free of arbitrage under any probability measure $\mathbb{P}$ that is equivalent to $\mathbb{Q}$ iff $d < 1+r < u$.*

Proof We first work under $\mathbb{Q}$. Assume that $d < 1 + r < u$, then we have that $\mathbb{Q}$ is a probability measure. Suppose that an arbitrage strategy exists with corresponding value process V. Then we have $V_0 = \bar{V}_0 = 0$, $V_N \geq 0$ and $\mathbb{Q}(V_N > 0) > 0$. Consequently, we also have $\mathbb{E}_{\mathbb{Q}} V_N > 0$ and $\mathbb{E}_{\mathbb{Q}} \bar{V}_N > 0$. But since $\bar{V}$ is a martingale under $\mathbb{Q}$, we then get $\bar{V}_0 = \mathbb{E}_{\mathbb{Q}} \bar{V}_0 = \mathbb{E}_{\mathbb{Q}} \bar{V}_N > 0$, a contradiction. [See also Exercise A.1.]

If $\mathbb{P}$ is a probability measure that is equivalent to $\mathbb{Q}$, then we have $\mathbb{P}(V_0 = 0) = 1$ iff $\mathbb{Q}(V_0 = 0) = 1$, $\mathbb{P}(V_N \geq 0) = 1$ iff $\mathbb{Q}(V_N \geq 0) = 1$, and $\mathbb{P}(V_N > 0) > 0$ iff $\mathbb{Q}(V_N > 0) > 0$. The result now follows from the previous case.

The proof of the necessity of $d < 1 + r < u$ follows as in the one period case (this is Exercise 1.1). □

The construction of a hedging strategy, the (x_N, y_N) as in (1.9), (1.10) and subsequently (x_n, y_n), shows an interesting aspect, the physical measure $\mathbb{P}$ plays no role. Suppose that two persons have different views (different $\mathbb{P}$) on the stock price movement in a CRR market with $r = 0$ and some $d < 1 < u$. One (the optimist) thinks that all events $\{Z_n = u\}$ have probability 0.99, the other (the pessimist) that these events have probability 0.01. Suppose that both want to buy a European call option with some maturity date and strike price. Although at first glance it seems reasonable to think that the optimist is willing to pay more for the option than the pessimist, we have seen above that their respective perceptions of the market movements are immaterial: if they both handle rationally, they will nevertheless agree on the same price for this option! This may sound surprising, but immediately becomes more understanding if one realizes that both optimist and pessimist also agree on the price of a share, regardless their different views on the future market behaviour.

For simple claims the expression for their values in (1.14) takes a simpler form. In this case the function F is specified by $F(s, Z_1, \dots, Z_N) = f(s\prod_{i=1}^N Z_i) = f(S_N)$, for some other given function f. The pricing formula is now given by

$$V_n =: v_n(S_n) = (1+r)^{-N+n} \sum_{i=0}^{N-n} f(S_n u^i d^{N-n-i}) \binom{N-n}{i} q_u^i q_d^{N-n-i}.$$

If we fix the price S_n at a known number s, then we can rewrite this equation as $v_n(s) = (1+r)^{-N+n}\mathbb{E}_{\mathbb{Q}} f(s\frac{S_N}{S_n})$. The discounted version of this expression is

$$\bar{v}_n(s) = \mathbb{E}_{\mathbb{Q}} \bar{f}(s\frac{S_N}{S_n}). \tag{1.19}$$

Since under the probability measure $\mathbb{Q}$ the ratio S_N/S_n and S_n are independent (why?), property (v) of Proposition A.12 says that $\bar{v}_n(s) = \mathbb{E}_{\mathbb{Q}}[\bar{f}(S_N)|S_n = s]$ and hence (1.19) is equivalent to

$$\bar{V}_n = \mathbb{E}_{\mathbb{Q}}[\bar{f}(S_N)|S_n].$$

For the special case of a European call option we have $f(x) = (x-K)^+$ so that in this case the pricing formula becomes

$$V_n = (1+r)^{-N+n} \sum_{i \in E(n)} (S_n u^i d^{N-n-i} - K) \binom{N-n}{i} q_u^i q_d^{N-n-i},$$

where $E(n)$ is the set of indices i for which $S_n u^i d^{N-n-i} > K$. Note that $E(n)$ is an interval in $\{0, \dots, N-n\}$, possibly empty in which case the above sum is zero. Assuming that $E(n) \neq \emptyset$, we get with $a_n = \min E(n)$,

$$\begin{aligned} V_n &= (1+r)^{-N+n} \sum_{i=a_n}^{N-n} (S_n u^i d^{N-n-i} - K) \binom{N-n}{i} q_u^i q_d^{N-n-i} \\ &= S_n \sum_{i=a_n}^{N-n} \binom{N-n}{i} \left(\frac{uq_u}{1+r}\right)^i \left(\frac{dq_d}{1+r}\right)^{N-n-i} \\ &\quad - K(1+r)^{n-N} \sum_{i=a_n}^{N-n} \binom{N-n}{i} q_u^i q_d^{N-n-i}. \end{aligned} \tag{1.20}$$

Observe that both sums above can be expressed in terms of binomial probabilities. With $\pi(k, p, a)$ the probability that a $\text{Bin}(k, p)$ distributed random variable is larger than or equal to a and $p = \frac{uq_u}{1+r}$ we can rewrite (1.20) as

$$S_n \pi(N-n, p, a_n) - K(1+r)^{-N+n} \pi(N-n, q_u, a_n). \tag{1.21}$$

In concrete cases, one can compute (1.21) using tables for the Binomial distribution, and for big values of $N-n$ approximate it using the Central Limit

Theorem. The latter we will do in Section 1.2.

We close this section with the so called *Put-Call parity*, which relates the fair price of a European call option to a European put option. The latter corresponds to the claim $p(S_N) = (K - S_N)^+$, whereas the former has payoff $c(S_N) = (S_N - K)^+$. Note that $c(S_N) - p(S_N) = S_N - K$. Denote the value of the call option at time n by C_n and that of the put option by P_n. Then one easily obtains (Exercise 1.6) the Put-Call parity formula

$$C_n - P_n = S_n - (1+r)^{n-N} K. \tag{1.22}$$

1.2 Limits in the CRR model

In this section we consider limit properties of the Cox-Ross-Rubinstein model by invoking the Central limit theorem (see Section A.5). In order to do so and to get sensible results we have to make judicious choices of the parameters involved. We consider a trading period which is the real interval $[0,T]$. Trading takes place at the time instants $t_n^N = n\Delta_N$ with $\Delta_N = T/N$. We consider now a sequence of discrete time CRR models indexed by N and in these models we let the parameters depend on N as follows. For a given $r, \sigma > 0$ we put

$$r_N = \exp(r\Delta_N) - 1 \tag{1.23}$$

$$u_N = \exp(\sigma\sqrt{\Delta_N}) \tag{1.24}$$

$$d_N = \exp(-\sigma\sqrt{\Delta_N}). \tag{1.25}$$

We are interested in asymptotics for $N \to \infty$, in which case we have $\Delta_N \to 0$. It follows from Proposition 1.6 that for all small enough Δ_N the CRR market is arbitrage free for any $\mathbb{P} \sim \mathbb{Q}$.

The consequences of the above choices for the parameters are straightforward for the bond price. Let us fix the parameter N for a while and consider the N-th CRR model with bond prices at fictitious times k given by B_k^N. The fictitious time instants k corresponds to the real time instants t_k^N with bond prices $B^N(t_k^N)$. The two bond prices are linked by the relation $B^N(t_k^N) = B_k^N$. For t in the interval $[t_k^N, t_{k+1}^N)$ we define $B^N(t) = B^N(t_k^N)$. Then at $t = k\Delta_N$ we have, using (1.23), $B^N(t) = B_k^N = (1 + r_N)^k = \exp(rt)$.

For arbitrary t we have a similar relation. Let $t \in [0,T]$ be fixed. Then $t \in [t_k^N, t_{k+1}^N)$ with $k = k(N) = [N\frac{t}{T}]$. Since $t_k^N \to t$ as $N \to \infty$, we get $B^N(t) \to B(t) := \exp(rt)$.

For the stock price movements things are more complicated. Let us first set the notation. By S_k^N we denote the stock prices at the discrete times k in the N-th CRR model. Like what we did for the bond price, we fix a time instant t and we define the stock price $S^N(t)$ as $S^N(t) = S_k^N$ with k such that $t \in [t_k^N, t_{k+1}^N)$.

Let us focus on the risk neutral probabilities, now indexed by N, $q_u(N)$ (and $q_d(N)$) in the N-th CRR model. We obtain from Equations (1.8), (1.24) and (1.25),

$$q_u(N) = \frac{e^{r\Delta_N} - \exp(-\sigma\sqrt{\Delta_N})}{\exp(\sigma\sqrt{\Delta_N}) - \exp(-\sigma\sqrt{\Delta_N})}. \tag{1.26}$$

We will consider what happens if $N \to \infty$. Using the Taylor expansion of the exponential function we get, performing some tedious calculations, for $N \to \infty$

$$q_u(N) = \frac{1}{2} + (r - \frac{1}{2}\sigma^2)\frac{\sqrt{\Delta_N}}{2\sigma} + O(\Delta_N), \tag{1.27}$$

and consequently $q_u(N) \to \frac{1}{2}$. Of course also $q_d(N) \to \frac{1}{2}$.

Like before we specialize to the pricing of a European call option. So we consider the claim with payoff $(S^N(T) - K)^+$. The fair price at any time $t \in [t_n^N, t_{n+1}^N)$ is given by formula (1.21) with the appropriate substitutions. So we define $p_N = \frac{u_N q_u(N)}{1+r_N}$ and $a_N(t) = \min\{i : S^N(t)u_N^i d_N^{N-n-i} > K\}$. Note that this set will eventually be non-empty for $N \to \infty$, which will be assumed from now on. We then have the inequalities

$$\frac{\log \frac{K}{S_N(t)d_N^{N-n}}}{\log \frac{u_N}{d_N}} < a_N(t) \leq \frac{\log \frac{K}{S_N(t)d_N^{N-n}}}{\log \frac{u_N}{d_N}} + 1. \tag{1.28}$$

We compute the limits of the probabilities $\pi(N-n, p_N, a_N(t))$ and $\pi(N-n, q_u(N), a_N(t))$ for $N \to \infty$ and $\frac{n}{N} \sim \frac{t}{T}$ so that $t_n^N \to t$. Let us introduce the auxiliary random variable Y_N which has a $\mathrm{Bin}(N-n, p_N)$ distribution. Note that $(N-n)\Delta_N \to T - t$. With the aid of Y_N we have $\pi(N-n, p_N, a_N(t)) = \Pr(Y_N > a_N(t))$, with Pr denoting probability.

It is our aim to apply the Central limit theorem, so we have to compute by standardization

$$\Pr(Y_N > a_N(t)) = \Pr(\frac{Y_N - \mathbb{E}Y_N}{\sqrt{\mathbb{V}\mathrm{ar}\, Y_N}} > \alpha_N(t)),$$

with $\alpha_N(t) = (\mathbb{V}\mathrm{ar}\, Y_N)^{-1/2}(a_N(t) - \mathbb{E}Y_N))$. Therefore we need expectation and variance of Y_N. It is easy to show that $\mathbb{V}\mathrm{ar}\, Y_N = \frac{1}{4}(N-n)(1 + O(\Delta_N))$. Furthermore we have

$$p_N = \frac{u_N q_u(N)}{1 + r_N} = \frac{1}{2} + (r + \frac{1}{2}\sigma^2)\frac{\sqrt{\Delta_N}}{2\sigma} + O(\Delta_N). \tag{1.29}$$

Hence we get (using (1.28))

$$\begin{aligned} a_N(t) - \mathbb{E}Y_N &= a_N(t) - (N-n)p_N \\ &= \frac{1}{2\sigma\sqrt{\Delta_N}}\left(\log\frac{K}{S^N(t)} - (T-t)(r + \frac{1}{2}\sigma^2 + O(\sqrt{\Delta_N}))\right) \end{aligned}$$

and therefore

$$\alpha_N(t) = \frac{\log\frac{K}{S^N(t)} - (T-t)(r + \frac{1}{2}\sigma^2 + O(\sqrt{\Delta_N}))}{\sigma\sqrt{T-t}}.$$

But then, with $S^N(t) = s$ we get (using Exercise A.10)

$$\pi(N-n, p_N, a_N(t)) \to \Phi(\frac{\log(s/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}).$$

The convergence of the probabilities $\pi(N-n, q_u(N), a_N(t))$ can be treated similarly (this is Exercise 1.8). The computations above are now summarized in

Theorem 1.7 *Under the assumptions of this section at time t when the stock price has the value $S^N(t) = x$, the fair price of a European call option with payoff $(S^N(T) - K)^+$ has the limiting expression*

$$C(t,x) = x\Phi(d_1(t,x)) - Ke^{-r(T-t)}\Phi(d_2(t,x)), \tag{1.30}$$

with $d_1(t,x) = \frac{\log(x/K)+(r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$ *and* $d_2(t,x) = d_1(t,x) - \sigma\sqrt{T-t}$.

Remark 1.8 Equation (1.30) is the famous *Black-Scholes formula*, to which we will return in Sections 3 and 5. For now, we only note that the function C satisfies for $t \in (0,T)$ and $x > 0$ the *Black-Scholes partial differential equation*

$$C_t(t,x) + \frac{1}{2}\sigma^2 x^2 C_{xx}(t,x) + rxC_x(t,x) - rC(t,x) = 0, \tag{1.31}$$

with boundary condition $C(T,x) = (x-K)^+$. See further Exercise 1.14.

Not only can we use the Central limit theorem to get a normal approximation for the price of a European call option, but also for the distribution of the stock price itself under the measure $\mathbb{Q}$. This comes as no surprise after the preceding calculations, since for all N also the probability distributions of the $S_N(t)$ are essentially determined by a binomial distribution. In order to make this statement precise we introduce some notation. Let $Z_k^N = S_k^N/S_{k-1}^N$ for $k = 1,\ldots,N$. Define for each $t \in [0,T]$ the random variable $W^N(t) = \sum_{k \le \frac{t}{T}N} \log Z_k^N$.

Proposition 1.9 *Let for each N the distribution of the $S_0^N,\ldots,S_N^N$ be determined under the probability measure $\mathbb{Q}^N$ that is such that $\mathbb{Q}^N(Z_k^N = u_N) = q_u(N)$ for all $k = 0,\ldots,N$ and that makes the $Z_1^N,\ldots,Z_N^N$ independent. Let $t_1,\ldots,t_n$ be a finite increasing sequence in $[0,T]$ and define $\Delta W_k^N = W^N(t_k) - W^N(t_{k-1})$. Then for $N \to \infty$ (under the probability measures $\mathbb{Q}^N$) the random variables ΔW_k^N converge in distribution to a random variable ΔW_k that has a $N\big((t_k - t_{k-1})(r - \frac{1}{2}\sigma^2), (t_k - t_{k-1})\sigma^2\big)$ distribution. Moreover, the distribution of the random n-vector $(\Delta W_1,\ldots,\Delta W_n)$ is such that its components are independent.*

Consequently, the n-vector with elements $\log S^N(t_k) - \log S^N(t_{k-1})$ (with $k = 1,\ldots,n$) converges in distribution to an n-vector with elements $\log S(t_k) - \log S(t_{k-1})$ that has a multivariate normal distribution that is such that all $\log S(t_k) - \log S(t_{k-1})$ have a normal $N\big((r - \frac{1}{2}\sigma^2)(t_k - t_{k-1}), \sigma^2(t_k - t_{k-1})\big)$ distribution. Moreover, the (limit) distribution of this vector is such that its components are independent.

Proof First we compute $\mathbb{E}\Delta W_k^N = \sum_{\frac{t_{k-1}}{T}N < j \le \frac{t_k}{T}N} \mathbb{E}\log Z_j^N$. Since $\log Z_k^N$ can only assume the two values $\sigma\sqrt{\Delta_N}$ and $-\sigma\sqrt{\Delta_N}$ with probabilities $q_u(N)$ and $q_d(N)$, we have $\mathbb{E}\log Z_k^N = \sigma\sqrt{\Delta_N}(q_u(N) - q_d(N))$. Using Equation (1.26) and the companion expression for $q_d(N)$, we get $\mathbb{E}\log Z_k^N = (r - \frac{1}{2}\sigma^2)\Delta_N + O(\Delta_N^{3/2})$. Since ΔW_k^N is the sum over approximately $(t_k - t_{k-1})/\Delta_N$ terms, we obtain $\mathbb{E}\Delta W_k^N = (r - \frac{1}{2}\sigma^2)(t_k - t_{k-1}) + O(\Delta_N^{1/2})$. Similarly one computes $\mathbb{V}\mathrm{ar}\,\Delta W_k^N = 4q_u(N)q_d(N)\sigma^2(t_k - t_{k-1}) \to \sigma^2(t_k - t_{k-1})$. Introducing $\Delta\tilde{W}_k^N = \Delta W_k^N - \mathbb{E}\Delta W_k^N$, we apply Theorem A.9 to have that the $\Delta\tilde{W}_k^N$ have a $N(0, \sigma^2(t_k - t_{k-1}))$ limit distribution. The assertion for the ΔW_k^N then follows from Exercise A.10. □

In Proposition 1.9 we have found the limit distribution for the log price process of the stock, it was such that the limit random variable $\log S_t$ follows a normal distribution; one also says that $S(t)$ follows a log-normal distribution. It is also possible to describe the limit distribution of the cumulative return process, see Equation (1.5) for the definition of the return process. Of course here we write R_n^N to denote the cumulative returns in the N-th CRR model and parallel to the notation that we previously used, we write $R^N(t) = R_n^N$ if $t \in [t_n^N, t_{n+1}^N)$.

Proposition 1.10 *Let $t_1, t_2, \ldots$ be an increasing sequence in $[0,T]$. Under the same assumptions as in Proposition 1.9 we have that the n-vector with elements $R^N(t_k) - R^N(t_{k-1})$ ($k = 1, \ldots, n$) converges in distribution to an n-vector with elements $R(t_k) - R(t_{k-1})$ that has a multivariate normal distribution that is such that all $R(t_k) - R(t_{k-1})$ have a normal $N\left(r(t_k - t_{k-1}), \sigma^2(t_k - t_{k-1})\right)$ distribution. Moreover, the (limit) distribution of this vector is such that its components are independent.*

Proof The proof proceeds along the same lines as that of Proposition 1.9. One shows that $\mathbb{E}(R^N(t_k) - R^N(t_{k-1}))$ converges to $r(t_k - t_{k-1})$ and that its variance has $\sigma^2(t_k - t_{k-1})$ as the limit. Invoking Theorem A.9 and Exercise A.10 will complete the proof. Details are left as Exercise 1.9. □

Compare the limit distributions of Propositions 1.9 and 1.10. We see that they are both normal, with the same variance, but with different expectations. The difference of the expectations of $R(t_k) - R(t_{k-1})$ and that of $\log S(t_k) - \log S(t_{k-1})$ is equal to $\frac{1}{2}\sigma^2(t_k - t_{k-1})$. Let us explain, why this is the case.

With the choice that we made in the present section for u_N and d_N we have first that $\lim_N u_N = \lim_N d_N = 1$, so that ΔR_k^N will be close to zero. A simple computation even yields $(\Delta R_k^N)^2 = \sigma^2 \Delta_N (1 + o(\Delta_N^{\frac{1}{2}}))$, no matter whether the stock price goes up or goes down. But then $\log Z_k^N = \log(1 + \Delta R_k^N) \approx \Delta R_k^N - \frac{1}{2}(\Delta R_k^N)^2 \approx \Delta R_k^N - \frac{1}{2}\sigma^2 \Delta_N (1 + o(\Delta_N^{\frac{1}{2}}))$. Consequently $\log S^N(t_k) - \log S^N(t_{k-1}) \approx R^N(t_k) - R^N(t_{k-1}) - \frac{1}{2}\sigma^2(t_k - t_{k-1})$, which accounts for the different expectations that we came across.

The assertion of Proposition 1.10 can be reflected in an appealing notation. Write $\Delta R^N(t_k)$ for the difference $R^N(t_k) - R^N(t_{k-1})$ and remember that in similar notation this can alternatively be written as $\Delta R^N(t_k) = (S^N(t_k) - S^N(t_{k-1}))/S^N(t_{k-1}) = \Delta S^N(t_k)/S^N(t_{k-1})$. Let $\Delta\mathcal{B}(t_k)$ be a random variable that has a normal $N(0, \Delta t_k)$ distribution with $\Delta t_k = t_k - t_{k-1}$. It then follows that we can write (for $N \to \infty$)

$$\Delta S^N(t_k) \approx S^N(t_{k-1})(r\Delta t_k + \sigma \Delta\mathcal{B}(t_k)). \tag{1.32}$$

We shall encounter later on a continuous time version of this approximate identity.

With the result of Proposition 1.9 in mind we show the limiting expression of the price at a time t of a simple claim $X = f(S_T^N)$, where f is a bounded continuous function (e.g. a European put option). Suppose that at t the stock price $S^N(t)$ is equal to a number s. Let us write U^N for $\log(S^N(T)/S^N(t))$. Then we have parallel to Equation (1.19)

$$\bar{v}_t^N(s) = \mathbb{E}_{\mathbb{Q}^N} \bar{f}(\exp(U^N)s).$$

From Proposition 1.9 we know that the limit distribution of U^N is normal with mean $(r - \frac{1}{2}\sigma^2)(T-t)$ and variance $\sigma^2(T-t)$. Hence, we can apply the portmanteau theorem (see Appendix, Theorem A.6) to conclude that $\bar{v}_t^N(s)$ converges the corresponding expectation in the limit model, i.e. to

$$\bar{v}_t(s) := \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \bar{f}(se^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma\sqrt{T-t}\,z})e^{-z^2/2}\,\mathrm{d}z. \tag{1.33}$$

We conclude this section by saying that we obtained a stochastic model for stock price movements, involving normal distributions, as a limit of simple discrete time models. In later sections we will arrive at the same model and some of its ramifications by a different and more direct approach. Note also that the convergence that we considered was in terms of finite dimensional distributions (we considered n-vectors in Propositions 1.9 and 1.10). It is possible to go beyond this and show *functional convergence* in terms of processes. What one does then is to consider the $S^N(\cdot)$ as random elements in a space of functions (one where all functions are right continuous and have left limits). The $S^N(\cdot)$ induce probability measures on this (infinite dimensional) function space, and one then studies weak convergence of these probability measures. This is, although possible, much harder to do and falls beyond the scope of the present course.

1.3 Exercises

1.1 Prove Proposition 1.3.

1.2 Consider in a CRR model the claim with final payoff S_N. Derive the fair price of this claim at time $n \leq N$.

1.3 Show that a portfolio is self-financing iff (1.2) holds and iff (1.3) holds.

1.4 Show the validity of (1.12).

1.5 Consider in the same CRR financial market with fixed terminal time N two European call options with strike prices $K_1 > K_2$. Which of the two has the highest price?

1.6 Consider the CRR model with a call and a put option. Derive the put-call parity Equation (1.22).

1.7 Show that a portfolio in a discrete time market is self-financing iff its discounted value process $\bar{V}$ is a martingale under $\mathbb{Q}$.

1.8 Show by using the Central limit theorem that with $n \sim \frac{t}{T}N$ one gets $\lim_{N\to\infty} \pi(N-n, q_u(N), a_N(t)) = \Phi(d_2(t))$ (notation as in Theorem 1.7).

1.9 Prove Proposition 1.10.

1.10 Show that the distribution of the vector $(R(t_1), \ldots, R(t_n))$ of Proposition 1.10 is multivariate normal. What are the expectation vector and covariance matrix?

1.11 Suppose a continuous model for the stock price is such that $\log(S(T)/S(t))$ has a normal $N((r - \frac{1}{2}\sigma^2)(T-t), \sigma^2(T-t))$ distribution (under $\mathbb{Q}$). Assume that at time t the price $S(t)$ is known to be equal to s. Then the price of the usual European call option at time t is known to be

$$e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}(s\frac{S(T)}{S(t)} - K)^+.$$

Show by computation of an integral that the explicit expression of this price is given by the Black-Scholes formula of Equation (1.30).

1.12 Use Equation (1.33) to compute explicitly the limit price of a European put option (with payoff $(K - S(T))^+$).

1.13 Consider the limiting CRR models and let the stock price at time t be equal to s. Let $C(t)$ be the limit of the price at t of a European call option with payoff $(S(T) - K)^+$ and $P(t)$ the price at t of the corresponding put option.

(a) Derive the limit put-call parity equation $C(t) - P(t) = s - e^{-r(T-t)}K$.

(b) Use the result of Exercise 1.12 to arrive at the Black-Scholes formula (1.30).

(c) Suppose one doesn't use the risk-neutral probabilities $q_u(N)$ and $q_d(N)$ in Theorem 1.7, but instead $p_u(N) = \frac{1}{2} + (\mu - \frac{1}{2}\sigma^2)\frac{\sqrt{\Delta}}{2\sigma}$ and the corresponding $p_d(N)$ for some $\mu \in \mathbb{R}$ and sufficiently small positive $\Delta = \frac{T}{N}$. What would then be the limit laws of $\log S_N(t)$ and $\log S_N(t) - \log S_N(s)$ (for $t > s$)?

1.14 Consider the function C defined in (1.30).

(a) Show that C satisfies the partial differential equation (1.31) for $t < T$ and $x > 0$. *Hint:* Show and use

$$\begin{aligned}
d_1 - d_2 &= \sigma\sqrt{T-t} \\
d_1 + d_2 &= \frac{2}{\sigma\sqrt{T-t}}\left(\log\frac{x}{K} + r(T-t)\right), \\
\phi(d_2) &= \frac{x}{K}\phi(d_1)\exp(r(T-t)), \\
\frac{\partial d_2}{\partial t} &= \frac{\partial d_1}{\partial t} + \frac{1}{2}\sigma(T-t)^{-1/2},
\end{aligned}$$

where ϕ denotes the density of the standard normal distribution.

(b) Show that $\lim_{t\uparrow T} C(t,x) = (x-K)^+$. Why is this to be expected?

2 Brownian motion

In this section we prove the existence of Brownian motion, perhaps the most famous example of a stochastic process. The technique that is used in the existence proof is based on linear interpolation properties for continuous functions.

2.1 Interpolation of continuous functions

Let a continuous function $f : [0,1] \to \mathbb{R}$ be given with $f(0) = 0$. We will construct an approximation scheme of f, consisting of continuous piecewise linear functions. To that end we make use of the dyadic numbers in $[0,1]$. Let for each $n \in \mathbb{N}$ the set D_n be equal to $\{k2^{-n} : k = 0, \ldots, 2^n\}$. The dyadic numbers in $[0,1]$ are then the elements of $\cup_{n=1}^{\infty} D_n$. To simplify the notation we write t_k^n for $k2^{-n} \in D_n$.

The interpolation starts with $f_0(t) \equiv tf(1)$ and then we define the other f_n recursively. Suppose f_{n-1} has been constructed by prescribing the values at the points t_k^{n-1} for $k = 0, \ldots, 2^{n-1}$ and by linear interpolation between these points. Look now at the points t_k^n for $k = 0, \ldots, 2^n$. For the even integers $2k$ we take $f_n(t_{2k}^n) = f_{n-1}(t_k^{n-1})$. Then for the odd integers $2k-1$ we define $f_n(t_{2k-1}^n) = f(t_{2k-1}^n)$. We complete the construction of f_n by linear interpolation between the points t_k^n. Note that for $m \geq n$ we have $f_m(t_k^n) = f(t_k^n)$.

The above interpolation scheme can be represented in a more compact way (to be used in Section 2.2) by using the so-called *Haar* functions H_k^n. These are defined as follows. $H_1^0(t) \equiv 1$ and for each n we define H_k^n for $k \in I(n) = \{1, \ldots, 2^{n-1}\}$ by

$$H_k^n(t) = \begin{cases} \frac{1}{2\sigma_n} & \text{if } t_{2k-2}^n \leq t < t_{2k-1}^n \\ -\frac{1}{2\sigma_n} & \text{if } t_{2k-1}^n \leq t < t_{2k}^n \\ 0 & \text{elsewhere} \end{cases} \tag{2.1}$$

where $\sigma_n = 2^{-\frac{1}{2}(n+1)}$. Next we put $S_k^n(t) = \int_0^t H_k^n(u)\,\mathrm{d}u$. Note that for $n \geq 1$ the support of S_k^n is the interval $[t_{2k-2}^n, t_{2k}^n]$ and that the graphs of the S_k^n are tent shaped with peaks of height σ_n at t_{2k-1}^n. For $n = 0$, $k = 1$ one has $S_1^0(t) = t$.

Next we will show how to cast the interpolating scheme in such a way that the Haar functions, or rather the *Schauder* functions S_k^n, are involved. Observe that not only the S_k^n are tent shaped, but also the consecutive differences $f_n - f_{n-1}$ on each of the intervals $(t_{k-1}^{n-1}, t_k^{n-1})$! Hence they are multiples of each other and to express the interpolation in terms of the S_k^n we only have to determine the multiplication constant. The height of the peak of $f_n - f_{n-1}$ on $(t_{k-1}^{n-1}, t_k^{n-1})$ is the value η_k^n at the midpoint t_{2k-1}^n. So $\eta_k^n = f(t_{2k-1}^n) - \frac{1}{2}(f(t_{k-1}^{n-1}) + f(t_k^{n-1}))$. Then we have for $t \in (t_{2k-2}^n, t_{2k}^n)$ the simple formula

$$f_n(t) - f_{n-1}(t) = \frac{\eta_k^n}{\sigma_n} S_k^n(t),$$

and hence we get for all t

$$f_n(t) = f_{n-1}(t) + \sum_{k \in I(n)} \frac{\eta_k^n}{\sigma_n} S_k^n(t). \tag{2.2}$$

Summing Equation (2.2) over n leads with $I(0) = \{1\}$ to the following representation of f_n on the whole interval $[0,1]$:

$$f_n(t) = \sum_{m=0}^{n} \sum_{k \in I(m)} \frac{\eta_k^m}{\sigma_m} S_k^m(t). \tag{2.3}$$

Theorem 2.1 *Let f be a continuous function on $[0,1]$. With the f_n defined by (2.3) we have $||f - f_n|| \to 0$, where $||\cdot||$ denotes the sup norm.*

Proof Let $\varepsilon > 0$ and choose N such that we have $|f(t) - f(s)| \leq \varepsilon$ as soon as $|t - s| < 2^{-N}$. It is easy to see that then $|\eta_k^n| \leq \varepsilon$ if $n \geq N$. On the interval $[t_{2k-2}^n, t_{2k}^n]$ we have that

$$|f(t) - f_n(t)| \leq |f(t) - f(t_{2k-1}^n)| + |f_n(t_{2k-1}^n) - f_n(t)| \leq \varepsilon + \eta_k^n \leq 2\varepsilon.$$

This bound holds on any of the intervals $[t_{2k-2}^n, t_{2k}^n]$. Hence $||f - f_n|| \to 0$. □

Corollary 2.2 *For arbitrary $f \in C[0,1]$ we have*

$$f = \sum_{m=0}^{\infty} \sum_{k \in I(m)} \frac{\eta_k^m}{\sigma_m} S_k^m, \tag{2.4}$$

where the infinite sum converges in the sup norm.

2.2 Existence of Brownian motion

Suppose that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given. A stochastic process X with time set $\mathcal{T}$ is a collection of random variables $\{X(t), t \in \mathcal{T}\}$ (so all the $X(t)$ are measurable functions on Ω). One may alternatively view X as a function $(\omega, t) \mapsto X(\omega, t)$ on $\Omega \times \mathcal{T}$, so that every $X(t)$ is defined by $X(t) : \omega \to X(\omega, t)$. For fixed ω we consider the functions $t \mapsto X(\omega, t)$. These functions called the sample paths of X.

Definition 2.3 A standard Brownian motion, also called Wiener process, is a stochastic process W with time index set $\mathcal{T} = [0, \infty)$ with the following properties.

(i) $W(0) = 0$;

(ii) the increments $W(t) - W(s)$ have a normal $N(0, t-s)$ distribution for all $t > s$;

(iii) the increments $W(t) - W(s)$ are independent of all $W(u)$ with $u \leq s < t$;

(iv) the paths of W are continuous functions.

Remark 2.4 One can show that part (iii) of Definition 2.3 is equivalent to the following. For all finite sequences $0 \leq t_0 \leq \ldots \leq t_n$ the random variables $W(t_k) - W(t_{k-1})$ $(k = 1, \ldots, n)$ are independent.

Having posed the definition of Brownian motion, we ask whether it exists in a precise mathematical sense. Clearly, the underlying outcome space Ω has to be big enough (on an Ω which contains only finitely many elements it is certainly not possible to define a Brownian motion). Since in the definition we required continuity of the paths, a good candidate for Ω should be the set $C[0,\infty)$ of continuous functions on $[0,\infty)$. In the sequel we will see that it is possible to define Brownian motion on a space Ω that is such all the paths become continuous, so that we can identify this Ω with $C[0,\infty)$. In the construction below, a different Ω will be used, a countable product of copies of $\mathbb{R}$. This suffices, since a continuous function is fixed as soon as we know its values on a countable dense subset of $\mathbb{R}$.

The method we use is a kind of converse of the interpolation scheme of Section 2.1. We will define what is going to be Brownian motion recursively on the time interval $[0,1]$ by attributing values at the dyadic numbers in $[0,1]$. A crucial part of the construction is the following fact. Supposing that we have shown that Brownian motion exists we consider the random variables $W(s)$ and $W(t)$ with $s < t$. Draw independent of these random variables a random variable ξ with a standard normal distribution and define $Z = \frac{1}{2}(W(s)+W(t)) + \frac{1}{2}\sqrt{t-s}\xi$. Then Z also has a normal distribution, whose expectation is zero and whose variance can be shown to be $\frac{1}{2}(t+s)$ (this is Exercise 2.1). Hence Z has the same distribution as $W(\frac{1}{2}(t+s))$! This fact lies at the heart of the construction of Brownian motion by a kind of 'inverse interpolation' that we will present now.

Let, as in Section 2.1, $I(0) = \{1\}$ and $I(n)$ be the set $\{1,\ldots,2^{n-1}\}$ for $n \geq 1$. Take a sequence of independent standard normally distributed random variables ξ_k^n that are all defined on some probability space Ω with $k \in I(n)$ and $n \in \mathbb{N}\cup\{0\}$ (it is a result in probability theory that one can take for Ω a countable product of copies of $\mathbb{R}$, endowed with a product σ-algebra and a product measure). With the aid of these random variables we are going to construct a sequence of continuous processes W^n as follows. Let, also as in Section 2.1, $\sigma_n = 2^{-\frac{1}{2}(n+1)}$. Put

$$W^0(t) = t\xi_1^0.$$

For $n \geq 1$ we get the following recursive scheme

$$W^n(t_{2k}^n) = W^{n-1}(t_k^{n-1}) \tag{2.5}$$

$$W^n(t_{2k-1}^n) = \frac{1}{2}\left(W^{n-1}(t_{k-1}^{n-1}) + W^{n-1}(t_k^{n-1})\right) + \sigma_n\xi_k^n. \tag{2.6}$$

For other values of t we define $W^n(t)$ by linear interpolation between the values of W^n at the points t_k^n. As in Section 2.1 we can use the Schauder functions for a compact expression of the random functions W^n. We have

$$W^n(t) = \sum_{m=0}^{n} \sum_{k\in I(m)} \xi_k^m S_k^m(t). \tag{2.7}$$

Note the similarity of this equation with (2.3). The main result of this section is

Theorem 2.5 *For almost all ω the functions $t \mapsto W^n(\omega,t)$ converge uniformly to a continuous function $t \mapsto W(\omega,t)$ and the process $W : (\omega,t) \to W(\omega,t)$ is Brownian motion on $[0,1]$.*

Proof We start with the following result. If Z has a standard normal distribution and $x > 0$, then (Exercise 2.2)

$$\mathbb{P}(|Z| > x) \leq \sqrt{\frac{2}{\pi}} \frac{1}{x} \exp(-\frac{1}{2}x^2). \tag{2.8}$$

Let $\beta_n = \max_{k \in I(n)} |\xi_k^n|$. Then $b_n := \mathbb{P}(\beta_n > n) \leq 2^{n-1}\sqrt{\frac{2}{\pi}} \frac{1}{n} \exp(-\frac{1}{2}n^2)$. Observe that $\sum_n b_n$ is convergent and that hence by virtue of the Borel-Cantelli lemma (Exercise 2.3) $\mathbb{P}(\limsup\{\beta_n > n\}) = 0$, and then $\mathbb{P}(\liminf\{\beta_n \leq n\}) = 1$.

Hence $\tilde{\Omega} := \liminf\{\beta_n \leq n\}$ is a subset of Ω with $\mathbb{P}(\tilde{\Omega}) = 1$, and such that for all $\omega \in \tilde{\Omega}$ there exists a natural number $n(\omega)$ with the property that all $|\xi_k^n(\omega)| \leq n$ if $n \geq n(\omega)$ and $k \in I(n)$. Consequently, for $\omega \in \tilde{\Omega}$ and for $p > n \geq n(\omega)$ we have

$$\sup_t |W^n(\omega, t) - W^p(\omega, t)| \leq \sum_{m=n+1}^{\infty} m\sigma_m < \infty. \tag{2.9}$$

This shows that the sequence $W^n(\omega, \cdot)$ with $\omega \in \tilde{\Omega}$ is Cauchy in $C[0, 1]$, so that it converges to a continuous limit, which we call $W(\omega, \cdot)$. For ω's not in $\tilde{\Omega}$ we define $W(\omega, \cdot) = 0$. So we now have continuous functions $W(\omega, \cdot)$ for all ω with the property $W(\omega, 0) = 0$.

As soon as we have verified properties (ii) and (iii) of Definition 2.3 we know that W is a Brownian motion. We will verify these two properties at the same time by showing that all increments $\Delta_j := W(t_j) - W(t_{j-1})$ with $t_j > t_{j-1}$ are independent $N(0, t_j - t_{j-1})$ distributed random variables. Thereto (cf. Section A.3) we will prove that the characteristic function $\mathbb{E} \exp(\mathrm{i} \sum_j \lambda_j \Delta_j)$ is equal to $\exp(-\frac{1}{2} \sum_j \lambda_j^2 (t_j - t_{j-1}))$.

We use an important property of the Haar functions: they form a Complete Orthonormal System of $\mathcal{L}^2[0, 1]$ (see Exercise 2.5). So every function $f \in \mathcal{L}^2[0, 1]$ has the representation $f = \sum_{n,k} \langle f, H_k^n \rangle H_k^n = \sum_{n=0}^{\infty} \sum_{k \in I(n)} \langle f, H_k^n \rangle H_k^n$, where $\langle \cdot, \cdot \rangle$ denotes the inner product of $\mathcal{L}^2[0, 1]$ and where the infinite sum is convergent in $\mathcal{L}^2[0, 1]$. As a result we have for any two functions f and g in $\mathcal{L}^2[0, 1]$ the Parseval identity $\langle f, g \rangle = \sum_{n,k} \langle f, H_k^n \rangle \langle g, H_k^n \rangle$. Taking the specific choice $f = \mathbf{1}_{[0,t]}$ and $g = \mathbf{1}_{[0,s]}$ results in $\langle \mathbf{1}_{[0,t]}, H_k^n \rangle = S_k^n(t)$ and

$$t \wedge s = \langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \rangle = \sum_{n,k} S_k^n(t) S_k^n(s). \tag{2.10}$$

Since for all fixed t we have $W^n(t) \to W(t)$ a.s., we have $\mathbb{E} \exp(\mathrm{i} \sum_j \lambda_j \Delta_j) = \lim_{n \to \infty} \mathbb{E} \exp(\mathrm{i} \sum_j \lambda_j \Delta_j^n)$ with $\Delta_j^n = W^n(t_j) - W^n(t_{j-1})$, in view of Theorem A.3(ii).

Note that $\sum_j \lambda_j \Delta_j^n = \sum_{m \leq n} \sum_k [\sum_j \lambda_j (S_k^m(t_j) - S_k^m(t_{j-1}))] \xi_k^m)$ is by independence of the standard normal ξ_k^m again normally distributed with mean zero and variance $\sum_{m \leq n} \sum_k [\sum_j \lambda_j (S_k^m(t_j) - S_k^m(t_{j-1}))]^2$. Recall that $\mathbb{E} \exp(\mathrm{i}Y) = \exp(-\frac{1}{2}\sigma^2)$ if Y has a $N(0, \sigma^2)$ distribution. We now compute

$$\begin{aligned}
\mathbb{E} \exp(\mathrm{i} \sum_j \lambda_j \Delta_j^n) &= \mathbb{E} \exp(\mathrm{i} \sum_{m \leq n} \sum_k [\sum_j \lambda_j (S_k^m(t_j) - S_k^m(t_{j-1}))] \xi_k^m) \\
&= \exp(-\frac{1}{2} \sum_{m \leq n} \sum_k (\sum_j \lambda_j (S_k^m(t_j) - S_k^m(t_{j-1}))^2).
\end{aligned}$$

Write now the triple sum in the exponential as

$$\sum_{m\le n}\sum_k\sum_{i,j}\lambda_j\lambda_i(S^m_k(t_j)-S^m_k(t_{j-1}))(S^m_k(t_i)-S^m_k(t_{i-1})),$$

and by swapping the summation order as

$$\sum_{i,j}\lambda_j\lambda_i\sum_{m\le n}\sum_k(S^m_k(t_j)-S^m_k(t_{j-1}))(S^m_k(t_i)-S^m_k(t_{i-1})).$$

Fix i and j and consider the sum over $m \le n$ and $k \in I(m)$ of

$$S^m_k(t_j)S^m_k(t_i)-S^m_k(t_{j-1})S^m_k(t_i)-S^m_k(t_j)S^m_k(t_{i-1})+S^m_k(t_{j-1})S^m_k(t_{i-1}).$$

This sum converges by virtue lf (2.10) for $n \to \infty$ to

$$t_j\wedge t_i - t_{j-1}\wedge t_i - t_j\wedge t_{i-1} + t_{j-1}\wedge t_{i-1}.$$

Considering the different cases $i < j$, $i > j$ and $i = j$, one sees that only the latter case gives a nonzero contribution, which is equal to $t_j - t_{j-1}$. Hence the expectation $\mathbb{E}\exp(\mathrm{i}\sum_j\lambda_j\Delta^n_j)$ converges to $\exp(-\frac{1}{2}\sum_j\lambda_j^2(t_j-t_{j-1}))$ as $n\to\infty$. This completes the proof. □

Having constructed Brownian motion on $[0,1]$, we proceed to show that it also exists on $[0,\infty)$. Take for each $n \in \mathbb{N}$ a probability space $(\Omega_n,\mathcal{F}_n,\mathbb{P}_n)$ that supports a Brownian motion W^n on $[0,1]$. Consider then $\Omega=\prod_n\Omega_n$, $\mathcal{F}=\prod_n\mathcal{F}_n$, $\mathbb{P}=\prod_n\mathbb{P}_n$ and the product probability space $(\Omega,\mathcal{F},\mathbb{P})$. Take for granted that the infinite products make sense and that $(\Omega,\mathcal{F},\mathbb{P})$ is indeed a probability space. On this product space the Brownian motions W^n are independent by construction, since $\mathbb{P}$ is a product measure. Let $\omega=(\omega_1,\omega_2,\ldots)$ and define then

$$W(\omega,t)=\sum_{n\ge 0}\mathbf{1}_{[n,n+1)}(t)\left(\sum_{k=1}^n W_k(\omega_k,1)+W_{n+1}(\omega_{n+1},t-n)\right). \tag{2.11}$$

This obviously defines a process with continuous paths and for all t the random variable $W(t)$ is the sum of independent normal random variables. It is not hard to establish that the process W defined by (2.11) has independent increments; this is Exercise 2.8. It is (almost) immediate that $\mathbb{E}W(t)=0$ and that $\mathbb{V}\mathrm{ar}\,W(t)=t$.

Apart from standard Brownian motion that we just defined, we will also consider simple transformations of it. Here is a list of processes that we will often encounter. We always use W to denote standard Brownian motion.

Brownian motion with variance parameter σ^2 is the process σW. Brownian motion with linear drift is a process X with $X(t)=bt+\sigma W(t)$, where b and σ are real constants. Note that also X has independent increments and that $\mathbb{E}X(t)=bt$ and $\mathbb{V}\mathrm{ar}\,X(t)=\sigma^2 t$.

Very important in this course is geometric Brownian motion. This is a process S with $S(t)=\exp(X(t))$, where X is a Brownian motion with drift, as we just defined it. Different from Brownian motion, here the ratios $S(t)/S(u)$ $(t>u)$ are independent from the past of the process before time u. Moreover we have (this is Exercise 2.9)

$$\mathbb{E}S(t)=\exp(bt+\frac{1}{2}\sigma^2 t) \tag{2.12}$$

$$\mathbb{V}\mathrm{ar}\,S(t)=\exp(2bt+\sigma^2 t)(\exp(\sigma^2 t)-1). \tag{2.13}$$

2.3 Properties of Brownian motion

Although we have defined Brownian motion as a process with continuous paths, the paths are very irregular. For instance, they are (almost surely) of *unbounded variation* over nonempty intervals. It is possible to prove this directly, but it is also a rather simple consequence, Corollary 2.8, of Proposition 2.6. The content of the proposition is that Brownian motion is of bounded *quadratic variation* on compact intervals. Let us introduce some notation. Consider an interval $[0,t]$ and let $\Pi_n = \{t_0^n, \ldots, t_{k_n}^n\}$ be a partition of it with $0 = t_0^n < \cdots < t_{k_n}^n = t$ and denote by μ_n its mesh: $\mu_n = \max\{t_j^n - t_{j-1}^n : j = 1, \ldots, k_n\}$. Let $V_n^2 = \sum_{j=1}^{k_n}(W(t_j^n) - W(t_{j-1}^n))^2$.

Proposition 2.6 *If $\mu_n \to 0$, then $\mathbb{E}(V_n^2 - t)^2 \to 0$. If moreover $\sum_{n=1}^\infty \mu_n < \infty$, then also $V_n^2 \to t$ a.s.*

Proof V_n^2 is the sum of squares of independent random variables that are $N(0, t_j^n - t_{j-1}^n)$ distributed. Hence $\mathbb{E}(W(t_j^n) - W(t_{j-1}^n))^2 = t_j^n - t_{j-1}^n$, so that $\mathbb{E}V_n^2 = t$. We proceed by computing $\mathbb{V}\text{ar}\, V_n^2$. Recall to that end that $\mathbb{V}\text{ar}\,(X^2) = 2\sigma^4$ for X having a $N(0,\sigma^2)$ distribution.

We then have $\mathbb{V}\text{ar}\, V_n^2 = \sum_{j=1}^{k_n} \mathbb{V}\text{ar}\,(W(t_j^n) - W(t_{j-1}^n)^2 = 2\sum_{j=1}^{k_n}(t_j^n - t_{j-1}^n)^2$ and this is less than or equal to $2\mu_n \sum_{j=1}^{k_n}(t_j^n - t_{j-1}^n) = 2\mu_n t$. This proves the first assertion of the theorem. Note that by Chebychev's inequality we also have $V_n^2 \to t$ in probability. Indeed $\mathbb{P}(|V_n^2 - t| > \varepsilon) \leq \frac{1}{\varepsilon^2}\mathbb{V}\text{ar}\, V_n^2 \to 0$.

To prove almost sure convergence we use the Borel-Cantelli lemma (again). With $E_n = \{|V_n^2 - t| > \varepsilon\}$ we have under the stipulated condition $\sum_{n=1}^\infty \mathbb{P}(E_n) < \infty$ and hence $\mathbb{P}(\limsup E_n) = 0$, equivalently $\mathbb{P}(\liminf E_n^c) = 1$. Since for every $\omega \in \liminf E_n$ we can find $N(\omega, \varepsilon)$ such that for all $n > N(\omega,\varepsilon)$ we have $|V_n^2(\omega) - t| \leq \varepsilon$, which completes the proof. □

Remark 2.7 The proposition has the interpretation that the paths of Brownian motion over an interval $[0,t]$ have quadratic variation t, denoted $\langle W \rangle_t = t$. Note that the quadratic variation is the same for all paths, since it doesn't depend on ω, unlike the 'prelimiting' random variables V_n^2.

Corollary 2.8 *With $V_n^1 = \sum_{j=1}^{k_n} |W(t_j^n) - W(t_{j-1}^n)|$ we have $V_n^1 \to \infty$ a.s., if $\mu_n \to 0$.*

Proof This follows from Exercise 2.10. □

We can also say something more precise about the continuity of the sample paths of Brownian motion.

Proposition 2.9 *The paths of Brownian motion are a.s. Hölder continuous of any order γ with $\gamma < \frac{1}{2}$, i.e. for almost every ω and for every $\gamma \in (0, \frac{1}{2})$ there exists a $C > 0$ such that $|W(\omega,t) - W(\omega,s)| \leq C|t-s|^\gamma$ for all t and s.*

Proof Exercise 2.12. □

Having established a result on the continuity of the paths of Brownian motion, we now turn to the question of differentiability of these paths. Proposition 2.10 says that they are nowhere differentiable. To get some feeling for this non-differentiability, consider $(W(t+h) - W(t))/h$. It has a normal distribution with variance $\frac{1}{h}$. Hence for any positive real number N we have that $\mathbb{P}(|(W(t+h)-W(t))/h| \geq N) = 2(1-\Phi(N\sqrt{h})) \to 1$ as $h \to 0$. We cannot expect the difference quotient to have a limit in any reasonable sense.

Proposition 2.10 *Put* $D = \{\omega : t \mapsto W(\omega, t)$ *is differentiable at some* $s \in (0,1)\}$. *Then* D *is contained in a set of zero probability.*

Proof For positive integers j and k, let $A_{jk}(s)$ be the set

$$\{\omega : |W(\omega, s+h) - W(\omega, s)| \leq j|h|, \text{ for all } h \text{ with } |h| \leq 1/k\}$$

and $A_{jk} = \bigcup_{s\in(0,1)} A_{jk}(s)$. Then we have the inclusion $D \subset \bigcup_{jk} A_{jk}$. Fix j, k and s for a while, pick $n \geq 4k$, choose $\omega \in A_{jk}(s)$ and choose i such that $s \in (\frac{i-1}{n}, \frac{i}{n}]$. Note first for $l = 1,2,3$ the trivial inequalities $\frac{i+l}{n} - s \leq \frac{l+1}{n} \leq \frac{1}{k}$. The triangle inequality and $\omega \in A_{jk}(s)$ gives for $l = 1,2,3$

$$\begin{aligned}
&|W(\omega, \frac{i+l}{n}) - W(\omega, \frac{i+l-1}{n})| \\
&\leq |W(\omega, \frac{i+l}{n}) - W(\omega, s)| + |W(\omega, s) - W(\omega, \frac{i+l-1}{n})| \\
&\leq \frac{l+1}{n} j + \frac{l}{n} j = \frac{2l+1}{n} j.
\end{aligned}$$

It then follows that

$$A_{jk} \subset B_{jk} := \bigcap_{n\geq 4k} C_{nj},$$

where

$$C_{nj} = \bigcup_{i=1}^{n} \bigcap_{l=1,2,3} \{\omega : |W(\omega, \frac{i+l}{n}) - W(\omega, \frac{i+l-1}{n})| \leq \frac{2l+1}{n} j\}. \tag{2.14}$$

We proceed by showing that C_{nj} has probability tending to zero. We use the following auxiliary result: if X has a $N(0,\sigma^2)$ distribution, then $\mathbb{P}(|X| \leq x) < x/\sigma$ (Exercise 2.13). By the independence of the increments of Brownian motion the probability of the intersection in (2.14) is the product of the probabilities of each of the terms and this product is less then $105 j^3 n^{-3/2}$. Hence $\mathbb{P}(C_{nj}) \leq 105 j^3 n^{-1/2}$, which tends to zero for $n \to \infty$. Consequently, $\mathbb{P}(B_{jk}) = 0$ and then also $\mathbb{P}(\bigcup_{j,k} B_{jk}) = 0$. The conclusion now follows from $D \subset \bigcup_{j,k} A_{jk} \subset \bigcup_{j,k} B_{jk}$. □

All the above results are on properties of the paths of Brownian motion. We close this section by mentioning that Brownian motion has the important property of being a *Markov* process.

Proposition 2.11 *Let* $t, h > 0$ *and suppose that we know* $W(s)$ *for all* $s \leq t$. *The conditional distribution of* $W(t+h)$ *given all* $W(s), s \leq t$ *is the same as the conditional distribution of* $W(t+h)$ *given* $W(t)$ *only and it is determined by*

$$\mathbb{P}(W(t+h) \leq x | W(s), s \leq t) = \Phi(\frac{x - W(t)}{\sqrt{h}}).$$

Proof This fact is a straightforward consequence of the independent increments property of W. Write $W(t+h)$ as the sum of the independent random variables $W(t+h)-W(t)$ and $W(t)$ and recall that $W(t+h)-W(t)$ is also independent of all $W(s)$ with $s \leq t$. Since $W(t+h)-W(t)$ has the same distribution as $\sqrt{h}Z$ with Z a standard normal random variable (with distribution function Φ) we can compute, exploiting the mentioned independence and using Proposition A.12(v),

$$\mathbb{P}(W(t+h) \leq x | W(s), s \leq t) =$$
$$\mathbb{P}(W(t+h) - W(t) \leq x - W(t) | W(s), s \leq t) = \Phi(\frac{x - W(t)}{\sqrt{h}}).$$

In the same way we have $\mathbb{P}(W(t+h) \leq x | W(t)) = \Phi(\frac{x-W(t)}{\sqrt{h}})$. The assertion follows. □

Remark 2.12 Proposition 2.11 has as a consequence that conditional expectations of functions of $W(t+h)$ given $W(s)$, $s \leq t$ reduce to conditional expectations given $W(t)$. E.g. suppose that f is measurable and bounded, then for given $0 \leq t \leq t+h = T$ one has $\mathbb{E}[f(W(T))|W(s), s \leq t] = \mathbb{E}[f(W(T))|W(t)]$. The latter is a function of $W(t)$ (and t), $v(t, W(t))$ say.

2.4 Exercises

2.1 Show that the random variable Z on page 18 has a normal distribution with mean zero and variance equal to $\frac{1}{2}(s+t)$.

2.2 Prove inequality (2.8).

2.3 Prove the following part of the Borel-Cantelli lemma:
If (A_k) is a sequence of events with $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$, then $\mathbb{P}(\limsup A_k) = 0$. Here $\limsup A_k = \bigcap_{n \geq 1}^{\infty} \bigcup_{k \geq n}^{\infty} A_k$. *Hint:* With $V_m = \bigcup_{k \geq m} A_k$ the sequence (V_m) is decreasing and for all N it holds that $\mathbb{P}(\limsup A_k) \leq \mathbb{P}(\bigcap_{m=1}^{N} V_m)$.

2.4 Show that $C[0,1]$ is a complete normed space under the sup norm. (Use completeness of $\mathbb{R}$.)

2.5 The Haar functions form a Complete Orthonormal System in $\mathcal{L}^2[0,1]$. Show first that the Haar functions are orthonormal. To prove that the system is complete, you argue as follows. Let f be orthogonal to all $H_{k,n}$ and set $F = \int_0^{\cdot} f(u)du$. Show that F is zero in all $t = k2^{-n}$, and therefore zero on the whole interval. Conclude that $f = 0$ a.e. (The set $\{f \neq 0\}$ has Lebesgue measure zero.)

2.6 Let (X_n) be a sequence of random k-vectors that in the $\mathcal{L}^2$-sense converge to a random vector X, i.e. $\mathbb{E}||X_n - X||^2 \to 0$, where $||\cdot||$ denotes the Euclidean norm.

(a) Show that $\mathbb{E}X_n \to \mathbb{E}X$ and $\mathbb{C}\mathrm{ov}(X_n) \to \mathbb{C}\mathrm{ov}(X)$.
(b) Assume then that all X_n are (multivariate) normal. Show that also X is (multivariate) normal.

2.7 Consider the processes W^n of Section 2.2. Let $t_1, \ldots, t_k \in [0,1]$. Show that the sequence of random vectors $(W^n(t_1), \ldots, W^n(t_k))$ in the $\mathcal{L}^2$-sense converges to $(W(t_1), \ldots, W(t_k))$, i.e. $\sum_{i=1}^{k} \mathbb{E}(W^n(t_i) - W(t_i))^2 \to 0$. (*Hint:* this sequence is Cauchy in $\mathcal{L}^2$. Identify the limit.)

2.8 Show that the increments of the process W defined by Equation (2.11) are independent.

2.9 Show the validity of equations (2.12) and (2.13).

2.10 The p-th order variation of a function $f : [0,1] \to \mathbb{R}$ over a partition $\Pi = \{t_0, \ldots, t_n\}$ of $[0,1]$ with $0 = t_0 < t_1 < \ldots < t_n = 1$ is defined as $V^p(f;\Pi) = \sum_{i=1}^n |f(t_i) - f(t_{i-1})|^p$. Put $V^p(f) = \lim V^p(f;\Pi)$, where the limit is taken over partitions with mesh tending to zero.

(a) Let $p = 1$ and $f \in C^1[0,1]$, so with bounded (left/right in the endpoints) derivative. Argue (use Riemann sums) that $V^1(f) = \int_0^1 |f'(t)| \, \mathrm{d}t$.

(b) Prove: if $f \in C[0,1]$ and $V^p(f)$ is positive and finite for some $p > 0$, then for every $p' < p$ the variation $V^{p'}(f) = \infty$, whereas $V^{p'}(f) = 0$ for every $p' > p$.

2.11 Give a simple example of a function f such that $0 < V^p(f) < \infty$ for all $p > 0$. [*Hint:* You should not be in the situation of Exercise 2.10.]

2.12 Prove Proposition 2.9. *Hint:* Use that $|S_k^m(t) - S_k^m(s)| \leq 2^{\frac{1}{2}(m-1)}|t-s|$ and an inequality similar to (2.9).

2.13 Show that for a random variable X with a $N(0,\sigma^2)$ distribution it holds that $\mathbb{P}(|X| \leq x) < x/\sigma$, for $x, \sigma > 0$.

2.14 Let X be a Brownian motion with linear drift, $X(t) = bt + \sigma W(t)$, with $b \in \mathbb{R}$. The quadratic variation of a process X over an interval $[0,T]$ will be denoted by $\langle X \rangle_T$, defined as the limit in probability of the $V_n^2 = \sum_{j=1}^{k_n} (X(t_j^n) - X(t_{j-1}^n))^2$, just as at the beginning of Section 2.3. Show that the quadratic variation $\langle X \rangle_T$ of X over $[0,T]$ is equal to $\sigma^2 T = \sigma^2 \langle W \rangle_T$. [Note that the linear deterministic term bt in $X(t)$ plays no role in the expression for $\langle X \rangle_T$.]

2.15 Show that for W^n defined by (2.7) it holds that

$$\mathbb{C}\mathrm{ov}(W^n(t), W^n(s)) = \sum_{m=0}^{n} \sum_{k \in I(m)} S_k^m(t) S_k^m(s)),$$

and that this converges to $t \wedge s = \mathbb{C}\mathrm{ov}(W(t), W(s))$ for $n \to \infty$.

3 The heat equation

The heat equation, the subject of this section, is intimately related to Brownian motion. We show some connections, that prove to be useful in later sections. We do not give proofs of all results, this would carry us to far away from the subject of this course. For this we refer to a course on Partial Differential Equations, although there are also proofs that use advanced probabilistic techniques instead of ordinary analysis.

3.1 Some theory

Knowing that $W(t)$ has a $N(0,t)$ distribution for $t > 0$ if W is a Brownian motion, we can write down the density $p(t,\cdot)$ of $W(t)$. It is given by

$$p(t,x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}).$$

Computing the partial derivatives of p we can verify that p satisfies for all $x \in \mathbb{R}$ and $t > 0$ a partial differential equation, the *heat equation*,

$$\frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\frac{\partial u}{\partial x^2}(t,x). \tag{3.1}$$

We will encounter many partial differential equations below that are based on the heat equation, so it is important to study properties of solutions to this equation.

Let f be a sufficiently well behaving function (below we shall make this precise) and define $u(t,x) = \mathbb{E}f(W(t)+x)$. Then we readily see that $u(0,x) = f(x)$. Moreover, we have for $t > 0$ the explicit formula

$$u(t,x) = \int_{-\infty}^{\infty} f(z)p(t,x-z)\,\mathrm{d}z. \tag{3.2}$$

Now suppose that it is allowed to interchange integration and (partial) differentiation in (3.2). Then we obtain $\frac{\partial u}{\partial t}(t,x) = \int_{-\infty}^{\infty} f(z)\frac{\partial p}{\partial t}(t,x-z)\,\mathrm{d}z$ and $\frac{\partial u}{\partial x^2}(t,x) = \int_{-\infty}^{\infty} f(z)\frac{\partial p}{\partial x^2}(t,x-z)\,\mathrm{d}z$. From these expressions for the partial derivatives of u we see that also u satisfies the heat equation. Clearly, a rigorous mathematical treatment involves a precise condition on f. Here it is.

Condition 3.1 The function f is a Borel-measurable function and for some $a > 0$ the integral $\int_{-\infty}^{\infty} \exp(-ax^2)|f(x)|\,\mathrm{d}x$ is finite.

Proposition 3.2 *Let f satisfy Condition 3.1 and let $u : [0,\frac{1}{2a}) \times \mathbb{R} \to \mathbb{R}$ be defined by $u(t,x) = \mathbb{E}f(x+W(t))$. Then u has partial derivatives of all orders on $(0,\frac{1}{2a}) \times \mathbb{R}$, that can be obtained by differentiation under the integral sign and u satisfies the heat equation. If moreover f is continuous at x, then u is continuous at $(0,x)$.*

Proof Since we are mainly interested in showing that u satisfies the heat equation, we only consider the corresponding derivatives, for other derivatives the proof is similar. Fix $\beta, \varepsilon > 0$ and let $t_1 = 1/2(a+\varepsilon)$, $0 < t_0 < t_1$. Let $E = (t_0, t_1) \times (-\beta, \beta)$. First we compute the partial derivatives

$$\begin{aligned}\frac{\partial p}{\partial t}(t,x) &= \frac{1}{2}(\frac{x^2}{t^2} - \frac{1}{t})p(t,x)\\ \frac{\partial p}{\partial x}(t,x) &= -\frac{x}{t}p(t,x)\\ \frac{\partial p}{\partial x^2}(t,x) &= (\frac{x^2}{t^2} - \frac{1}{t})p(t,x).\end{aligned}$$

Note that there exists a constant $C = C(a, \varepsilon, t_0)$, such that on E all these derivatives are bounded in absolute value by $C(x^2+1)p(t,x)$ for all $t \geq t_0$. Let $q(t,x)$ denote any of these partial derivatives and consider

$$\begin{aligned}|f(y)q(t,x-y)| &\leq C|f(y)|((x-y)^2+1)p(t,x-y)\\ &= C|f(y)|\exp(-ay^2)\times\\ &\qquad ((x-y)^2+1)\exp(ay^2)p(t,x-y). \end{aligned} \tag{3.3}$$

Let's concentrate on the last factor in (3.3), which is

$$((x-y)^2+1)\frac{1}{\sqrt{2\pi t}}\exp(ay^2 - \frac{(x-y)^2}{2t}).$$

We will show that it is bounded for all $y \in \mathbb{R}$ and $(t,x) \in E$. First we have for the exponent

$$\begin{aligned}ay^2 - \frac{(x-y)^2}{2t} &\leq ay^2 - (a+\varepsilon)(x-y)^2\\ &= -\varepsilon y^2 - (a+\varepsilon)x^2 + 2(a+\varepsilon)xy\\ &\leq 2(a+\varepsilon)\beta|y| - \varepsilon y^2,\end{aligned}$$

and hence its exponential quickly tends to zero if $|y| \to \infty$. Note that $(x-y)^2 \leq 2(\beta^2 + y^2)$ if $|x| < \beta$. From this we obtain that $((x-y)^2+1)\exp(ay^2)p(t,x-y)$ is bounded by a constant, D say, if $(t,x) \in E$ and $y \in \mathbb{R}$. Hence

$$|f(y)q(t,x-y)| \leq CD\exp(-ay^2)|f(y)|,$$

which is integrable by virtue of Condition 3.1. We then obtain that $(t,x) \mapsto \int f(y)q(t,x-y)\,\mathrm{d}y$ is continuous in (t,x) on E by the dominated convergence theorem, and then also on $(0, 1/2a) \times \mathbb{R}$ by letting $\beta \to \infty$ and $\varepsilon, t_0 \to 0$.

Let us now focus on the partial derivative of u w.r.t. x. By the mean value theorem we have with $q = \frac{\partial p}{\partial x}$,

$$\begin{aligned}\frac{1}{h}(u(t,x+h) - u(t,x)) &= \int_{-\infty}^{\infty} f(y)\frac{1}{h}(p(t,x+h-y) - p(t,x-y))\,\mathrm{d}y\\ &= \int_{-\infty}^{\infty} f(y)q(t,\theta_h(t,x,y))\,\mathrm{d}y,\end{aligned} \tag{3.4}$$

with $\theta_h(t,x,y)$ between $x-y$ and $x+h-y$. Letting $h \to 0$ and invoking the just proved continuity result of the integral in (3.4), we obtain

$$\frac{\partial u}{\partial x}(t,x) = \int_{-\infty}^{\infty} f(y)q(t,x-y)\,\mathrm{d}y,$$

which is what we had to show. The analogous results for the other derivatives can be derived similarly, from which it follows that u satisfies (3.1) for $(t,x) \in (0, \frac{1}{2a}) \times \mathbb{R}$.

We now show the continuity of u in $(0, x_0)$, where x_0 is a continuity point of f. Assume without loss of generality (why?) that $f(x_0) = 0$. Then we have to show $u(t,x) \to 0$ if $(t,x) \to (0, x_0)$.

Fix $\varepsilon > 0$, then there is $\delta > 0$ such that $|f(y)| \le \varepsilon$ for $|y - x_0| < \delta$. We will show that $|u(t,x)|$ is small for $0 < t < 1/2a$, $t \to 0$, and $x \in (x_0 - \frac{1}{2}\delta, x_0 + \frac{1}{2}\delta)$. Obviously we have

$$|u(t,x)| \le \varepsilon + \int_{-\infty}^{x_0-\delta} |f(y)| p(t, x-y)\,\mathrm{d}y + \int_{x_0+\delta}^{\infty} |f(y)| p(t, x-y)\,\mathrm{d}y. \quad (3.5)$$

We treat the latter integral in detail, the other one can be dealt with similarly. Write it as $\frac{1}{\sqrt{2\pi t}} \int_{x_0+\delta}^{\infty} |f(y)| \exp(-ay^2) \times \exp(ay^2 - \frac{(x-y)^2}{2t})\,\mathrm{d}y$ and look at $h(y) = ay^2 - \frac{(x-y)^2}{2t}$ for $y \in \mathbb{R}$. For $t < 1/2a$ the function h is maximal at $y_0 = \frac{x}{1-2at} \to x$, when $t \to 0$. For small enough t we have $y_0 < x_0 + \frac{1}{2}\delta$, which lies outside the integration interval $(x_0 + \delta, \infty)$. Hence it follows that, for small t, h is decreasing on $(x_0 + \delta, \infty)$, and so h is maximal at the boundary point $x_0 + \delta$ with value $a(x_0+\delta)^2 - \frac{(x_0+\delta-x)^2}{2t}$, which is less than $a(x_0+\delta)^2 - \frac{\delta^2}{8t}$. Hence,

$$\int_{x_0+\delta}^{\infty} |f(y)| p(t, x-y)\,\mathrm{d}y \le \int |f(y)| \exp(-ay^2)\,\mathrm{d}y \times \exp(a(x_0+\delta)^2) p(t, \frac{1}{2}\delta),$$

which tends to 0 for $t \to 0$, as $p(t, \frac{1}{2}\delta)$ then tends to zero. □

The consequence of Proposition 3.2 is that every continuous f that satisfies Condition 3.1 gives through Equation (3.2) a solution u of the heat equation with $u(0, \cdot) = f$. Two questions arise: are these the only possible solutions of the heat equation and are solutions unique? The answer is no in general, but by imposing some regularity conditions on the solution we get what we want. We state a theorem that gives affirmative answers to these questions.

Theorem 3.3 *(i) Let u_1 and u_2 be two functions in $C^{1,2}((0,T) \times \mathbb{R})$ that solve the heat equation and such that are such that* $\sup\{\exp(-ax^2)|u_i(t,x)| : (t,x) \in (0,T) \times \mathbb{R}\} < \infty$ *for some $a > 0$ ($i = 1,2$) and such that* $\lim_{t\downarrow 0, y \to x} u_1(t,y) = \lim_{t\downarrow 0, y\to x} u_2(t,y)$ *for all x. Then u_1 and u_2 coincide.*
(ii) Let u be a nonnegative function in $C^{1,2}((0,T) \times \mathbb{R})$ with $0 \le T \le \infty$ that solves the heat equation and assume that for all $x \in \mathbb{R}$ the limit $f(x) :=$ $\lim_{t\downarrow 0, y\to x} u(t,y)$ *exists. Then u has for $t > 0$ the integral representation $u(t,x) = \int_{-\infty}^{\infty} f(y) p(t, x-y)\,\mathrm{d}y$.*

The first assertion of Theorem 3.3 is proved in the Appendix, Section A.8. Under the conditions of this theorem and Proposition 3.2, the function u in Equation (3.2) is the unique solution to the heat equation with boundary condition $u(0, \cdot) = f$. The second assertion of the theorem is also of importance for us, since we will mainly deal with nonnegative solutions of the heat (and related) equations. It says that (under the given conditions) every positive solution to the heat equation is of the form $u(t,x) = \mathbb{E} f(x + W(t))$ with W a Brownian motion. As a side remark we mention that a similar statement holds for bounded solutions.

Next to the heat equation (3.1) we also consider the *backward heat equation*

$$\frac{\partial v}{\partial t}(t,x) + \frac{1}{2}\frac{\partial v}{\partial x^2}(t,x) = 0. \tag{3.6}$$

We then have the following corollary to Theorem 3.3.

Corollary 3.4 *Let v be a nonnegative function in $C^{1,2}((0,T)\times\mathbb{R})$ with $0 \leq T < \infty$ that solves the backward heat equation and assume that for all $x \in \mathbb{R}$ the limit $\lim_{t\uparrow T, y\to x} v(t,y) =: f(x)$ exists. Then v has the integral representation $v(t,x) = \int_{-\infty}^{\infty} f(x+y)p(T-t,y)\,\mathrm{d}y$.*

Proof Simply define $u(t,x) = v(T-t,x)$ (T is finite here!) and apply Theorem 3.3. □

The probabilistic interpretation of solutions v of the backward heat equation is $v(t,x) = \mathbb{E}f(x + W(T-t))$. We give this a different appearance.

Proposition 3.5 *Under the conditions of Corollary 3.4 one has*

$$v(t,W(t)) = \mathbb{E}[f(W(T))|W(t)], \tag{3.7}$$

$$v(t,x) = \mathbb{E}[f(W(T))|W(t)=x]. \tag{3.8}$$

Proof Because $W(T-t)$ has the same distribution as $W(T) - W(t)$ we can also write $v(t,x) = \mathbb{E}f(x + W(T) - W(t))$. But then, using property (v) of Proposition A.12, we immediately arrive at (3.7) and (3.8). □

We will come back to this interpretation of v later on, but recall in the mean time Remark 2.12 in Section 2.3 where we have seen that

$$v(t,W(t)) = \mathbb{E}[f(W(T))|W(s), s \leq t].$$

Now we can add that v in the above display is the solution to the backward heat equation with terminal condition $v(T,x) = f(x)$.

Remark 3.6 The backward heat equation (3.6) is not immediately of prime interest from a financial point of view, as opposed to the Black-Scholes equation (1.31). In Exercises 3.5 and 3.6 you will study how these two equations are related.

3.2 Exercises

3.1 Let f be a bounded continuous function. Let u be as in Equation (3.2). Show (not referring to Proposition 3.2) that $\lim_{t\downarrow 0} u(t,x) = f(x)$ by application of the bounded convergence theorem.

3.2 Let u be given by (3.2). Show that under the conditions of Proposition 3.2 we have $u_t(t,x) = \int_{-\infty}^{\infty} f(y)p_t(t,x-y)\,\mathrm{d}y$.

3.3 The Hermite polynomials in two variables $H_n(t,x)$ ($n \geq 0$) are defined by $H_n(t,x) = \frac{\partial^n}{\partial\alpha^n}\exp(\alpha x - \frac{1}{2}\alpha^2 t)|_{\alpha=0}$.

(a) Explain that the $H_n(t,x)$ are polynomial in t and x and compute $H_n(t,x)$ for $n = 0, 1, 2, 3$.

(b) Show that $\frac{\partial}{\partial x} H_n(t,x) = nH_{n-1}(t,x)$ for $n \geq 1$ and that the H_n satisfy the backward heat equation.

3.4 Let $f(x) = \exp(\gamma x^2)$. Compute u defined by Equation (3.2) explicitly.

(a) For which points (t,x) is $u(t,x)$ well defined? Check that your answer is in agreement with Proposition 3.2.

(b) Is u a solution to the heat equation?

3.5 Let $v \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R}^2)$ and define u by $u(t,x) = v(t, e^{x - \frac{1}{2}\sigma^2 t})$.

(a) Suppose that v is a solution of

$$v_t(t,x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t,x) = 0. \tag{3.9}$$

Show that u satisfies

$$u_t(t,x) + \frac{1}{2}\sigma^2 u_{xx}(t,x) = 0. \tag{3.10}$$

(b) Conversely, let u be a solution of (3.10). For $x > 0$ we put $v(t,x) = u(t, \log x + \frac{1}{2}\sigma^2 t)$. Show that v satisfies (3.9).

3.6 Suppose that the function v satisfies for $t, x > 0$ the equation

$$v_t(t,x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t,x) = 0 \tag{3.11}$$

(a) Define $w(t,x) = v(t, xe^{-rt})e^{rt}$. Show that w satisfies

$$w_t(t,x) + \frac{1}{2}\sigma^2 x^2 w_{xx}(t,x) + rxw_x(t,x) - rw(t,x) = 0. \tag{3.12}$$

(b) Conversely, if w is a solution of (3.12) and $v(t,x) = w(t, xe^{rt})e^{-rt}$, then v is a solution to (3.11). *Remark*: Equation (3.12) is the *Black-Scholes partial differential equation* already encountered as (1.31), to which we will return in later sections.

3.7 Let W be a standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $a \in \mathbb{R}$ and let the process X be defined by $X_t = at + W_t$. Assume that f, a measurable function on $\mathbb{R}$, satisfies Condition 3.1. Define v by $v(t,x) := \mathbb{E}f(x + X_t)$. Show that v has all the properties of the function u mentioned in Proposition 3.2, except that v satisfies the partial differential equation $v_t = av_x + \frac{1}{2}v_{xx}$.

3.8 Let $p(t,x) = \frac{1}{\sqrt{2\pi t}} \exp(-x^2/2t)$. Show that all partial derivatives of p satisfy the heat equation. Is this in agreement with Theorem 3.3? Explain why (not).

4 Girsanov's theorem in a simple case

Having defined Brownian motion with drift at the end of Section 2.2, we study in this section in some detail how distributional properties of this process change under an *absolutely continuous change of probability measures*.

4.1 Measure transformations and Girsanov's theorem

Assume that on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a standard Brownian motion W is defined and we consider the process X given by

$$X(t) = x_0 + at + \sigma W(t), \tag{4.1}$$

where x_0 is a given real number (the initial value of the process X) and a and σ other real constants. It follows from the properties of Brownian motion, that for $t > s$ the increments $X(t) - X(s)$ have a normal distribution with mean $a(t-s)$ and variance $\sigma^2(t-s)$. Moreover, the increments over disjoint intervals are independent.

Since the paths of X are, like those of W, continuous functions, the process X induces a probability measure, also called the *law* of the process, on the space $C[0,T]$ of continuous functions on $[0,T]$, where T is some fixed terminal time. Of course we also need a σ-algebra on this space, we take the Borel σ-algebra that is induced by the sup norm. Clearly, the law depends on a and σ^2. What can we say of the relation between these laws, if we let the parameters a and σ vary? We will see that for fixed σ^2 the laws for varying a are *equivalent*, and for varying σ they are mutually *singular*. Let us define this terminology.

Let $\mathcal{L}^0_+$ be the set of all nonnegative random variables on a measurable space $(\Omega, \mathcal{F})$. If $(\Omega, \mathcal{F})$ has two probabilities $\mathbb{P}$ and $\mathbb{Q}$ defined on it, we say that $\mathbb{Q}$ is absolutely continuous w.r.t. $\mathbb{P}$, which is denoted by $\mathbb{Q} \ll \mathbb{P}$, if there exists a random variable Z such that $\mathbb{P}(Z \geq 0) = 1$ and such that

$$\mathbb{E}_{\mathbb{Q}} X = \mathbb{E}_{\mathbb{P}} ZX, \text{ for all } X \in \mathcal{L}^0_+, \tag{4.2}$$

where $\mathbb{E}_{\mathbb{Q}}$ denotes expectation under the probability measure $\mathbb{Q}$ and $\mathbb{E}_{\mathbb{P}}$ denotes expectation under the probability measure $\mathbb{P}$. For Z we also use the notation $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. Note that necessarily we have $\mathbb{E}_{\mathbb{P}} Z = 1$ (take $X = 1$). Every nonnegative random variable Z with $\mathbb{E}_{\mathbb{P}} Z = 1$ can be used to define a new probability measure $\mathbb{Q}$ through Equation (4.2), by taking X equal to the indicator of an event, i.e. for $F \in \mathcal{F}$ one defines

$$\mathbb{Q}(F) := \mathbb{E}_{\mathbb{P}} Z \mathbf{1}_F. \tag{4.3}$$

Countable additivity of a such defined $\mathbb{Q}$ follows from the monotone convergence theorem.

Let us also observe that $\mathbb{Q}(Z > 0) = 1$, because $\mathbb{Q}(Z = 0) = \mathbb{E}_{\mathbb{P}} \mathbf{1}_{\{Z=0\}} Z = \mathbb{E}_{\mathbb{P}} 0 = 0$ follows from (4.3). Is $\mathbb{P}(Z > 0) = 1$? In general not, see Exercise 4.3. But if $\mathbb{P}(Z = 0) = 0$, then we also have $\mathbb{P} \ll \mathbb{Q}$ and $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q} = 1/Z$. Indeed, since $\mathbb{P}(Z\frac{1}{Z} = 1) = 1$ we have $\mathbb{E}_{\mathbb{P}} X = \mathbb{E}_{\mathbb{P}} Z(\frac{1}{Z}X) = \mathbb{E}_{\mathbb{Q}} \frac{1}{Z} X$. Two measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ are called *equivalent* if both $\mathbb{P} \ll \mathbb{Q}$ and $\mathbb{Q} \ll \mathbb{P}$, in which case one writes $\mathbb{P} \sim \mathbb{Q}$.

Note that if $\mathbb{Q} \ll \mathbb{P}$, then $\mathbb{Q}(F) = 0$ if $\mathbb{P}(F) = 0$ as follows from (4.3) by taking $X = \mathbf{1}_F$. It is a main result (*Radon-Nikodym theorem*) in measure theory that also the converse is true. For this theorem we refer to a course in measure theory, although it is simple to prove it for discrete probability spaces, see Exercise 4.5. The random variables $\frac{d\mathbb{Q}}{d\mathbb{P}}$ and $\frac{d\mathbb{P}}{d\mathbb{Q}}$ are called Radon-Nikodym derivatives.

We will attack the following problem. Let X as above, $X(t) = at + W(t)$ (we take $x_0 = 0$ and $\sigma = 1$), defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Can we find a probability measure $\mathbb{Q}$ such that $\mathbb{Q} \ll \mathbb{P}$ and such that X becomes a Brownian motion under the probability $\mathbb{Q}$? The answer is yes, and we will show how to do this under the limitation that we restrict the time set to a bounded interval $[0, T]$. A non-dynamic version of this phenomenon is the content of Exercise 4.1, a forerunner of Proposition 4.2 below.

We will use in the remainder of this section the filtration $\mathbb{F} = \mathbb{F}^W = \{\mathcal{F}_t, t \in [0, T]\}$, where $\mathcal{F}_t = \sigma(W(s), s \leq t)$ is the smallest σ-algebra that makes all $W(s)$, for $s \leq t$, random variables. In particular we will use the σ-algebra $\mathcal{F}_T$. Define the random variable

$$Z(T) = \exp(-aW(T) - \frac{1}{2}a^2T). \tag{4.4}$$

Clearly we have $Z(T) \geq 0$ and moreover $\mathbb{E}_{\mathbb{P}} Z(T) = 1$ (see Exercise A.5). Hence we can use $Z(T)$ to define a new probability measure $\mathbb{Q}$ on $\mathcal{F}_T$ by

$$\mathbb{Q}(F) = \mathbb{E}_{\mathbb{P}} \mathbf{1}_F Z(T) \text{ for all } F \in \mathcal{F}_T. \tag{4.5}$$

Let us investigate some more properties of $Z(T)$. Let $t \in [0, T]$ be arbitrary and write

$$Z(T) = Z(t) \exp\left(-a(W(T) - W(t)) - \frac{1}{2}a^2(T - t)\right),$$

with $Z(t) = \exp(-aW(t) - \frac{1}{2}a^2t)$. Note that $Z(t)$ is $\mathcal{F}_t$-measurable. From properties of the normal distribution (Exercise A.5 again) we obtain

$$\mathbb{E}_{\mathbb{P}} \exp(-a(W(T) - W(t)) - \frac{1}{2}a^2(T - t)) = 1.$$

Note also that $\exp(-a(W(T) - W(t)) - \frac{1}{2}a^2(T - t)$ is independent of $\mathcal{F}_t$ by the property that Brownian motion has independent increments. Hence, using properties (iv) and (v) of Proposition A.12, we get

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[Z(T)|\mathcal{F}_t] &= Z(t)\mathbb{E}_{\mathbb{P}}[\exp(-a(W(T) - W(t)) - \frac{1}{2}a^2(T - t)|\mathcal{F}_t] \\ &= Z(t). \end{aligned} \tag{4.6}$$

Equation (4.6) tells us that the process $\{Z(t) : t \in [0, T]\}$ is a martingale (w.r.t. $\mathbb{F}$ under the probability measure $\mathbb{P}$).

Suppose that we want to play the same game, finding a new probability measure, but now on $\mathcal{F}_t$ instead of $\mathcal{F}_T$. We can do this by copying the above approach and work with $Z(t)$ as a Radon-Nikodym derivative. Let us say that this gives us a probability $\mathbb{Q}_t$. Instead, we could also simply restrict $\mathbb{Q}$ to the smaller σ-algebra $\mathcal{F}_t$, call the restriction $\mathbb{Q}|_t$. The two approaches turn out to be equivalent:

Proposition 4.1 *The two probabilities $\mathbb{Q}_t$ and $\mathbb{Q}|_t$ on $\mathcal{F}_t$ are the same.*

Proof Exercise 4.8. $\square$

We know that $\mathbb{E}_\mathbb{P} W(T) = 0$, but what is $\mathbb{E}_\mathbb{Q} W(T)$? We have

$$\begin{aligned}\mathbb{E}_\mathbb{Q} W(T) &= \mathbb{E}_\mathbb{P}[W(T)Z(T)] \\ &= \mathbb{E}_\mathbb{P}\left(W(T)\exp(-aW(T))\right)\exp(-\frac{1}{2}a^2T) \\ &= -aT,\end{aligned}$$

because

$$\mathbb{E}_\mathbb{P}\left(W(T)\exp(-aW(T))\right) = -aT\exp(\frac{1}{2}a^2T),$$

which follows from Exercise 4.6. Using the same arguments that led us to the result of (4.6), we can also compute $\mathbb{E}_\mathbb{Q} W(t) = -at$ (Exercise 4.7). We conclude that under the measure $\mathbb{Q}$ the process W is not a Brownian motion, since we get nonzero expectation. To remedy this problem we could consider the process $W^\mathbb{Q}$ defined by

$$W^\mathbb{Q}(t) = W(t) + at, \tag{4.7}$$

at least it has expectation zero. As a matter of fact we have the following simple version of what is known as *Girsanov's theorem.*

Proposition 4.2 *The process $W^\mathbb{Q}$ defined by (4.7) is a Brownian motion on the time domain $[0,T]$ under the probability $\mathbb{Q}$ defined by (4.5).*

Proof We only have to show that the increments of $W^\mathbb{Q}$ are independent (see Remark 2.4) and that $W^\mathbb{Q}(t) - W^\mathbb{Q}(s)$ has a normal distribution with variance $t-s$ under the (new) probability $\mathbb{Q}$. We can do both things at the same time by computing the joint characteristic function of any finite vector with elements $W^\mathbb{Q}(t_k) - W^\mathbb{Q}(t_{k-1})$, $k = 1,\ldots,n$, where $t_0 = 0 \leq t_1, \ldots, \leq t_n \leq T$. We compute, using Proposition 4.1 with $t = t_n$, upon noting that $Z(t_n) = \prod_{k=1}^n \exp\big(-a(W(t_k) - W(t_{k-1})) - \frac{1}{2}a^2(t_k - t_{k-1})\big)$,

$$\begin{aligned}&\mathbb{E}_\mathbb{Q}\exp(\mathrm{i}\sum_k \lambda_k(W^\mathbb{Q}(t_k) - W^\mathbb{Q}(t_{k-1}))) = \\ &\quad \mathbb{E}_\mathbb{P}[\exp(\mathrm{i}\sum_k \lambda_k(W^\mathbb{Q}(t_k) - W^\mathbb{Q}(t_{k-1}))\times \\ &\quad \exp(-a\sum_k(W(t_k) - W(t_{k-1}) - \frac{1}{2}a^2\sum_k(t_k - t_{k-1}))].\end{aligned}$$

Use now the definition of $W^\mathbb{Q}$ and a bit of rearranging the terms in the exponential to see that the last expectation becomes

$$\mathbb{E}_\mathbb{P}\exp(\sum_k(\mathrm{i}\lambda_k - a)(W(t_k) - W(t_{k-1})))\exp(\sum_k(a\mathrm{i}\lambda_k - \frac{1}{2}a^2)(t_k - t_{k-1})).$$

But under $\mathbb{P}$ the increments of W are independent and the computation of the expectation of the product reduces to the computation of the product of the expectations. Using that for a complex number z and $N(0,\sigma^2)$ distributed random variable X one has $\mathbb{E}e^{zX} = e^{\frac{1}{2}z^2\sigma^2}$ we get

$$\mathbb{E}_{\mathbb{P}} \exp(\sum_k (\mathrm{i}\lambda_k - a)(W(t_k) - W(t_{k-1}))) = \exp(\frac{1}{2}\sum_k (\mathrm{i}\lambda_k - a)^2(t_k - t_{k-1})).$$

We finally conclude that

$$\mathbb{E}_{\mathbb{Q}} \exp(\mathrm{i}\sum \lambda_k(W^{\mathbb{Q}}(t_k) - W^{\mathbb{Q}}(t_{k-1}))) = \exp(-\frac{1}{2}\sum_k \lambda_k^2(t_k - t_{k-1})). \quad (4.8)$$

From Equation (4.8) we deduce the independence under $\mathbb{Q}$ of the increments of $W^{\mathbb{Q}}$ and normality with the right parameters (see also Section A.3). □

Proposition 4.2 has the following

Corollary 4.3 *Let X be the process, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, given by $X(t) = at + \sigma W(t)$, where W is Brownian motion (w.r.t. $\mathbb{P}$) and $\sigma > 0$. Define the probability $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z(T) = \exp(-\gamma W(T) - \frac{1}{2}\gamma^2 T)$, with $\gamma = \frac{a-b}{\sigma}$. Then $X(t) = bt + \sigma W^{\mathbb{Q}}(t)$, where $W^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$ on $[0,T]$.*

Proof Exercise 4.9. □

Remark 4.4 It is possible to show that the probability measure $\mathbb{Q}$ of Corollary 4.3 is the unique probability measure on $\mathcal{F}_T$ such that X can be written as $X(t) = bt + \sigma W^{\mathbb{Q}}(t)$, where $W^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$. Hence, also the $Z(T)$ of this corollary is the unique random variable that gives a new probability measure $\mathbb{Q}$ such that X has this representation.

What we have seen here is that the *absolutely continuous measure transformations* change the drift parameter of the process X, but not the parameter σ. Can we also make a measure transformation, such that σ changes into another parameter, τ say, so that if $X(t) = at + \sigma W^{\mathbb{P}}(t)$ with $W^{\mathbb{P}}$ Brownian motion under $\mathbb{P}$ and so that $X(t) = bt + \tau W^{\mathbb{Q}}$ with $W^{\mathbb{Q}}$ Brownian motion under $\mathbb{Q}$?

Suppose that the two above representations of X hold with $\sigma > 0$ and $\tau > 0$. We have seen in Exercise 2.14 that for $X(t) = at + \sigma W^{\mathbb{P}}(t)$ the quadratic variation of X over an interval $[0,T]$ is almost surely equal to $\sigma^2 T$, so $\mathbb{P}(\langle X\rangle_T = \sigma^2 T) = 1$. Similarly, we have $\mathbb{Q}(\langle X\rangle_T = \tau^2 T) = 1$. Hence, assuming that $\sigma^2 \neq \tau^2$, there is set $E \in \mathcal{F}_T$ (namely $E = \{\langle X\rangle_T = \sigma^2 T\}$), such that $\mathbb{P}(E) = 1$ and $\mathbb{Q}(E) = 0$. If such a set can be found (as it is the case here), the measures $\mathbb{P}$ and $\mathbb{Q}$ are called *mutually singular*, the extreme opposite of being absolutely continuous w.r.t. to each other. One can say that $\mathbb{P}$ is concentrated on the event E (because $\mathbb{P}(E^c) = 0$), whereas for $\mathbb{Q}$ the opposite holds true.

Equation (4.2) tells us how to express expectations under a new probability measure $\mathbb{Q}$ in terms of expectation under $\mathbb{P}$. Do we have a similar device for conditional expectations? Here is the answer.

Proposition 4.5 *Let $\mathbb{Q}$ be defined by (4.5) on $\mathcal{F}_T$. Let X be a $\mathcal{F}_T$-measurable random variable such that $\mathbb{E}_{\mathbb{P}}|X|Z(T) < \infty$. Then*

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t] = \frac{\mathbb{E}_{\mathbb{P}}[XZ(T)|\mathcal{F}_t]}{Z(t)}, \ \mathbb{Q}\text{-a.s.} \quad (4.9)$$

Proof We will repeatedly use properties of the conditional expectation, see Section A.6 of the Appendix. Let $F \in \mathcal{F}_t$. We compute

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_F \mathbb{E}_{\mathbb{P}}[XZ(T)|\mathcal{F}_t]\right) &= \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_F XZ(T)\right)\\ &= \mathbb{E}_{\mathbb{Q}}\left(\mathbf{1}_F X\right)\\ &= \mathbb{E}_{\mathbb{Q}}\left(\mathbf{1}_F \mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]\right)\\ &= \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_F Z(t)\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]\right),\end{aligned}$$

where we used Proposition 4.1 in the last equality. Comparing the extreme sides of this chain of equations, which holds for any $F \in \mathcal{F}_t$, we conclude that $\mathbb{E}_{\mathbb{P}}[XZ(T)|\mathcal{F}_t] = Z(t)\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$. Since $\mathbb{Q}(Z(t) > 0) = 1$, we can divide by $Z(t)$ to arrive at (4.9). □

4.2 Exercises

4.1 Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that X is random variable having the $N(0,1)$ distribution. Let ϕ denote the density of the $N(0,1)$ distribution and let ϕ_a denote the density of the $N(a,1)$ distribution. Let $z(x) = \frac{\phi_a(x)}{\phi(x)}$. Define $\mathbb{Q}(F) = \mathbb{E}(\mathbf{1}_F Z)$ for $F \in \mathcal{F}$, where $Z = z(X)$.

(a) Compute $\mathbb{E}Z = 1$ (use an integral).

(b) Show that $\mathbb{Q}(X \leq x) = \Phi(x-a)$, where Φ is the distribution function of $N(0,1)$.

(c) Conclude that $X^{\mathbb{Q}} := X - a$ has the standard normal distribution under $\mathbb{Q}$.

(d) Show that $z(x) = \exp(ax - \frac{1}{2}a^2)$ and that Z has the same distribution as $\exp(aW(1) - \frac{1}{2}a^2)$, where W is a Brownian motion. Note the resemblance with (4.4). How to choose a in the current setup such that Z has exactly the same distribution as $Z(T)$?

We see that the change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is 'neutralized' by replacing X with $X - a$ in the sense that $X - a$ has, under $\mathbb{Q}$, the same distribution as X under $\mathbb{P}$.

4.2 Show that Brownian motion is a martingale w.r.t. its own filtration.

4.3 Take $\Omega = \{1,2,3\}$ and $\mathbb{P}$ the uniform probability measure on this set and let $\mathbb{Q}$ be such that $\mathbb{Q}(\{1\}) = \mathbb{Q}(\{2\}) = \frac{1}{2}$. Show that $\mathbb{Q} \ll \mathbb{P}$ and determine the Radon-Nikodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{P}$. Show that $\mathbb{P}(Z > 0) < 1$.

4.4 Let $\mathbb{Q}$ and $\mathbb{P}$ be two probabilities on a measurable space $(\Omega, \mathcal{F})$ and assume that $\mathbb{Q} \ll \mathbb{P}$. Show that $\mathbb{E}_{\mathbb{P}}Z = 1$ (Z as we used in Equation (4.2)).

4.5 Let $\Omega = \{\omega_1, \omega_2, \ldots\}$, $\mathcal{F}$ the power set of Ω and $\mathbb{P}$ and $\mathbb{Q}$ be two probabilities on $(\Omega, \mathcal{F})$. Let $\tilde{\Omega} = \{\omega \in \Omega : \mathbb{P}(\{\omega\}) > 0\}$. Assume that $\mathbb{Q}(F) = 0$ as soon as $\mathbb{P}(F) = 0$. Then in particular $\mathbb{Q}(\{\omega\}) = 0$ if $\omega \in \tilde{\Omega}^c$. Define $Z(\omega) = \mathbb{Q}(\{\omega\})/\mathbb{P}(\{\omega\})$ for $\omega \in \tilde{\Omega}$ and zero otherwise. Show by direct computation that $\mathbb{E}_{\mathbb{Q}}X = \mathbb{E}_{\mathbb{P}}XZ$ for all nonnegative random variables X.

4.6 Let Z be $N(0, \tau^2)$ distributed. Show that for all $\alpha \in \mathbb{R}$ the equality $\mathbb{E}(Z\exp(\alpha Z)) = \alpha\tau^2\exp(\frac{1}{2}\alpha^2\tau^2)$ holds.

4.7 Let W be a standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let for every $t \in \mathbb{R}$ $\mathcal{F}_t$ be the σ-algebra generated by $W(s)$ with $s \leq t$. Put $Z = \exp(-aW(T) - \frac{1}{2}a^2T)$ for a certain $T > 0$. Let $\mathbb{Q}$ be the probability on $\mathcal{F}_T$ defined by $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} = \mathrm{Z}$. Show that $\mathbb{E}_{\mathbb{Q}}W(t) = -at$ for $t \leq T$ and that $\mathbb{E}_{\mathbb{Q}}[W(T)|\mathcal{F}_t] = W(t) - a(T-t)$ by using Proposition 4.5.

4.8 Prove Proposition 4.1 by writing $Z(T) = Z(t)\exp(-a(W_T - W_t) - \frac{1}{2}a^2(T - t))$ and exploiting the independence (under $\mathbb{P}$) of the increments of W.

4.9 Prove Corollary 4.3 (it is easy).

4.10 Let W be a Brownian motion under $\mathbb{P}$ and define $\mathbb{Q}$ by (4.5) with $Z(T)$ as in (4.4). Show, use (4.9), that

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t] = \mathbb{E}_{\mathbb{P}}[Xe^{-a(W(T)-W(t))}|\mathcal{F}_t]e^{-\frac{1}{2}a^2(T-t)}.$$

4.11 Let $\varepsilon_1, \ldots, \varepsilon_n$ be random variables that under a probability measure $\mathbb{P}$ are independent and standard normally distributed. Let $\mu_1, \ldots, \mu_n$ and $\sigma_1, \ldots, \sigma_n$ be real numbers and $Y_i = \mu_i + \sigma_i\varepsilon_i$, $i = 1, \ldots, n$. Assume that under a probability measure $\mathbb{Q}$ (on $\sigma(Y_1, \ldots, Y_n)$) the Y_i become independent with each a $N(0, \sigma_i^2)$ distribution. Show that the likelihood ratio (Radon-Nikodym derivative) $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(-\sum_{i=1}^n \frac{\mu_i}{\sigma_i}\varepsilon_i - \frac{1}{2}\sum_{i=1}^n (\frac{\mu_i}{\sigma_i})^2\right)$.

4.12 Let $0 = t_0 < t_1 < \cdots < t_n = T$ and $a(t) = \sum_{i=1}^n a_i \mathbf{1}_{(t_{i-1}, t_i]}(t)$, where the a_i are real numbers. Let W be a standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and X be a Brownian motion with piecewise constant drift defined by $X(t) = \int_0^t a(s)\,\mathrm{d}s + \sigma W(t)$.
Let $\Delta W_i = W(t_i) - W(t_{i-1})$ and $Z = \exp\left(-\frac{1}{\sigma}I(a) - \frac{1}{2}\frac{1}{\sigma^2}\int_0^T a^2(s)\,\mathrm{d}s\right)$, where $I(a) = \sum_{i=1}^n a_i\Delta W_i$. Define $\mathbb{Q}$ by $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z$.
Argue that under $\mathbb{Q}$ the process X becomes a Brownian motion on $[0, T]$ with variance parameter σ^2. *Hint: Copy the proof of Proposition 4.2 and use that*

$$Z(T) = \prod_{i=1}^n \exp\left(-\sum_{i=1}^n \frac{a_i}{\sigma}\Delta W_i - \sum_{i=1}^n \frac{a_i^2}{2\sigma^2}(t_i - t_{i-1})\right).$$

You may even assume w.l.o.g. (think why!) that for the increments over intervals $(t_{k-1}, t_k]$, the t_k are the t_i in the definition of $a(t)$.
Remark: A more sophisticated argument, involving conditional characteristic functions, shows that X is even a Brownian motion, if the a_i are bounded $\mathcal{F}_{t_{i-1}}$-measurable random variables.

4.13 Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable Z with the properties $\mathbb{P}(Z \geq 0) = 1$ and $\mathbb{E}Z = 1$. Define $\mathbb{Q}$ on $\mathcal{F}$ by $\mathbb{Q}(F) = \mathbb{E}[\mathbf{1}_F Z]$.

(a) Show that $\mathbb{Q}$ is a probability measure on $\mathcal{F}$. [For countable additivity you need the Monotone convergence theorem.]
(b) Show that $\mathbb{Q}(F) = 0$ if $\mathbb{P}(F) = 0$.
(c) Show that $\mathbb{P} \ll \mathbb{Q}$ if and only if $\mathbb{P}(Z > 0) = 1$.

5 Black Scholes market

We have established in Proposition 1.9 the (finite dimensional) limit distributions of the Cox-Ross-Rubinstein model. Let us recall what we found. The limit stochastic variables $S(t_k)$ were such that the increments of the log-price process $\log S(t)$ became independent and moreover the distribution of $\log S(t_k) - \log S(t_{k-1})$ was normal with mean $(r-\frac{1}{2}\sigma^2)(t_k-t_{k-1})$ and variance $\sigma^2(t_k-t_{k-1})$. Having in mind the distribution of a process X, a Brownian motion with drift, that is given by $X(t) = x_0 + at + \sigma W(t)$ (see the end of Section 2.2), we postulate the following model for the log-price process under the *equivalent martingale measure* $\mathbb{Q}$.

$$\log S(t) = \log s + (r - \frac{1}{2}\sigma^2)t + \sigma W^{\mathbb{Q}}(t), \tag{5.1}$$

where $W^{\mathbb{Q}}$ is a Brownian motion under the probability measure $\mathbb{Q}$. Note that for this model increments $\log S(t_k) - \log S(t_{k-1})$ are independent and that they have the normal distributions as specified above.

5.1 Models with equivalent measures

Let us first explain (in this context) the term equivalent martingale measure as we used it above. With the exposition of Section 1 in mind, it should be the case that the discounted price process $\bar{S}$ is a martingale. Since the discount factor at each time t is the bond price $B(t) = e^{rt}$, we get from Equation (5.1) that

$$\bar{S}(t) = s \exp\big(\sigma W^{\mathbb{Q}}(t) - \frac{1}{2}\sigma^2 t\big). \tag{5.2}$$

We conclude from the discussion of Girsanov's theorem, compare to Equation 4.6, that this indeed gives a martingale and therefore talking about an equivalent *martingale* measure thus makes sense. What about *equivalent*?

Recall from Section 1 that in the CRR set up, the probability measure $\mathbb{Q}$ was an artefact, useful to compute prices of financial derivatives, but the 'physical' measure, $\mathbb{P}$ say, that attributes the 'real' probabilities to up and down movements of the stock price may be very different. Nevertheless we assumed equivalence of these probability measures in the sense that price paths that had positive probability under one measure had positive probability under the other as well. In the present context we therefore look for any other probability measure $\mathbb{P}$ that is equivalent to $\mathbb{Q}$ and we wonder how such a measure equation will change (5.1).

The solution of this problem can be obtained by application of Girsanov's theorem, as we discussed it in Section 4. So we conclude from e.g. Corollary 4.3 that any other 'physical' probability measure $\mathbb{P}$ that gives $\log S$ a constant drift parameter, different from $r - \frac{1}{2}\sigma^2$, is equivalent to $\mathbb{Q}$, when restricted to $\mathcal{F}_T$, with the T the terminal time. Here we take $\mathcal{F}_T$ to be the smallest σ-algebra that makes the $W^{\mathbb{Q}}(s)$, $s \leq T$, measurable functions on Ω. For $t \leq T$ we have a similar definition of $\mathcal{F}_t$. Note that $\mathcal{F}_T$ is also the smallest σ-algebra that makes all $S(t)$, $t \leq T$, measurable functions (why?).

Let us note that any other probability measure that gives $\log S$ a constant drift can arise as the limit of discrete time processes in the way we treated this subject in Section 1. Indeed, replacing the $q_u(N)$ that we used there with a probability $p_u(N) = \frac{e^{(\alpha+\frac{1}{2}\sigma^2)\Delta_N} - d_N}{u_N - d_N}$ eventually leads to a limit process S that can be represented as $\log S(t) = \log s + \alpha t + \sigma W^{\mathbb{P}}(t)$.

There are however many more probabilities that are equivalent to $\mathbb{Q}$. It follows from Exercise 4.12 that also a probability measure $\mathbb{P}$ under which $\log S$ has a piecewise constant drift is equivalent to $\mathbb{Q}$. It is a *theorem* that the piecewise constant functions are dense in $\mathcal{L}^2[0,T]$ and this enables one to prove that any probability measure $\mathbb{P}$ that is such that

$$\log S(t) = \log S(0) + \int_0^t a(s)\,\mathrm{d}s + \sigma W^{\mathbb{P}}(t), \tag{5.3}$$

with $a \in \mathcal{L}^2[0,T]$ and $W^{\mathbb{P}}$, a Brownian motion under $\mathbb{P}$, is equivalent to $\mathbb{Q}$ when restricted to $\mathcal{F}_T$. We don't discuss the details of this result at the present stage. One of the reasons is that with the tools that we have developed so far, it is unclear how the Radon-Nikodym derivatives $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$ and $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ would look like. We come back to this later on in Section 6.3. For now it is sufficient to work with probabilities $\mathbb{P}$ under which we have Equation (5.3) and to know that these probabilities are equivalent to $\mathbb{Q}$ when we work with a fixed time horizon T. Together with the bond price process B given by $B(t) = \exp(rt)$ we are in the framework of the *Black-Scholes* model. The two equations for S and B are then said to describe the Black-Scholes market.

As a final remark to this section we mention that there are still many other probability measures $\mathbb{P}$, equivalent to $\mathbb{Q}$. Under these other measures Equation (5.3) will change in the sense that also certain *random* functions a are allowed.

5.2 Arbitrage

As in Section 1 we will consider portfolios that consist of a certain real number of stocks and a certain real number of bonds. The portfolios will be dynamic and therefore the numbers of stocks and bonds depend on time. For each fixed time t they are allowed to depend on the behavior and values of the stock price prior to t, but not on the future stock prices after t. Denoting (at time t) these numbers by x_t and y_t we thus have that they are in fact random variables and the portfolio process (x, y) is thus a bivariate stochastic process. The value of the portfolio at time t is denoted by $V(t)$ and obviously

$$V(t) = x_t S(t) + y_t B(t). \tag{5.4}$$

We will also work with the discounted values of stochastic processes. Like in Section 1, we write $\bar{Y}$ for the discounted version of a process Y and it is defined by $\bar{Y}(t) = Y(t)/B(t) = Y(t)e^{-rt}$. In particular we use the discounted value $\bar{V}(t)$ of the portfolio, which is of course

$$\bar{V}(t) = x_t \bar{S}(t) + y_t. \tag{5.5}$$

Let us for a while agree on the following definition (in the spirit of Exercise 1.7) of what we will call a *martingale portfolio*. It is such that the discounted value process $\bar{V}$ is a martingale under the measure $\mathbb{Q}$. In the discrete time setting a martingale portfolio was a self-financing portfolio and vice versa (Exercise 1.7).

This is (under appropriate conditions) harder to establish in the continuous time case. But it is relatively easy to give an expression for the value process V of such a portfolio. By definition we have $\bar{V}(t) = \mathbb{E}_{\mathbb{Q}}[\bar{V}(T)|\mathcal{F}_t]$ and thus $V(t) = e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[V(T)|\mathcal{F}_t]$. Conditional expectations of this type have been analyzed in Section 3, where we used partial differential equations with a boundary condition at time T, which could be cast as $V(T) = f(W(T))$. So we can apply the results of that section as soon as we can formulate a boundary condition as a function of $W(T)$. This will be done in Section 5.3.

We aim at a definition of self-financing portfolios and we let us inspire by results of Section 1, more precisely by the considerations that led to equations (1.11) and (1.12). Whereas in Section 1 we emphasized the value process, here we turn our attention to the structure of the self-financing portfolio process. *Important:* we set $r = 0$ in the motivation below of the definition of a self-financing portfolio.

We make a digression and consider first an arbitrary portfolio in discrete time with value process V. Let us introduce the notation $V_n(u)$ as the value of V_n if $S_n = S_{n-1}u$, $V_n(d)$ likewise and, similarly, we write $S_n(u) = S_{n-1}u$ and $S_n(d) = S_{n-1}d$. Since x_n and y_n don't depend on the value of S_n we have the two identities

$$V_n(u) = x_nS_n(u) + y_n$$
$$V_n(d) = x_nS_n(d) + y_n.$$

We introduce the difference operator D, that applied to V_n by definition results into $DV_n = V_n(u) - V_n(d)$ and applied to S_n results in $DS_n = S_n(u) - S_n(d)$. Now we can write

$$x_n = \frac{DV_n}{DS_n}. \tag{5.6}$$

Recall that for $r = 0$ a portfolio in discrete time is self-financing iff its value process V is a martingale under the equivalent martingale measure $\mathbb{Q}$ (Exercise 1.7). Suppose that we have a self-financing portfolio such that its terminal value V_N is a function of S_N, $V_N = f(S_N)$, say. Then we have $V_n = \mathbb{E}_{\mathbb{Q}}[f(S_N)|S_0, \dots, S_n]$, which reduces, by the Markov property of S under the probability $\mathbb{Q}$, to a function of S_n only, $V_n = v_n(S_n)$ say. We can then write $V_n(u) = v_n(S_{n-1}u)$ and $V_n(d) = v_n(S_{n-1}d)$. In this case Equation (5.6) becomes

$$x_n = \frac{v_n(S_{n-1}u) - v_n(S_{n-1}d)}{S_{n-1}u - S_{n-1}d}. \tag{5.7}$$

We consider what happens if $u \downarrow 1$, $d \uparrow 1$ so that $u - d \to 0$. Assume that we can extend the domain of the function v_n to $\mathbb{R}$ and that the derivative v_n' of v_n exists. Then a Taylor expansion in (5.7) yields the expression

$$x_n \approx v_n'(S_{n-1}).$$

This result will be the basis of our definition of a self-financing portfolio in continuous time, in particular for a portfolio that hedges a simple claim.

We now turn to the limiting procedure of Section 1.2. So we take a large number N and compare the market in continuous time with time horizon $[0, T]$

with the fictitious, or approximating, CRR market with time set $\{0, \ldots, N\}$, where N and T are related by $N\Delta_N = T$. Fix a time $t \in [0, T]$ and let n be the corresponding time instant in the fictitious market, so $n = [\frac{t}{T}N]$ and we have $t_n^N \leq t < t_{n+1}^N$, with $t_n^N = \frac{n}{N}T$.

By $\mathbb{Q}_N$ we denote the equivalent martingale measure in the N-th approximating CRR market and, in general, we endow any quantity related to these market with an upper or lower index N. Let us now focus on the values $V_n^N = v_n^N(S_n^N)$. Recall, use Equation (1.19) with $r = 0$, that we can write these as $\mathbb{E}_{\mathbb{Q}_N} f(\frac{S_N^N}{S_n^N} s)$ if S_n^N has the value s. Assume for a moment that f is a bounded continuous function. Then we have from the convergence results of Section 1.2, that $v_n^N(s)$ converges to $v(t, s) := \mathbb{E}_{\mathbb{Q}} f(\frac{S(T)}{S(t)} s)$, where $\mathbb{Q}$ is now the probability such that Equation (5.1) holds. Next we replace s with su_N or with sd_N. If v is differentiable in the second variable with derivative v_x we find, parallel to what we did just after (5.7), the approximation $v_n^N(su_N) \approx v(t, s) + s(u_N - 1)v_x(t, s)$ and a similar approximation for $v_n^N(sd_N)$. If x_n^N is the amount of stocks at time n in the N-th CRR market, we thus obtain from (5.7) the approximation

$$x_n^N = \frac{DV_n^N(s)}{DS_n^N} \approx v_x(t, s).$$

The arguments that we used to obtain the last approximation can be made rigorous under suitable assumptions on f, for instance if f is such that a condition equivalent to Condition 3.1 holds for all $a > 0$, but this is not our aim. However, it suggests the following definition of a self-financing portfolio under special circumstances. The temporary assumption $r = 0$ will shortly be seen to be superfluous. The general case will be treated in Section 7.

Definition 5.1 A martingale portfolio in the Black-Scholes market that is such that the value process V can be written as $V(t) = v(t, S(t))$, with v having appropriate differentiability properties, is called *self-financing* if the amount x_t invested in the stock at any time t satisfies the relation $x_t = v_x(t, S(t))$.

This definition was suggested by a limit procedure under the condition that the interest was zero. But it is possible to reduce the case of non-zero interest rate to this case and we will show how to do this. So, we consider a portfolio with value process V that is given as $V(t) = v(t, S(t))$. The discounted value process is then $\bar{V}(t) = \bar{v}(t, \bar{S}(t))$ where the functions v and $\bar{v}$ are related by $\bar{v}(t, x) = e^{-rt} v(t, e^{rt} x)$. Observe that in the discounted Black-Scholes market the interest rate is equal to zero. Suppose that we have a portfolio that is self-financing in the sense of Definition 5.1 in the discounted market. Then we know that $x_t = \bar{v}_x(t, \bar{S}(t))$. It is now easy to deduce that also $x_t = v_x(t, S(t))$ (Exercise 5.1). What this result says can also be stated as 'a portfolio is self-financing in the Black-Scholes market iff it is self-financing in the discounted Black-Scholes market'. Moreover Definition 5.1 now also makes sense in a market with nonzero interest rate. One might wonder whether the partial derivative w.r.t. the second variable of the function v in Definition 5.1 exists. This is guaranteed by the results of Section 3. For more details we refer to Section 5.3.

Knowing what a self-financing portfolio is, we can now define arbitrage portfolios. A portfolio is an arbitrage portfolio over the interval $[0, T]$ if it is self-financing and such that $V(0) = 0$ with $\mathbb{P}$-probability equal to 1, $V(T) \geq 0$ with

$\mathbb{P}$-probability equal to 1 and $V(T)$ is strictly positive with positive $\mathbb{P}$-probability. Note that this (verbal) definition is essentially the same as Definition 1.2 for a discrete time market.

Proposition 5.2 *The Black-Scholes market is free of arbitrage.*

Proof Clearly, the definition of an arbitrage portfolio could alternatively be stated in terms of the discounted value process instead of the value process. Since $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, we also have $\mathbb{Q}(V(0) = 0) = 1$, $\mathbb{Q}(\bar{V}(T) \geq 0) = 1$ and $\mathbb{Q}(\bar{V}(T) > 0) > 0$. Suppose that we have a self-financing portfolio. Since its discounted value $\bar{V}$ is a martingale under $\mathbb{Q}$, we have in particular $V(0) = \bar{V}(0) = \mathbb{E}_{\mathbb{Q}}\bar{V}(T)$. If $\mathbb{Q}(\bar{V}(T) \geq 0) = 1$ and $\mathbb{Q}(\bar{V}(T) > 0) > 0$ we necessarily have $\mathbb{E}_{\mathbb{Q}}\bar{V}(T) > 0$ so that $V(0) > 0$. This portfolio is thus not an arbitrage possibility. □

Having established that the Black-Scholes market is free of arbitrage, we can adopt the following pricing principle. If we have two financial products that at a time T give exactly the same payoff, then their fair prices at any time t before T are the same. We have encountered this already as *the law of one price*. We will use this principle in Section 5.3, when we discuss products that can be hedged. We thus defined what is called *no arbitrage pricing*.

5.3 Hedging

Consider a contingent claim (simple or not) whose payoff at the maturity time T is equal to X. By definition, X is an $\mathcal{F}_T$-measurable random variable. Assume that the claim has finite expectation w.r.t. $\mathbb{Q}$. For technical reasons, unless stated otherwise, we will mostly restrict our attention to *simple* claims X, which are by definition claims of the form $X = F(S(T))$, where F is a measurable function on $\mathbb{R}$. Since below we work with expectations of claims, we also need to impose sufficient integrability conditions. We will sometimes state these explicitly, otherwise these are tacitly understood to be satisfied.

We say that a claim X can be *hedged* if there is a self-financing portfolio process (x_t, y_t), $t \in [0, T]$, such that the terminal value of the portfolio is under $\mathbb{Q}$ almost surely equal to X, so $V(T) = X$, $\mathbb{Q}$-a.s. This portfolio is called the *hedging* or *replicating* portfolio. By the no arbitrage pricing principle (and the law of one price), the fair price of a hedgeable claim at any time $t \leq T$ is equal to the value $V(t)$ at that time of the hedging portfolio. But by the definition of a self-financing portfolio as a special kind of martingale portfolio, we know that $\bar{V}(t) = \mathbb{E}_{\mathbb{Q}}[\bar{V}(T)|\mathcal{F}_t]$. Hence, with bond prices $B(t) = \exp(rt)$, we get that $V(t) = e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[V(T)|\mathcal{F}_t]$, so that the fair price at time t of the hedgeable claim X becomes $e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$.

Assume that hedging is possible. The above formula for the fair price is so far nothing else but a representation. What we aim at is, if possible, an explicit expression for the fair price. So we need to calculate the conditional expectation $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$ explicitly. As a first step we show that the value process V of a simple claim can be written as $V(t) = v(t, S(t))$ for some function v that will be specified later on. What is needed to show this, is that the process S is Markov under the probability $\mathbb{Q}$. This is straightforward. Since $S(t) = s\exp((r - \frac{1}{2}\sigma^2)t + \sigma W^{\mathbb{Q}}(t))$ and $W^{\mathbb{Q}}$ is Markov, this readily follows (you check!). Because $\bar{V}(t) = \mathbb{E}_{\mathbb{Q}}[\bar{F}(S(T))|\mathcal{F}_t]$, we see that this conditional expectation reduces to the conditional expectation $\mathbb{E}_{\mathbb{Q}}[\bar{F}(S(T))|S(t)]$. Defining $\bar{v}(t,x) = \mathbb{E}_{\mathbb{Q}}[\bar{F}(S(T))|\bar{S}(t) = x]$, we obtain that $\bar{v}(t,\bar{S}(t)) = \bar{V}(t)$ and $V(t) = e^{rt}\bar{v}(t,\bar{S}(t))$, so that $v(t,x) = \bar{v}(t,xe^{-rt})e^{rt}$.

The next step is to specify the function v or, equivalently, the function $\bar{v}$. To that end we introduce the auxiliary function f, defined by

$$f(x) = e^{-rT}F(se^{\sigma x+(r-\frac{1}{2}\sigma^2)T}). \tag{5.8}$$

We thus have that $\bar{F}(S(T)) = f(W^{\mathbb{Q}}(T))$. With another auxiliary function w, defined by $w(t,y) = \mathbb{E}_{\mathbb{Q}}[f(W^{\mathbb{Q}}(T))|W^{\mathbb{Q}}(t) = y]$, we have that $w(t,y) = \bar{v}(t,se^{\sigma y-\frac{1}{2}\sigma^2 t})$. The function w is known to us from Section 3, it satisfies the backward heat equation (3.6) under the appropriate conditions on f. But then we know parallel to Exercise 3.5 also which partial differential equation $\bar{v}$ satisfies and from Exercise 3.6 the partial differential equation that v itself satisfies. We summarize this paragraph as follows.

Proposition 5.3 *Any simple claim $F(S(T))$ with finite expectation that can be hedged is such that the value $V(t)$ of the claim at time t can be written as $V(t) = v(t,S(t))$, where v is the solution of the Black-Scholes partial differential equation*

$$v_t(t,x) + \frac{1}{2}\sigma^2x^2v_{xx}(t,x) + rxv_x(t,x) - rv(t,x) = 0, \tag{5.9}$$

with boundary condition $v(T,x) = F(x)$.

The natural question to ask is now 'which simple claims can be hedged?' The answer is that we can hedge any simple claim $F(S(T))$ under certain regularity conditions on the function F.

Theorem 5.4 *The Black-Scholes market is complete in the sense that every simple claim $F(S(T))$ can be hedged, if f defined by Equation (5.8) satisfies Condition 3.1.*

Before proving this theorem we make some comments. The theorem clearly differs from the analogous statement for the CRR market where every claim, simple or not, could be hedged. We confined ourselves to the restricted case of simple claims for technical reasons. It is also true that *every* claim (subject to certain integrability conditions) in the Black-Scholes market can be hedged. This result however requires a rather deep theorem in probability theory (the Martingale Representation Theorem), to which we return in Section 7.3.

Proof of Theorem 5.4 Let $F(S(T))$ be the claim under consideration. We have to show that we can find a self-financing portfolio that is such that its terminal value $V(T)$ is equal to $F(S(T))$. From the discussion preceding Proposition 5.3 we know that $v(t,x) := e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[F(S(T))|S(t) = x]$ is the solution to Equation (5.9). In particular, v_x exists and we can define $x_t = v_x(t,S(t))$. With $y_t = e^{-rt}(v(t,S(t)) - x_tS(t))$ we obtain a self-financing portfolio with value process $V(t) = v(t,S(t))$ and the property that $v(T,S(T)) = F(S(T))$. This portfolio is thus the hedge portfolio of our claim. □

Many of the results, like the investment $x_t = v_x(t, S(t))$, of Definition 5.1 in this section are based on rather heuristic arguments. A more fundamental approach will be encountered in Section 7. To use that approach we need to develop the theory of Stochastic integration. This will be the topic of Section 6, with applications to more general measure transformations of Girsanov type in Section 6.3.

5.4 Exercises

5.1 Show that a portfolio is self-financing in the Black-Scholes market iff it is self-financing in the discounted Black-Scholes market.

5.2 Consider the Black-Scholes market with $r > 0$. Suppose that a portfolio is such that the value process V is of the form $V(t) = v(t, S(t))$. Take as an alternative portfolio the one in which $x_t \equiv 0$ and $y_t \equiv \bar{v}(t, S(t))$. Show that this portfolio process can only be self-financing if $V(t)$ does not depend on $S(t)$.

5.3 Consider the Black-Scholes model and a self-financing portfolio that replicates a non-negative claim. Show that the value process of the hedge portfolio is non-negative as well.

5.4 Consider the Black-Scholes model and determine the hedge portfolios as well as the value functions for each of the following simple claims.

(a) $F(S(T)) = S(T) - K$.

(b) $F(S(T)) = (S(T) - K)^2$. *Hint:* Reduce the value function to an expectation of a random variable that is a function of $S(T)/S(t)$. Compute the expectation by using Exercise A.5.

(c) $F(S(T)) = (\log S(T) - K)^+$.

(d) $F(S(T)) = 100\,\mathbf{1}_{\{S(T)>100\}}$.

(e) $F(S(T)) = (S(T) - K)^+$. Consider first the case $r = 0$. For $t < T$ the price C_t of this claim is then given by

$$C_t = S_t \Phi(\frac{\log \frac{S_t}{K} + \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}) - K\Phi(\frac{\log \frac{S_t}{K} - \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}),$$

where Φ is the distribution function of the standard normal distribution. Find also C_T, and x_T and y_T.

(f) Use the results for $r = 0$ under (e) to obtain the expression for C_t in the general case $r \geq 0$. What are x_t and y_t now?

6 Elementary Itô calculus

The modern theory of option pricing uses the technology and calculus of the *Stochastic integral*, also called *Itô integral*. In the presentation thus far we could circumvent this theory, although in Section 5 we defined self-financing portfolios in a special case only and we could not treat hedging of composite claims. It is impossible to do this without knowledge of Itô integrals. Let us first go back to the definition of a self-financing portfolio in discrete time. It was such that, in the notation that has by now become familiar to us, $\Delta\bar{V}_n = x_n \Delta\bar{S}_n$. With the limit procedure of Section 1.2 in mind we could replace this by the continuous time analogue, 'differences become differentials', $\mathrm{d}\bar{V}(t) = x_t \mathrm{d}\bar{S}(t)$. Now we are immediately faced with the question how to interpret this equation, and what is meant by a 'differential' like $\mathrm{d}\bar{V}(t)$? If $\bar{V}$ and $\bar{S}$ would be a differentiable function of t, we could rely on ordinary differential calculus to give a meaning to this equation, $\frac{\mathrm{d}}{\mathrm{d}t}\bar{V}(t) = x_t \frac{\mathrm{d}}{\mathrm{d}t}\bar{S}(t)$. In the present context it turns out to be the case that functions $\bar{V}$ and $\bar{S}$ are not differentiable w.r.t. t. Especially for the latter, this is relatively easy to see. If it were differentiable, the same would be the case with $\log \bar{S}$. But we have modeled this as a Brownian motion with (linear) drift under $\mathbb{Q}$, see Equation (5.1). Brownian motion itself is not differentiable, as we already observed in Section 2.3. Nevertheless, we will see that it is possible to derive a *stochastic* calculus, that is built on *stochastic* integrals.

In all what follows we assume that we work with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a Brownian motion W is defined. By $\mathcal{F}_t$ we denote the σ-algebra that is generated by $W(s), s \leq t$.

6.1 The Itô integral, an informal introduction

The purpose of this section is to define an integral of a stochastic process a w.r.t. a Brownian motion W. The integral over a time interval $[0, T]$ will be denoted by

$$\int_0^T a(s)\,\mathrm{d}W(s) \tag{6.1}$$

and will be called the *stochastic integral* or the *Itô integral* of a w.r.t. W. The form of the expression (6.1) reminds us of a Stieltjes, or perhaps of a Lebesgue integral. In both cases integrals were defined as limits of sums. And this is, in spite of some subtle differences, the approach that we also follow here, en passant touching upon the differences between the conventional set up and the present one.

First we define *simple* processes. We call a process a simple over a fixed interval $[0, T]$ if it is bounded and can be written as

$$a(t) = \sum_{j=1}^{n} a_j \mathbf{1}_{(t_{j-1}, t_j]}(t), \tag{6.2}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ and where the a_j are $\mathcal{F}_{t_{j-1}}$-measurable bounded random variables. The class of simple processes will be denoted by $\mathcal{S}$. $\mathcal{S}$ will be endowed with a norm $||\cdot||$, defined by $||a|| = (\mathbb{E}\int_0^T a(s)^2\,\mathrm{d}s)^{1/2}$.

Strictly speaking, this not a norm, since $||a|| = 0$ does not imply that $a(s) = 0$ for all s. But by identifying functions whose difference has norm zero (so we actually look at a quotient space), it becomes one.

For a simple process a there is basically only one proper choice of what the integral $\int_0^T a(s)\,\mathrm{d}W(s)$ should be. We define it as the sum

$$\int_0^T a(s)\,\mathrm{d}W(s) = \sum_{j=1}^{n} a_j(W(t_j) - W(t_{j-1})). \tag{6.3}$$

It is not very difficult to show, Exercise 6.1, that this integral is independent of the chosen representation of a, like in (6.2).

We haven't used yet any specific property of W as a Brownian motion. Therefore we take the liberty to temporarily move away from the Brownian context and give a financial interpretation. Suppose then that W would be the price process of some asset. Suppose that the t_j are trading times at which one can purchase a number of assets, the number being a_j at time t_{j-1} and that this number is held constant on the time interval $(t_{j-1}, t_j]$ and that at t_j the number of assets is immediately replaced with a_{j+1}. A negative a_j is to be interpreted as selling assets. When we move in time from t_{j-1} to t_j the price changes from $W(t_{j-1})$ to $W(t_j)$, and the profit (or loss) of the holder of the assets in that time period becomes $a_j(W(t_j) - W(t_{j-1}))$. As a result $\int_0^T a(s)\,\mathrm{d}W(s)$ as in (6.3) is the total profit (or loss) somebody would incur if a_j assets are held during every time interval $(t_{j-1}, t_j]$.

We also want to define an integral with T replaced with $t < T$. For that case we define

$$\int_0^t a(s)\,\mathrm{d}W(s) := \int_0^T \mathbf{1}_{(0,t]}(s)a(s)\,\mathrm{d}W(s).$$

Note that also $a\mathbf{1}_{(0,t]}$ is a simple process, if a is such. An alternative formula (Exercise 6.2) is

$$\int_0^t a(s)\,\mathrm{d}W(s) = \sum_{j=1}^{n} a_j(W(t_j \wedge t) - W(t_{j-1} \wedge t)). \tag{6.4}$$

Let us write $I(a)$ for the process defined on $[0, T]$ by $I_t(a) = \int_0^t a(s)\,\mathrm{d}W(s)$. We immediately mention some properties of the thus defined integral process.

Proposition 6.1 *For $a \in \mathcal{S}$ we have*

(i) *The process $I(a)$ is a continuous martingale with expectation zero.*

(ii) $\mathbb{E}[(I_t(a) - I_s(a))^2|\mathcal{F}_s] = \mathbb{E}[\int_s^t a(s)^2\,\mathrm{d}s|\mathcal{F}_s]$.

(iii) $\mathbb{E}(I_t(a) - I_s(a))^2 = \mathbb{E}\int_s^t a(u)^2\,\mathrm{d}u$ *for $t > s$.*

(iv) *I_T considered as an operator on the space of simple processes is linear and an isomorphism from $\mathcal{S}$ with $||\cdot||$ into $\mathcal{L}^2(\Omega, \mathcal{F}_T, \mathbb{P})$.*

Proof (i) Without loss of generality we may assume that s and t are among the t_i. The martingale property now follows from Exercise A.17. Continuity is obvious.

(iii) follows from (ii), which *you* do as Exercise 6.4 by assuming again that the s and t are among the t_i.

(iv) That I_T is linear is obvious. Taking $s = 0$ and $t = T$ in (ii) gives us $\mathbb{E}I_T(a)^2 = \mathbb{E}\int_0^T a(s)^2\,\mathrm{d}s$. □

Remark 6.2 Since we know from property (i) of Proposition 6.1 that $I(a)$ is a martingale, we can formulate property (iii) as a consequence of Exercise A.27 also as $\mathbb{E}[I_t(a)^2 - I_s(a)^2|\mathcal{F}_s] = \mathbb{E}[\int_s^t a(s)^2\,\mathrm{d}s|\mathcal{F}_s]$. By rearranging one gets $\mathbb{E}[I_t(a)^2 - \int_0^t a(s)^2\,\mathrm{d}s|\mathcal{F}_s] = I_s(a)^2 - \int_0^s a(s)^2\,\mathrm{d}s$. This shows that the process $\{I_t(a)^2 - \int_0^t a(s)^2\,\mathrm{d}s, t \in [0,T]\}$ is a martingale.

Example 6.3 Let the times t_i be fixed and consider the process $W^n(t) = \sum_{i=1}^n W(t_{i-1})\mathbf{1}_{(t_{i-1},t_i]}(t)$. Application of the definition of the integral results after some manipulation, the basic identity $a(b-a) = \frac{1}{2}\big(b^2 - a^2 - (b-a)^2\big)$ is useful here, in

$$\int_0^T W^n(s)\,\mathrm{d}W(s) = \frac{1}{2}W(T)^2 - \frac{1}{2}\sum_{i=1}^n (W(t_i) - W(t_{i-1}))^2. \tag{6.5}$$

Note that, although piecewise constant, the W^n are not simple, since the $W(t_{i-1})$ are not bounded. Nevertheless, the resulting expression is just as correct as (6.3) for simple processes. See also Exercise 6.3.

Here we pause for a while in our treatment to outline why we cannot use Stieltjes of or Lebesgue integration theory. The main theorem for Stieltjes integrals is that the integral $\int_0^T a(s)\,\mathrm{d}W(s)$ exists for *functions* a and W if a is continuous and W of bounded variation. By a partial integration trick, one can show that the integral also exists if W is continuous and a of bounded variation. One then writes

$$\int_0^T a(s)\,\mathrm{d}W(s) = a(T)W(T) - \int_0^T W(s)\,\mathrm{d}a(s).$$

This trick is clearly of no help if we make the special choice W to be a typical path of Brownian motion and $a = W$. We have seen in Section 2.3 that the paths of Brownian motion are not of bounded variation, therefore Stieltjes integration theory is useless at this point. The same holds for Lebesgue theory. To attach to a function W a signed measure one again needs this function to be of bounded variation. So we find ourselves in a dead end street, it seems. Nevertheless we will be able to find a way out by exploiting the special structure of simple processes, the a_i were required to be $\mathcal{F}_{t_{i-1}}$-measurable, combined with the fact that the paths of Brownian motion are of bounded quadratic variation and the isometry property mentioned in Proposition 6.1.

To define $\int_0^T a(s)\,\mathrm{d}W(s)$ for a wider class of processes we will use a limit procedure. We don't treat the most general class of processes for which it is possible to define the stochastic integral. Since the integrands that we encounter are mainly such that they have continuous paths or piecewise continuous paths, we restrict ourselves, next to simple processes, to piecewise continuous integrands. Recall that a function $a : [0,T] \to \mathbb{R}$ is piecewise continuous if there are finitely many intervals $[s_{i-1}, s_i]$ with union $[0,T]$ such that a is continuous on the (s_{i-1}, s_i) and has finite left and right limits at the s_i. Note that a piecewise continuous function is bounded on $[0, T]$.

A process is said to be piecewise continuous if it has paths that are piecewise continuous functions, and we take the s_i the same for each path. These paths are then bounded functions, but the bounds will depend on the particular ω, and so the process will in general not be (uniformly) bounded. Think here of Brownian motion, the random variables $W(t)$ can assume arbitrary large values. The following proposition shows that we can approximate piecewise continuous processes by simple processes.

Proposition 6.4 *Let $\mathcal{P}$ be the set of piecewise continuous adapted processes a with $||a|| < \infty$. Then the set $\mathcal{S}$ is dense in $\mathcal{P}$ w.r.t. the norm $||\cdot||$.*

Proof Consider first the case where the process a is continuous on $[0, T]$ and bounded. Let $0 = t_0^n < \ldots < t_n^n = T$ be a subdivision of $[0, T]$. Consider the approximating (adapted) simple processes a^n defined by

$$a^n(\omega, t) = \sum_{i=1}^{n} a(\omega, t_{i-1}^n)\mathbf{1}_{(t_{i-1}^n, t_i^n]}(t). \tag{6.6}$$

Since a is bounded, by M say, also the a^n are bounded by M. Furthermore Since the paths of a are also *uniformly* continuous (why?), we can find for each of them and for each $\varepsilon > 0$ a $\delta > 0$ such that for each subdivision whose mesh is less than δ we have $|a(\omega, t) - a(\omega, t_{i-1}^n)| < \varepsilon$, if $t \in (t_{i-1}^n, t_i^n]$. Hence for such subdivisions we have $\sup\{|a^n(t) - a(t)| : t \in [0, T]\} < \varepsilon$, so that we have $a^n(t, \omega) \to a(t, \omega)$. Boundedness allows us to invoke the dominated convergence theorem to establish $||a^n - a|| \to 0$, which shows the assertion of the proposition for bounded continuous a.

If a is bounded, say again by M, and only piecewise continuous, we cut up the interval $[0, T]$ in a finite union of open intervals (s_{i-1}, s_i) on which a is continuous, with right and left limits at the endpoints. Treating these limits as the values of a at the endpoints we can repeat the above argument on the closed intervals $[s_{i-1}, s_i]$.

If a is not bounded we approximate a first with the bounded process $a^N = \max\{\min\{N, a\}, -N\}$ $(N > 0)$, which is piecewise continuous. Consider then

$$\begin{aligned}
||a - a^N||^2 &= \mathbb{E}\int_0^T |a(t) - a^N(t)|^2 \,\mathrm{d}t \\
&= \mathbb{E}\int_0^T (a(t) - N)^2 \mathbf{1}_{\{a(t) > N\}} \,\mathrm{d}t \\
&\quad + \mathbb{E}\int_0^T (a(t) + N)^2 \mathbf{1}_{\{a(t) < -N\}} \,\mathrm{d}t \\
&\leq \mathbb{E}\int_0^T a(t)^2 \mathbf{1}_{\{|a(t)| > N\}} \,\mathrm{d}t,
\end{aligned}$$

which tends to zero as N tends to infinity by the dominated convergence theorem. The a^N we can approximate arbitrarily well by the previous step. □

We are now ready to define the stochastic integral $\int_0^T a(t)\,\mathrm{d}W(t)$ for processes in $\mathcal{P}$. Let a be such a process, and let a^n be a sequence of simple processes such that $||a^n - a|| \to 0$. The integrals $I_T(a^n)$ are well defined and by linearity we have $\mathbb{E}(I_T(a^n) - I_T(a^m))^2 = \mathbb{E}I_T(a^n - a^m)^2$, which is by the isometry property equal to $||a^n - a^m||^2$. It becomes arbitrarily small for n, m large enough. This shows that the $I_T(a^n)$ form a Cauchy sequence in $\mathcal{L}^2(\Omega, \mathcal{F}_T, \mathbb{P})$, which is a Hilbert space if we identify a random variable as zero if its norm is zero. Hence the sequence has a limit in this space. We note that this limit is independent of the a approximating sequence. For, if we take another sequence of $b^n \in \mathcal{S}$ that approximate a, we would also get a limit. By mixing the two sequences we would get a limit again. But any subsequence of the latter will then have the same limit, in particular the subsequence consisting of the $I_T(a^n)$ and the subsequence consisting of the $I_T(b^n)$ that thus have the same limit. We conclude the construction of the stochastic integral by

Theorem 6.5 *For every $a \in \mathcal{P}$ there exists an a.s. unique random variable $I_T(a)$ with the property that $\mathbb{E}I_T(a)^2 = ||a||^2$. If we define the process $I(a)$ by $I_t(a) = I_T(a\mathbf{1}_{(0,t]})$, then assertions (i)-(iv) of Proposition 6.1 are still valid. Furthermore, for any sequence (a^n) that converges in $\mathcal{P}$ to a limit a one also has the convergence of $I_T(a^n)$ to $I_T(a)$ in $\mathcal{L}^2(\Omega, \mathcal{F}_T, \mathbb{P})$.*

Proof The existence and uniqueness of $I_T(a)$ has been shown above. In all what follows the a^n are processes in $\mathcal{S}$ such that $||a^n - a|| \to 0$. We will need the properties that also $\mathbb{E}(I_t(a^n) - I_t(a))^2 \to 0$ and that $\mathbb{E}I_t(a^n)^2 \to \mathbb{E}I_t(a)^2$ for $t \leq T$. Below we follow the numbering of Proposition 6.1 and we also exploit the contents of this proposition, the properties hold for simple processes.

(i) First we discuss the continuity property of $I(a)$. Since all $I(a^n)$ are continuous, we expect the same for $I(a)$. One easily obtains that $\mathbb{E}(I_t(a) - I_s(a))^2 = \mathbb{E}\int_s^t a(u)^2 \, \mathrm{d}u$, which implies already some form of continuity. To get that almost all sample paths are continuous, one needs results from probability that we don't discuss here, that imply uniform convergence of almost all paths of $I(a^n)$ to those of $I(a)$.

To show that $I(a)$ is a martingale we prove, analogous to Equation (A.12), that $\mathbb{E}\mathbf{1}_A I_t(a) = \mathbb{E}\mathbf{1}_A I_s(a)$ for all $A \in \mathcal{F}_s$. We have, exploiting the martingale property of $I(a_n)$,

$$
\begin{aligned}
\mathbb{E}\mathbf{1}_A I_t(a) &= \mathbb{E}\mathbf{1}_A(I_t(a) - I_t(a^n)) + \mathbb{E}\mathbf{1}_A I_t(a^n) \\
&= \mathbb{E}\mathbf{1}_A(I_t(a) - I_t(a^n)) + \mathbb{E}\mathbf{1}_A I_s(a^n) \\
&= \mathbb{E}\mathbf{1}_A(I_t(a) - I_t(a^n)) + \mathbb{E}\mathbf{1}_A(I_s(a^n) - I_s(a)) + \mathbb{E}\mathbf{1}_A I_s(a).
\end{aligned}
$$

Application of the Cauchy-Schwarz inequality gives $(\mathbb{E}\mathbf{1}_A(I_t(a) - I_t(a^n)))^2 \leq \mathbb{E}(I_t(a) - I_t(a^n))^2$, which tends to zero. Similarly, $\mathbb{E}\mathbf{1}_A(I_s(a^n) - I_s(a))$ has limit zero. So the first two terms on the right in the above display vanish, which was our goal.

(ii) We have seen above that $I_t(a^n) \stackrel{\mathcal{L}^2}{\to} I_t(a)$. But then also $I_s(a^n) \stackrel{\mathcal{L}^2}{\to} I_s(a)$ and hence $\mathbf{1}_A(I_t(a^n) - I_s(a^n)) \stackrel{\mathcal{L}^2}{\to} \mathbf{1}_A(I_t(a) - I_s(a))$ for any event $A \in \mathcal{F}_s$, nothing else but $\mathbb{E}\mathbf{1}_A(I_t(a^n) - I_s(a^n))^2 \to \mathbb{E}\mathbf{1}_A(I_t(a) - I_s(a))^2$. Moreover, as $||a^n - a|| \to 0$, we also have $||a^n|| \to ||a||$. Therefore we have for any $A \in \mathcal{F}_s$, using the already established property for simple a^n,

$$
\mathbb{E}\mathbf{1}_A(I_t(a) - I_s(a))^2 = \lim \mathbb{E}\mathbf{1}_A(I_t(a^n) - I_s(a^n))^2
$$

$$= \lim \mathbb{E}\mathbf{1}_A \int_s^t a^n(u)^2 \,\mathrm{d}u$$
$$= \mathbb{E}\mathbf{1}_A \int_s^t a(u)^2 \,\mathrm{d}u.$$

As $A \in \mathcal{F}_s$ is arbitrary, this implies that

$$\mathbb{E}[(I_t(a) - I_s(a))^2 | \mathcal{F}_s] = \mathbb{E}[\int_s^t a(u)^2 \,\mathrm{d}u | \mathcal{F}_s].$$

(iii) The third property $\mathbb{E} I_T(a)^2 = \mathbb{E} \int_0^T a(u)^2 \,\mathrm{d}u$ now immediately follows.

(iv) Linearity of the operator I_T on $\mathcal{P}$ follows similarly by approximation with simple processes for which we already know it. The last assertion of the theorem, isomorphism, now follows, as we already have proved in the previous step that norms are preserved. □

Henceforth we denote the stochastic integral $I_T(a)$ by $\int_0^T a(s) \,\mathrm{d}W(s)$ and $I_t(a)$ by $\int_0^t a(s) \,\mathrm{d}W(s)$, often abbreviated as $\int_0^t a \,\mathrm{d}W$.

Example 6.6 Let us study the integral $\int_0^T W(s) \,\mathrm{d}W(s)$. Take as approximations the functions W^n of Example 6.3. Note that we have

$$\begin{aligned} ||W - W^n||^2 &= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \mathbb{E}(W(t) - W(t_{i-1}))^2 \,\mathrm{d}t \\ &= \sum_{i=1}^n \int_{t_{i-1}}^{t_1} (t - t_{i-1}) \mathrm{d}t \\ &= \frac{1}{2} \sum_{i=1}^n (t_i - t_{i-1})^2 \to 0, \end{aligned}$$

if the mesh of the subdivision tends to zero. Hence to get $\int_0^T W(s) \,\mathrm{d}W(s)$ we have to compute the limit of the sum in Equation (6.5). But this we have done in Section 2.3, when we discussed the quadratic variation of Brownian motion. With this in mind we conclude

$$\int_0^T W(s) \,\mathrm{d}W(s) = \frac{1}{2} W(T)^2 - \frac{1}{2} T. \tag{6.7}$$

By the properties of the Itô integral we know that $\mathbb{E} \int_0^T W(s) \,\mathrm{d}W(s) = 0$, which also follows from the right hand side of (6.7).

Next we will find the quadratic variation of the process $I(a)$.

Proposition 6.7 *Let $a \in \mathcal{P}$. The quadratic variation of $I(a)$ over the interval $[0, t]$ is given by $\langle I(a) \rangle_t := \int_0^t a(u)^2 \,\mathrm{d}u$.*

Proof Consider first the case where a is piecewise constant. The result then immediately follows from the definition of the Itô integral for simple processes and Proposition 2.6.

Let now $a \in \mathcal{P}$ be arbitrary but bounded (we omit the proof of the general case). Consider partitions $0 = t_0^n < \cdots < t_n^n = t$ of $[0,t]$ with mesh tending to zero and a^n the processes that are constant on each of the intervals $(t_{k-1}^n, t_k^n]$ and equal to $a(t_{k-1}^n)$. It holds that $\int_0^t a^n(u)^2 \, \mathrm{d}u \to \int_0^t a(u)^2 \, \mathrm{d}u$ a.s. and also $\mathbb{E} \int_0^t (a^n(u) - a(u))^2 \, \mathrm{d}u \to 0$, because we can apply the dominated convergence theorem, see Theorem A.3 (ii) and Remark A.4, since a is bounded. Note that the latter implies $\mathbb{E} \int_0^t a^n(u)^2 \, \mathrm{d}u \to \mathbb{E} \int_0^t a(u)^2 \, \mathrm{d}u$.

We will show, see (6.9) below, that the difference

$$Q_n(a) - \int_0^t a^n(u)^2 \, \mathrm{d}u$$

tends to zero in the $\mathcal{L}^2$-sense, where $Q_n(a) = \sum_k (\int_{t_{k-1}^n}^{t_k^n} a \, \mathrm{d}W)^2$. First we write

$$Q_n(a) = \sum_k (\int_{t_{k-1}^n}^{t_k^n} (a - a^n) \, \mathrm{d}W + \sum_k \int_{t_{k-1}^n}^{t_k^n} a^n \, \mathrm{d}W)^2.$$

Next we write $Q_n(a)$ as the sum of the three terms

$$\begin{aligned} Q_n(a - a^n) &= \sum_k (\int_{t_{k-1}^n}^{t_k^n} (a - a^n) \, \mathrm{d}W)^2, \\ Q_n(a^n) &= \sum_k (\int_{t_{k-1}^n}^{t_k^n} a^n \, \mathrm{d}W)^2, \\ C_n &= 2 \sum_k (\int_{t_{k-1}^n}^{t_k^n} (a - a^n) \, \mathrm{d}W \int_{t_{k-1}^n}^{t_k^n} a^n \, \mathrm{d}W. \end{aligned}$$

The expectation $\mathbb{E} Q_n(a - a^n)$ equals $\mathbb{E} \int_0^t (a(u) - a^n(u))^2 \, \mathrm{d}u$ by Theorem 6.5 and Property (iii) of Proposition 6.1, applied to $a - a^n$ instead of a and $s = 0$, and tends to zero by the approximation property of the a^n. Below we will prove

$$Q_n(a^n) \xrightarrow{\mathcal{L}^2} \int_0^t a(u)^2 \, \mathrm{d}u, \tag{6.8}$$

which then also implies $Q_n(a^n) \xrightarrow{\mathcal{L}^1} \int_0^t a(u)^2 \, \mathrm{d}u$ and $\mathbb{E} Q_n(a^n) \to \mathbb{E} \int_0^t a(u)^2 \, \mathrm{d}u$. Under the assumption that (6.8) holds, the expectation of $|C_n|$ tends to zero by a version of the Cauchy-Schwarz inequality,

$$(\mathbb{E}|C_n|)^2 \leq \mathbb{E} Q_n(a - a^n) \, \mathbb{E} Q_n(a^n).$$

Let us therefore now focus on (6.8). As

$$Q_n(a^n) - \int_0^t a(u)^2 \, \mathrm{d}u = Q_n(a^n) - \int_0^t a^n(u)^2 \, \mathrm{d}u + \int_0^t a^n(u)^2 - \int_0^t a(u)^2,$$

it is sufficient to show that

$$\mathbb{E}(\sum_k (\int_{t_{k-1}^n}^{t_k^n} a^n \, \mathrm{d}W)^2 - \int_0^t a^n(u)^2 \, \mathrm{d}u)^2 \to 0, \tag{6.9}$$

as this also implies the $\mathcal{L}^1$-convergence of $\int_{t_{k-1}^n}^{t_k^n} a^n \, \mathrm{d}W)^2 - \int_0^t a^n(u)^2 \, \mathrm{d}u$ and because we already established $\int_0^t (a^n(u))^2 \, \mathrm{d}u \xrightarrow{\mathcal{L}^1} \int_0^t a(u)^2 \, \mathrm{d}u$ at the beginning of this proof.

The ordinary integral in (6.9) are actually sums and the big summation in (6.9) can be written as

$$\sum_k a^n(t_{k-1}^n)^2((\Delta W_k^n)^2 - \Delta t_k^n),$$

with $\Delta W_k^n = W(t_k^n) - W(t_{k-1}^n)$ and $\Delta t_k^n = t_k^n - t_{k-1}^n$. Since a is bounded, by a constant C say, we have that the expectation in (6.9) is by application of the Cauchy-Schwarz inequality for sums bounded by $C^4 \sum_k \mathbb{E}((\Delta W_k^n)^2 - \Delta t_k^n)^2$ (check!), and we can finish as we did in the proof of Proposition 2.6.

Wrapping up, we have

$$Q_n(a) - \int_0^t a(u)^2 \, \mathrm{d}u = Q_n(a - a^n) + \left(Q_n(a^n) - \int_0^t a(u)^2 \, \mathrm{d}u \right) + C_n,$$

where all three summands on the right tend to zero in $\mathcal{L}^1$ by the earlier established convergence of $Q_n(a - a^n)$ and C_n and the just shown convergence of the middle term. □

Remark 6.8 The Itô integral $I_T(a)$ can be shown to be also well defined for the wider class of processes a that are adapted and piecewise continuous, and satisfy $\int_0^T a(s)^2 \, \mathrm{d}s < \infty$ a.s. (there is no expectation here). In this case the assertion of Proposition 6.7 holds too.

A process X will be called a *semimartingale* if it can be written as X_0 plus the sum of the two integrals $\int_0^\cdot a(s) \, \mathrm{d}W(s)$ and $\int_0^\cdot b(s) \, \mathrm{d}s$ with a piecewise continuous, $\mathbb{E} \int_0^T a(s)^2 \, \mathrm{d}s < \infty$ (so $a \in \mathcal{P}$) and b an adapted process such that $\int_0^T |b(s)| \, \mathrm{d}s < \infty$ a.s. for all $T > 0$. So, a semimartingale can be decomposed as

$$X = X(0) + \int_0^\cdot b(s) \, \mathrm{d}s + \int_0^\cdot a(s) \, \mathrm{d}W(s). \tag{6.10}$$

We need the following lemma.

Lemma 6.9 *Let f and g be two continuous functions on an interval $[0, T]$. Assume that f has finite variation over this interval and that g has finite quadratic variation over this interval. Then the quadratic variation of $f + g$ is equal to the quadratic variation of g.*

Consequently, if X is a semimartingale as in Equation (6.10), *its quadratic variation $\langle X \rangle_t$ over an interval $[0, t]$ is given by*

$$\langle X \rangle_t = \int_0^t a(s)^2 \, \mathrm{d}s.$$

Proof Exercise 6.7. □

A consequence of this lemma and Proposition 6.7 is that the decomposition of a continuous semimartingale is unique (Exercise 6.8). Hence we can unambiguously *define* the stochastic integral of a process $h \in \mathcal{P}$ w.r.t. the semimartingale X as

$$\int_0^T h(t)\,\mathrm{d}X(t) = \int_0^T h(t)b(t)\,\mathrm{d}t + \int_0^T h(t)a(t)\,\mathrm{d}W(t), \tag{6.11}$$

where the latter is the stochastic integral, and it is well-defined under the condition $\mathbb{E}\int_0^T h(t)^2 a(t)^2\,\mathrm{d}t < \infty$. For Equation (6.11) we also use the shorthand notation in stochastic differentials

$$h(t)\mathrm{d}X(t) = h(t)b(t)\,\mathrm{d}t + h(t)a(t)\,\mathrm{d}W(t). \tag{6.12}$$

We close this section by mentioning that it is the preservation of the martingale property that makes the Itô integral such a powerful tool from a probabilistic point of view. This preservation is the consequence of the special structure of the simple processes. There are however alternatives to the Itô integral, the most prominent one being the Stratonovich integral. Contrary to the Itô integral, Stratonovich integrals w.r.t. Brownian motion are in general not martingales, see Exercise 6.12.

6.2 The Itô rule

Knowing how to understand the stochastic integral w.r.t. Brownian motion, we can now develop the important calculus rules. The first one is the basis of all the others to follow.

Lemma 6.10 *Let f be a twice continuously differentiable function, such that $\mathbb{E}\int_0^T f'(W(t))^2\,\mathrm{d}t$ and $\mathbb{E}\int_0^T f''(W(t))^2\,\mathrm{d}t$ are both finite. Then $f(W)$ is a semimartingale and one has a.s.*

$$f(W(T)) = f(W(0)) + \int_0^T f'(W(t))\,\mathrm{d}W(t) + \frac{1}{2}\int_0^T f''(W(t))\,\mathrm{d}t. \tag{6.13}$$

Proof (sketch) Let $\Pi = \{0 = t_0 < \cdots < t_n = T\}$ be a partition of $[0,T]$, which we will make as fine as needed. We use a Taylor expansion to write with $\Delta_i = t_i - t_{i-1}$, $\Delta W_i = W(t_i) - W(t_{i-1})$ and W_i^* an appropriate convex combination of $W(t_i)$ and $W(t_{i-1})$.

$$\begin{aligned} f(W(T)) - f(W(0)) &= \sum_{i=1}^n (f(W(t_i) - f(W(t_{i-1}))) \\ &= \sum_{i=1}^n f'(W(t_{i-1}))\Delta W_i + \frac{1}{2}\sum_{i=1}^n f''(W_i^*)\Delta W_i^2. \end{aligned} \tag{6.14}$$

By properties of the stochastic integral the first term converges in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ to $\int_0^T f'(W(t))\,\mathrm{d}W(t)$ if we make the partition finer and finer. The second term we split as

$$\frac{1}{2}\sum_{i=0}^{n-1} f''(W_i^*)\Delta_i + \frac{1}{2}\sum_{i=0}^{n-1} f''(W_i^*)(\Delta W_i^2 - \Delta_i).$$

Here the first sum converges pathwise to the Riemann integral $\int_0^T f''(W(t))\,\mathrm{d}t$. To treat the second one, we state that we can replace $f''(W_i^*)$ with $f_i = f''(W(t_{i-1}))$, making an negligeble error only (without making this precise). Below we shall use an argument like in the proof of Proposition 2.6, specifically $\mathbb{E}(\Delta W_i^2 - \Delta_i)^2 = \mathbb{V}\mathrm{ar}\,(\Delta W_i^2 - \Delta_i) = 2\Delta_i^2$. If we also assume that $\mathbb{E}f_i^2 < \infty$, we have, by Cauchy-Schwarz again, $\mathbb{E}(\sum_{i=0}^{n-1} f_i(\Delta W_i^2 - \Delta_i))^2 = 2\sum_{i=1}^n \mathbb{E}f_i^2\Delta_i^2 \leq 2\mu\sum_{i=1}^n \mathbb{E}f_i^2\Delta_i$, where μ is the mesh of Π. Consider now a family of partitions Π with $\mu \to 0$. Since $\sum_{i=1}^n \mathbb{E}f_i^2\Delta_i \to \mathbb{E}\int_0^T f''(W(t))^2\,\mathrm{d}t$, which is finite by assumption and $\mu \to 0$, we have $\mathbb{E}(\sum_{i=0}^{n-1} f_i(\Delta W_i^2 - \Delta_i))^2 \to 0$. This finishes the sketch proof. □

Lemma 6.10 can be extended to functions f of two variables.

Lemma 6.11 *Let f be a function of two variables t and x and suppose that f is continuously differentiable w.r.t. t with partial derivative f_t, twice continuously differentiable function w.r.t. x with partial derivatives f_x and f_{xx}, such that $\mathbb{E}\int_0^T f_x(t, W(t))^2\,\mathrm{d}t < \infty$. Then one has a.s.*

$$f(T, W(T)) = f(0, W(0)) + \int_0^T f_t(t, W(t))\,\mathrm{d}t + \int_0^T f_x(t, W(t))\,\mathrm{d}W(t) + \frac{1}{2}\int_0^T f_{xx}(t, W(t))\,\mathrm{d}t. \qquad (6.15)$$

Proof One can use the same techniques as in the proof of Lemma 6.10, again based on a Taylor expansion but now for a function of two variables. We omit the details. □

For Equation (6.15) we use, similar to (6.12), the abbreviation

$$\mathrm{d}f(t, W(t)) = f_t(t, W(t))\,\mathrm{d}t + f_x(t, W(t))\,\mathrm{d}W(t) + \frac{1}{2}f_{xx}(t, W(t))\,\mathrm{d}t. \qquad (6.16)$$

Remark 6.12 The conditions involving finite expectations in Lemmas 6.10 and 6.11 are rather annoying. It can be shown that these are actually not needed. For this we need the notion of *stopping times*, which is not part of this course.

Now we are in the position to state a version of the stochastic chain rule.

Theorem 6.13 *Let $X(t) = f(t, W(t))$ and $Y(t) = g(t, X(t))$ where f and g are supposed to satisfy the appropriate differentiability conditions and that are such that the (stochastic) integrals below are well-defined. Then X and Y are semimartingales and it holds that*

$$\mathrm{d}Y(t) = g_t(t, X(t))\,\mathrm{d}t + g_x(t, X(t))\,\mathrm{d}X(t) + \frac{1}{2}g_{xx}(t, X(t))f_x(t, W(t))^2\,\mathrm{d}t,$$

where the term with $dX(t)$ has to be understood as a shorthand notation for an expression like (6.12).

Proof Let h be the decomposition $h(t, x) = g(t, f(t, x))$. Compute the partial derivatives of h, express them in those of f and g and use Equation (6.12). □

Now we have seen in Theorem 6.13 that functions and compositions of functions of time and Brownian motion are again semimartingales, we can use Lemma 6.9 to compute the quadratic variation of X and Y as they appear in this theorem.

Proposition 6.14 *Let X and Y be as in Theorem 6.13. Then X has quadratic variation process given by $\langle X\rangle_t = \int_0^t f_x(s, W(s))^2 \,\mathrm{d}s$ and Y has quadratic variation given by $\langle Y\rangle_t = \int_0^t h_x(s, W(s))^2 \,\mathrm{d}s$, where $h(t, x) \equiv g(t, f(t, x))$. Moreover these quadratic variations are related by $\langle Y\rangle_t = \int_0^t g_x(s, X(s))^2 \,\mathrm{d}\langle X\rangle_s$.*

Proof The expressions for $\langle X\rangle_t$ and $\langle Y\rangle_t$ follow from Lemma 6.9. The relation between the two quadratic variation processes then immediately follows by $h_x(s, x) = g_x(s, f(x))f_x(s, x)$. □

Corollary 6.15 *In the situation of Theorem 6.13 we have*

$$\mathrm{d}Y(t) = g_t(t, X(t))\,\mathrm{d}t + g_x(t, X(t))\,\mathrm{d}X(t) + \frac{1}{2}g_{xx}(t, X(t))\mathrm{d}\langle X\rangle_t. \qquad (6.17)$$

Proof Just use the expression for $\langle X\rangle_t$ of Proposition 6.14. □

Corollary 6.15 can be shown to be generalized to the situation, where $Y(t) = g(A(t), X(t))$ with A a process given by $A(t) = \int_0^t a(s)\,\mathrm{d}s$, where a is a continuous process. We state, rather informally, the result and omit the proof.

Proposition 6.16 *Let X be a semimartingale, A a process of the form $A(t) = \int_0^t a(s)\,\mathrm{d}s$ and g a function that is supposed to satisfy the appropriate differentiability conditions. Let Y be defined by $Y(t) = g(A(t), X(t))$. Assuming that the (stochastic) integrals below are well-defined, we have that Y is a semimartingale as well and*

$$\begin{aligned}\mathrm{d}Y(t) = {} & g_t(A(t), X(t))a(t)\,\mathrm{d}t \\ & + g_x(A(t), X(t))\,\mathrm{d}X(t) + \frac{1}{2}g_{xx}(A(t), X(t))\mathrm{d}\langle X\rangle_t. \end{aligned} \qquad (6.18)$$

Example 6.17 We can apply Proposition 6.16 to the following interesting case. Let $X(t)$ be of the form $X(t) = f(t, W(t))$, $A(t) = \int_0^t f_x(s, W(s))^2 \,\mathrm{d}s$, $g(t, x) = \exp(x - \frac{1}{2}t)$ and $Y(t) = g(A(t), X(t))$. Then we have (check it yourself!)

$$Y(t) = Y(0) + \int_0^t Y(s)\,\mathrm{d}X(s). \qquad (6.19)$$

If $X(0) = 0$, then one has

$$Y(t) = 1 + \int_0^t Y(s)\,\mathrm{d}X(s). \qquad (6.20)$$

For processes X and Y that are related by Equation (6.20) we use the notation $Y = \mathcal{E}(X)$. The process Y is said to be the *Doléans exponential* of X. Equation (6.19) also results when $Y(t) = \exp(X(t) - \frac{1}{2}\langle X\rangle_t)$ for an arbitrary semimartingale X, the precise form of its decomposition is not relevant. Important for the application of the Itô rule of Proposition 6.16 is to realize that the quadratic variation of $X - \frac{1}{2}\langle X\rangle$ is $\langle X\rangle$. Working out the details is the content of Exercise 6.22. It may be instructive to compare this example to the analogous formulas in the discrete time setting of Equations (1.6) and (1.7).

The Itô rules that we have given above, in particular (6.18) can be extended to more dimensional semimartingales X. To give an example, take $X = (X_1, X_2)$, where X_1 and X_2 are semimartingales. Let $Y(t) = f(t, X_1(t), X_2(t))$, with f having the appropriate continuous partial derivatives. The version of the Itô formula for Y involves the *quadratic covariation process* $\langle X_1, X_2\rangle$. For $T > 0$, and a partition $\Pi = \{0 = t_0 < \cdots < t_n = T\}$, one considers

$$
\begin{aligned}
C(\Pi)_T &= \sum_{i=1}^{n} (X_1(t_i) - X_1(t_{i-1}))(X_2(t_i) - X_2(t_{i-1})) \\
&=: \sum_{i=1}^{n} \Delta X_{1,i} \Delta X_{2,i} \\
&= \frac{1}{2} \sum_{i=1}^{n} \left((\Delta X_{1,i} + \Delta X_{2,i})^2 - \Delta X_{1,i}^2 - \Delta X_{2,i}^2 \right).
\end{aligned}
$$

If $C(\Pi)_T$ has a limit (in probability) for partitions Π with mesh tending to zero, then this limit is denoted $\langle X_1, X_2\rangle_T$, and one should have the relation

$$
\langle X_1, X_2\rangle_T = \frac{1}{2} \left(\langle X_1 + X_2\rangle_T - \langle X_1\rangle_T - \langle X_2\rangle_T \right). \tag{6.21}
$$

Indeed, for semimartingales X_1 and X_2 this limit exists, and Equation (6.21) is valid, the content of the next proposition.

Proposition 6.18 *Let X_1 and X_2 be semimartingales with decompositions as in* (6.10) *involving the processes a_i and b_i ($i = 1, 2$). Then $\langle X_1, X_2\rangle_T$ exists, it satisfies* (6.21) *for any $T > 0$, and is explicitly given by*

$$
\langle X_1, X_2\rangle_T = \int_0^T a_1(s) a_2(s) \, \mathrm{d}s.
$$

Proof First we observe that, in analogy with Lemma 6.9, we can ignore the ordinary integral. Following the proof of Proposition 6.7, we then replace the $Q_n(a)$ there with $C(\Pi)_T$ and mimic the remainder of the proof. Showing the validity of (6.21) is straightforward. □

Turning back to the multivariate setting $Y(t) = f(t, X_1(t), X_2(t))$, it holds that Y is a semimartingale as well and in shorthand notation, where the arguments $t, X_1(t)$ and $X_2(t)$ of the partial derivatives have been *omitted* for clarity of the displayed formula, we have

$$
\begin{aligned}
\mathrm{d}Y(t) &= f_t \, \mathrm{d}t + f_{x_1} \, \mathrm{d}X_1(t) + f_{x_2} \, \mathrm{d}X_2(t) \\
&\quad + \frac{1}{2} \big(f_{x_1 x_1} \, \mathrm{d}\langle X_1\rangle_T + 2 f_{x_1 x_2} \, \mathrm{d}\langle X_1, X_2\rangle_T + f_{x_2 x_2} \, \mathrm{d}\langle X_2\rangle_T \big).
\end{aligned} \tag{6.22}
$$

We notice that we have seen already an instance of this formula, namely (6.18). You check why! One special case deserves our attention.

Consider $f(t, x_1, x_2) = x_1 x_2$. Then Equation (6.22) becomes the *product formula* for semimartingales,

$$
\mathrm{d}(X_1 X_2) = X_1 \, \mathrm{d}X_2 + X_2 \, \mathrm{d}X_1 + \mathrm{d}\langle X_1, X_2\rangle,
$$

which is the shorthand version of

$$X_1(t)X_2(t) - X_1(0)X_2(0) = \\ \int_0^t X_1(s)\,\mathrm{d}X_2(s) + \int_0^t X_2(s)\,\mathrm{d}X_1(s) + \langle X_1, X_2\rangle_t. \tag{6.23}$$

See also Exercise 6.21 for the discrete time version of Equation (6.23). This equation can be specialized further. Let $X_1 = X$ be an arbitrary semimartingale and $X_2 = A$ a semimartingale of the form $A = \int_0^\cdot a(t)\,\mathrm{d}t$. It follows that in this case one has (check why)

$$\mathrm{d}(X(t)A(t)) = X(t)a(t)\,\mathrm{d}t + A(t)\,\mathrm{d}X(t). \tag{6.24}$$

Next we present some formal computations that facilitate the computations of quadratic variation and quadratic covariation. The following symbolic multiplication rules are useful, where W is a Brownian motion and X, X_1, X_2 are semimartingales.

$$\begin{aligned}(\mathrm{d}t)^2 &= 0\\ \mathrm{d}t \times \mathrm{d}W(t) &= 0\\ (\mathrm{d}W(t))^2 &= \mathrm{d}t\\ (\mathrm{d}X(t))^2 &= \mathrm{d}\langle X\rangle_t\\ \mathrm{d}X_1(t) \times \mathrm{d}X_2(t) &= \mathrm{d}\langle X_1, X_2\rangle_t\end{aligned}$$

Note that all quantities above have no mathematical meaning, e.g. $\mathrm{d}t$ is not a number, but is only used under the integral sign. Similar comments apply to the other quantities, but as before they are also conveniently used in differential notation. The justification of these multiplications can be traced back to the computations in the proofs of all kinds of quadratic variations and covariations, where we encountered expressions like sums of $(X(t_i) - X(t_{i-1}))^2$, which looks like $(\mathrm{d}X(t))^2$. The limit of such a sum was seen to be $\langle X\rangle_t$, whose 'differential' is $\mathrm{d}\langle X\rangle_t$.

The usefulness of these rules is illustrated by consider semimartingales X_1 and X_2 as in Proposition 6.18. In differential notation, we have

$$\begin{aligned}\mathrm{d}X_1(t) &= b_1(t)\,\mathrm{d}t + a_1(t)\,\mathrm{d}W(t)\\ \mathrm{d}X_2(t) &= b_2(t)\,\mathrm{d}t + a_2(t)\,\mathrm{d}W(t).\end{aligned}$$

With the above rule we find

$$\begin{aligned}\mathrm{d}\langle X_1, X_2\rangle_t &= \mathrm{d}X_1(t) \times \mathrm{d}X_2(t)\\ &= (b_1(t)\,\mathrm{d}t + a_1(t)\,\mathrm{d}W(t))(b_2(t)\,\mathrm{d}t + a_2(t)\,\mathrm{d}W(t))\\ &= a_1(t)a_2(t)\,\mathrm{d}t,\end{aligned}$$

which we recognize as the result of Proposition 6.18 in differential form.

6.3 Girsanov's theorem revisited

We have seen in Section 4 an absolutely continuous change of measure that transformed a given Brownian motion in a Brownian motion with linear drift.

In Exercise 4.12 an example was given of a change measure that transformed Brownian motion into a Brownian motion with piecewise linear drift (and vice versa). In that exercise $Z = \exp\left(-\frac{1}{\sigma}I(a) - \frac{1}{2}\frac{1}{\sigma^2}\int_0^T a^2(s)\,\mathrm{d}s\right)$, with $I(a) = \sum_{i=1}^n a_i \Delta W_i$, played a role. The sum we now understand as an Itô integral of the simple function $a(t) = \sum_{i=1}^n a_i \mathbf{1}_{(t_{i-1},t_i]}(t)$, where the a_i are real numbers and $0 = t_0 < t_1 < \cdots < t_n = T$.

Having a closer look at the exercise, we suspect that any random variable $Z = \exp(-\int_0^T a(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^T a(s)^2\,\mathrm{d}s)$ with $a \in \mathcal{P}$ can be used to define a new probability measure $\mathbb{Q}$ with $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z$. However, at this point we have to be careful! Whatever the situation is, we can always define a measure $\mathbb{Q}$ on $\mathcal{F}_T$ by $\mathrm{d}\mathbb{Q} = Z(T)\mathrm{d}\mathbb{P}$, i.e. by $\mathbb{Q}(F) = \mathbb{E}_{\mathbb{P}}\mathbf{1}_F(T)$. But to have that $\mathbb{Q}$ is a *probability* measure, we need $\mathbb{E}_{\mathbb{P}}Z(T) = 1$. If a is a non-random function in $\mathcal{L}^2[0,T]$, this follows from Exercise 6.10. But for random processes a this is in general not true. In this section we will have a closer look at Girsanov transformations.

Let us consider a simple process $a \in \mathcal{S}$ (in particular, a is bounded) and let us look at $Z(t) = \exp(-\int_0^t a(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t a(s)^2\,\mathrm{d}s)$. Application of the Itô rule gives

$$Z(t) = 1 - \int_0^t Z(s)a(s)\,\mathrm{d}W(s). \tag{6.25}$$

Using reconditioning properties of conditional expectation, one can show that in this case the process Z is a martingale on $[0,T]$. Moreover, one may show that even $\mathbb{E}_{\mathbb{P}}Z(T)^2 < \infty$ for any bounded process a, Exercise 6.20. Observe by the way, that we can use the Doléans exponential to write $Z = \mathcal{E}(Y)$, with $Y = -\int_0^{\cdot} a(s)\,\mathrm{d}W(s)$, see Example 6.17.

Next we look at a process $a \in \mathcal{P}$ and define, as above,

$$Z(t) = \exp(-\int_0^t a(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t a(s)^2\,\mathrm{d}s).$$

Without checking technical conditions, we simply apply the Itô rule again, to get the same representation (6.25). It is in this case not clear whether Z is a martingale. For instance, if we have processes $a^n \in \mathcal{S}$ that converge to a w.r.t. the norm of Section 6.1 we have $\mathcal{L}^2$-convergence of the resulting stochastic integrals $I_T(a^n)$ to $I_T(a)$. With $Z^n = \mathcal{E}(Y^n)$, $Y^n = -\int_0^{\cdot} a^n(s)\,\mathrm{d}W(s)$, it is possible to show that $Z^n(T)$ converges to $Z(T)$ in probability, but this type of convergence is too weak to conclude $\mathbb{E}_{\mathbb{P}}Z(T) = 1$. From Fatou's lemma we can only deduce that $\mathbb{E}_{\mathbb{P}}Z(T) \leq \liminf \mathbb{E}_{\mathbb{P}}Z^n(T) = 1$. So we need stronger assumptions.

Look at Equation (6.25). To have that Z is a martingale, it is sufficient to impose that aZ belongs to $\mathcal{P}$ (we then also have $\mathbb{E}_{\mathbb{P}}\int_0^T Z(s)^2 a(s)^2\,\mathrm{d}s < \infty$). But this is equivalent to $\mathbb{E}_{\mathbb{P}}Z(T)^2 < \infty$. So, from now on we assume that a is such that this condition is satisfied. This condition is a bit unsatisfactory, because it is not an explicit condition in terms of a only. On the other hand it is satisfied if a is deterministic and in $\mathcal{L}^2[0,T]$, or if a is a simple process.

Theorem 6.19 *Let $a \in \mathcal{P}$, in particular $\mathbb{E}_{\mathbb{P}}\int_0^T a(t)^2\,\mathrm{d}t < \infty$ and let $Z(t) = \exp\left(-\int_0^t a(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t a(s)^2\,\mathrm{d}s\right)$. Assume that $\mathbb{E}_{\mathbb{P}}\int_0^T a(t)^2 Z(t)^2\,\mathrm{d}t < \infty$.*

Define the measure $\mathbb{Q}$ on $\mathcal{F}_T$ by $\mathrm{d}\mathbb{Q} = Z(T)\,\mathrm{d}\mathbb{P}$. Then $\mathbb{Q}$ is a probability measure and the process $W^{\mathbb{Q}}$ defined by $W^{\mathbb{Q}}(t) = W(t) + \int_0^t a(s)\,\mathrm{d}s$ is a Brownian motion under $\mathbb{Q}$ for $t \in [0,T]$.

Proof First we show that $\mathbb{Q}$ is a probability measure, which follows as soon as we know that $\mathbb{E}_{\mathbb{P}} Z(T) = 1$. From the assumption $\mathbb{E}_{\mathbb{P}} \int_0^T a(t)^2 Z(t)^2\,\mathrm{d}t < \infty$ and (6.25) we see that the process $\{Z(t), 0 \le t \le T\}$ is a martingale and hence $\mathbb{E}_{\mathbb{P}} Z(T) = \mathbb{E}_{\mathbb{P}} Z(0) = 1$.

We next show that $W^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$. For simplicity we further assume in this proof that a is a bounded process, $|a| \le K$ say. Consider the sequence of $a^n \in \mathcal{S}$ that converges to a as in (6.6) in the proof of Proposition 6.4. Then the a^n are also bounded by K, and by the Cauchy-Schwarz inequality also $\mathbb{E} \int_0^T |a^n(s) - a(s)|\,\mathrm{d}s \to 0$, hence $\int_0^T |a^n(s) - a(s)|\,\mathrm{d}s$ converges to zero in $\mathbb{P}$-probability, and then almost surely along a subsequence (see Proposition A.2 for all relations between different modes of convergence); this subsequence will henceforth be indexed by n again.

Put $Z^n(t) = \exp\Big(-\int_0^t a^n(s)\,\mathrm{d}W(s) - \frac{1}{2}\int_0^t a^n(s)^2\,\mathrm{d}s\Big)$. In the present situation all $\mathbb{E}_{\mathbb{P}} Z^n(T)^2$ and $\mathbb{E}_{\mathbb{P}} Z(T)^2$ are finite. We will show that $Z^n(T)$ converges to $Z(T)$ in $\mathcal{L}^2(\Omega, \mathcal{F}_T, \mathbb{P})$.

From Equation (6.25) and the analogous expression for Z^n we obtain

$$
\begin{aligned}
Z^n(T) - Z(T) &= -\int_0^T (a^n(t)Z^n(t) - a(t)Z(t))\,\mathrm{d}W(t) \\
&= -\int_0^T (a^n(t) - a(t))Z(t)\,\mathrm{d}W(t) \\
&\quad - \int_0^T (Z^n(t) - Z(t))a^n(t)\,\mathrm{d}W(t).
\end{aligned}
$$

Hence we get, by the elementary inequality $(a+b)^2 \le 2a^2 + 2b^2$,

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}|Z^n(T) - Z(T)|^2 &\le 2\mathbb{E}_{\mathbb{P}} \int_0^T |a^n(t) - a(t)|^2 Z(t)^2\,\mathrm{d}t \\
&\qquad + 2\mathbb{E}_{\mathbb{P}} \int_0^T |Z^n(t) - Z(t)|^2 a^n(t)^2\,\mathrm{d}t \\
&\le 2\mathbb{E}_{\mathbb{P}} \int_0^T |a^n(t) - a(t)|^2 Z(t)^2\,\mathrm{d}t \\
&\qquad + 2K^2\mathbb{E}_{\mathbb{P}} \int_0^T |Z^n(t) - Z(t)|^2\,\mathrm{d}t.
\end{aligned}
$$

If we write $F^n(t) = \mathbb{E}_{\mathbb{P}}|Z^n(t) - Z(t)|^2$, $A^n(T) = 2\mathbb{E}_{\mathbb{P}} \int_0^T |a^n(t) - a(t)|^2 Z(t)^2\,\mathrm{d}t$ and $B = 2K^2$, then we have

$$
F^n(T) \le A^n(T) + B\int_0^T F^n(t)\,\mathrm{d}t. \tag{6.26}
$$

It follows from Gronwall's inequality (Exercise 6.14) that we can deduce that $F^n(T) \le A^n(T)\exp(BT)$. So to achieve our aim, $\mathbb{E}_{\mathbb{P}}|Z^n(T) - Z(T)|^2 \to 0$, it is sufficient to show that A^n converges to zero, and this follows from the dominated convergence theorem, because $\mathbb{E}_{\mathbb{P}} \int_0^T Z(t)^2 \, \mathrm{d}t < \infty$ (check this!).

Recall from Exercise 4.12 that under $\mathbb{Q}^n$ we have that W^n defined by $W^n(t) = W(t) + \int_0^t a^n(s) \, \mathrm{d}s$ is a Brownian motion. This means that any event of the type $E^n = \{W^n(t_1) \leq x_1, \ldots, W^n(t_m) \leq x_m\}$ has a stationary $\mathbb{Q}^n$-probability, i.e. $\mathbb{Q}^n(E^n)$ is the same for all n. Let $E = \{W^{\mathbb{Q}}(t_1) \leq x_1, \ldots, W^{\mathbb{Q}}(t_m) \leq x_m\}$. We will show that $\mathbb{Q}(E)$ is equal to $\mathbb{Q}^n(E^n)$.

Consider

$$
\begin{aligned}
|\mathbb{Q}(E) - \mathbb{Q}^n(E^n)| &\leq \mathbb{E}_{\mathbb{P}} |Z(T)\mathbf{1}_E - Z^n(T)\mathbf{1}_{E^n}| \\
&\leq \mathbb{E}_{\mathbb{P}} |Z(T)(\mathbf{1}_E - \mathbf{1}_{E^n})| + \mathbb{E}_{\mathbb{P}} |Z^n(T) - Z(T)|. \qquad (6.27)
\end{aligned}
$$

The second term in (6.27) tends to zero, because of the just shown $\mathcal{L}^2$-convergence of $Z^n(T)$ to $Z(T)$, as it implies $\mathcal{L}^1$-convergence. Since (along a subsequence) the $\mathbf{1}_{E^n}$ converge to $\mathbf{1}_E$ $\mathbb{P}$-a.s. (Exercise 6.18), we can use a version of the dominated convergence theorem to conclude, or see Remark 6.20 below, that also the first term in (6.27) eventually vanishes, and hence $\mathbb{Q}(E) = \mathbb{Q}^n(E^n)$. The conclusion is that $W^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$. □

Remark 6.20 Instead of using dominated convergence at the end of the proof, we can argue as follows. Write $D_n = \mathbf{1}_E - \mathbf{1}_{E^n}$, and note that $|D_n| \leq 1$. Moreover, $D_n \stackrel{\mathbb{P}}{\to} 0$ (check). Then, for any $\delta > 0$,

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}} |Z(T)(\mathbf{1}_E - \mathbf{1}_{E^n})| &= \mathbb{E}_{\mathbb{P}} Z(T)|D_n|\mathbf{1}_{\{|D_n|>\delta\}} + \mathbb{E}_{\mathbb{P}} |Z(T)D_n|\mathbf{1}_{\{|D_n|\leq\delta\}} \\
&= \mathbb{E}_{\mathbb{P}} Z(T)\mathbf{1}_{\{|D_n|>\delta\}} + \delta \, \mathbb{E}_{\mathbb{P}} Z(T) \\
&\leq \left(\mathbb{E}_{\mathbb{P}} Z(T)^2 \mathbb{P}(|D_n| > \delta)\right)^{1/2} + \delta.
\end{aligned}
$$

Hence

$$
\limsup_{n\to\infty} \mathbb{E}_{\mathbb{P}} |Z(T)(\mathbf{1}_E - \mathbf{1}_{E^n})| \leq \delta,
$$

which yields the result, since δ is arbitrary.

Remark 6.21 Crucial for the property of $\mathbb{Q}$ being a probability measure is the requirement that $\mathbb{E}_{\mathbb{P}} Z(T) = 1$. A well known sufficient condition for this to be true, weaker than the one in Theorem 6.19, is $\mathbb{E}_{\mathbb{P}} \exp(\frac{1}{2} \int_0^T a(s)^2 \, \mathrm{d}s) < \infty$, known as *Novikov's condition.* This condition is sharp in the sense that if the factor $\frac{1}{2}$ in the exponent is replaced with $\alpha < \frac{1}{2}$, counterexamples exist in which $\mathbb{E}_{\mathbb{P}} Z(T) < 1$. The condition is obviously satisfied if a is deterministic and $\int_0^T a(s)^2 \, \mathrm{d}s < \infty$.

6.4 Exercises

6.1 Show that (the value of) the Itô integral of (6.3) doesn't depend on the specific representation of the simple process a. *Hint: If* $a(t) = \sum_{i=1}^m b_i \mathbf{1}_{(s_{i-1},s_i]}(t)$ *is another representation, consider* $a(t) = \sum_{i=1}^m \sum_{j=1}^n b_i \mathbf{1}_{(s_{i-1},s_i]\cap(t_{j-1},t_j]}(t)$.

6.2 Show the validity of Equation (6.4).

6.3 Here are some consequences of Example 6.3.

(a) Show Equality (6.5).

(b) What is the limit of its right hand side when the t_i come from a sequence of partitions of $[0,T]$ with mesh tending to zero? What is in the same situation the limit of $W^n(t)$?

(c) What is the 'reasonable' value of $\int_0^T W(s)\,\mathrm{d}W(s)$?

6.4 Prove assertion (ii) of Proposition 6.1

6.5 This exercise concerns Theorem 6.5.

(a) Show that $I_T(a)$ is a linear functional of $a \in \mathcal{S}$.

(b) Show that $I_T(a)$ is a linear functional of $a \in \mathcal{P}$.

(c) Consider the stochastic integrals $I_t(a)$ with $a \in \mathcal{P}$. Denote by $\langle I(a)\rangle$ the quadratic variation process of $I(a)$ (i.e. $\langle I(a)\rangle_t$ is the quadratic variation of $I(a)$ over the interval $[0,t]$). Show that $I(a)_t^2 - \langle I(a)\rangle_t$, $t \in [0,T]$ is a martingale.

6.6 Let $M(t) = \int_0^t W(s)\,\mathrm{d}W(s)$.

(a) Use Theorem 6.5 to show that $M(t)$ and $M(t)^2 - \int_0^t W(s)^2\,\mathrm{d}s$ ($t \in [0,T]$) are martingales. See also Remark 6.2.

(b) Let $A(t)$, $t \in [0,T]$, be an adapted positive process with piecewise continuous paths and $\mathbb{E}A(t) < \infty$ for all t. Show/argue that $\mathbb{E}[\int_0^t A(u)\,\mathrm{d}u|\mathcal{F}_s] = \int_0^s A(u)\,\mathrm{d}u + \int_s^t \mathbb{E}[A(u)|\mathcal{F}_s]\,\mathrm{d}u$, by using the definition of conditional expectation. (*You may use that for positive $A(u)$ it holds that $\mathbb{E}\int_0^T A(u)\,\mathrm{d}u = \int_0^T \mathbb{E}A(u)\,\mathrm{d}u$.*)

(c) Show by a direct computation, not involving stochastic integration theory, that $M(t)^2 - \int_0^t W(s)^2\,\mathrm{d}s$ ($t \in [0,T]$) is a martingale. (*Here it is often useful to write $W(t) = (W(t)-W(s)) + W(s)$ and that $W(t)-W(s)$ is independent of $\mathcal{F}_s$. Recall that $\mathbb{E}X^4 = 3\sigma^4$ if X has the $N(0,\sigma^2)$ distribution.*)

6.7 Prove Lemma 6.9 for the functions f and g.

6.8 Show that the semimartingale decomposition (6.10) is unique.

6.9 Compute from one of the Itô rules $\mathrm{d}Z_t$ in the following cases.

(a) $Z_t = e^{aW_t}$. Find a differential equation with $\mathbb{E}Z_t$ as solution.

(b) $Z_t = e^{aW_t - bt}$. Show that Z is a martingale in the special case $b = \frac{1}{2}a^2$.

(c) $Z_t = X_t^2$, for $X_t = W_t^2 - t$.

6.10 Let σ be a piecewise continuous function (non-random!) on any interval $[0,t]$. Let W be Brownian motion and put $X(t) = \int_0^t \sigma(s)\,\mathrm{d}W(s)$. Let $Z(t) = \exp(\mathrm{i}uX(t))$, for fixed $u \in \mathbb{R}$. Use the Itô rule to find

$$Z(t) = 1 + \int_0^t \mathrm{i}u\sigma(s)Z(s)\,\mathrm{d}W(s) - \frac{1}{2}u^2 \int_0^t \sigma(s)^2 Z(s)\,\mathrm{d}s.$$

Use this equation to compute the characteristic function $\phi(u) = \mathbb{E}\exp(\mathrm{i}uX(t))$. Deduce that $X(t)$ is normally distributed and determine its expectation and variance.

6.11 Let X be given by $X(t) = \exp(\gamma t + \sigma W(t))$, with γ a constant and W a Brownian motion. Find the semimartingale representation of X by using the Itô rule and show that the quadratic variation process of X is given by $\sigma^2 \int_0^t X(u)^2 \, \mathrm{d}u$.

6.12 Consider a Brownian motion W and a partition $0 = t_0^n < t_1^n < \cdots < t_n^n = t$ of that $[0,t]$. Define for $0 \leq \alpha \leq 1$ the random variables

$$S_n(\alpha) = \sum_{i=1}^n W_{(1-\alpha)t_{i-1}^n + \alpha t_i^n}(W_{t_i^n} - W_{t_{i-1}^n}).$$

(a) Show that for partitions whose mesh $\max\{t_i^n - t_{i-1}^n : i = 1, \ldots n\}$ converges to 0, it holds that $S_n(\alpha)$ in $\mathcal{L}^2$-sense converges to $J(\alpha)$ with

$$J(\alpha) = \frac{1}{2} W_t^2 + (\alpha - \frac{1}{2})t.$$

[*Remark*: for $\alpha = 0$ is $J(\alpha)$ the Itô integral $\int_0^t W_s \, \mathrm{d}W_s$ and for $\alpha = \frac{1}{2}$ one obtains what is known as the Stratonovich integral.]

(b) Put

$$S_n'(\alpha) = \sum_{i=1}^n (\alpha W_{t_i^n} + (1-\alpha) W_{t_{i-1}^n})(W_{t_i^n} - W_{t_{i-1}^n}).$$

Show that $S_n(\alpha)'$ in $\mathcal{L}^2$-sense converges to $J(\alpha)$.

6.13 Brownian motion W is, as we have seen, a continuous martingale with quadratic variation process $\langle W \rangle_t \equiv t$. It is in fact the *only* continuous martingale with this property, which can be shown as follows. Let M be such a martingale (it is continuous, $\langle M \rangle_t \equiv t$) and define for every $\lambda \in \mathbb{R}$ and $s \leq t$ the random variables $\phi(s,t) = \mathbb{E}[e^{\mathrm{i}\lambda(M(t)-M(s))} | \mathcal{F}_s]$. We have reached our goal as soon as we can prove $\phi(s,t) = e^{-\frac{1}{2}\lambda^2(t-s)}$ (why?). Let $Y(t) = e^{\mathrm{i}\lambda M(t)}$.

(a) Use the Itô rule to get

$$Y(t) - Y(s) = \mathrm{i}\lambda \int_s^t e^{\mathrm{i}\lambda M(u)} \, \mathrm{d}M(u) - \frac{1}{2}\lambda^2 \int_s^t e^{\mathrm{i}\lambda M(u)} \, \mathrm{d}u.$$

(b) Divide both sides of this equation by $Y(s)$ and take conditional expectation given $\mathcal{F}_s$ to arrive at the integral equation $\phi(s,t) = 1 - \frac{1}{2}\lambda^2 \int_s^t \phi(s,u) \, \mathrm{d}u$. Solve this equation.

6.14 Let $f : [0,T] \to [0,\infty)$ satisfy the inequality $f(t) \leq a(t) + \beta \int_0^t f(s) \, \mathrm{d}s$, with a an integrable function and $\beta \geq 0$. Show Gronwall's inequality, $f(t) \leq a(t) + \beta \int_0^t a(s) e^{\beta(t-s)} \, \mathrm{d}s$ for all $t \leq T$. If the function a is non-decreasing, then $f(t) \leq a(t) e^{\beta t}$.

6.15 For two functions X and Y, their *cross quadratic variation* $\langle X, Y \rangle$ over an interval $[0,T]$ is defined as the limit of sums $\sum_i (X(t_i) - X(t_{i-1}))(Y(t_i) - Y(t_{i-1}))$ for partitions of $[0,T]$ with mesh tending to zero. Show that $\langle X, Y \rangle = 0$ if X and Y are continuous, and Y has bounded (first order) variation over $[0,T]$.

6.16 Let $X(t) = v(t, W(t))$, where $v \in C^{1,2}(\mathbb{R}^+, \mathbb{R})$ and W is a Brownian motion. Find, by using the Itô formula, conditions on v (in particular it should be the solution to a well known PDE) in order that X is a martingale.

6.17 Proposition 6.4 also holds if we replace $\mathcal{P}$ by the larger class of processes a that are continuous on finitely many intervals (s_{i-1}, s_i) with existing left and right limits at the endpoints that may also be $\pm\infty$ and $\mathbb{E}\int_0^T a(s)^2 \,\mathrm{d}s < \infty$. Show this.

6.18 Show that $\mathbf{1}_{E^n} \stackrel{\mathbb{P}-\text{a.s.}}{\to} \mathbf{1}_E$ along a subsequence in the proof of Theorem 6.19.

6.19 Consider Example 6.17.

(a) Verify that Equation (6.19) is true.

(b) There is another way to arrive at (6.19) for a more general situation. Let X be a semimartingale and $\langle X \rangle$ its quadratic variation process. Consider the product $Y(t) = \exp(X(t)) \exp(-\frac{1}{2}\langle X \rangle_t)$. Use the Itô product formula to arrive at (6.19).

6.20 Let a be a bounded adapted piecewise continuous process and Z as in Equation (6.25) for $t \leq T$. Show that $\mathbb{E}Z(T)^2 < \infty$. *Hint: write* $Z(T)^2 = \mathcal{E}(U)_T \exp(\int_0^T a(s)^2 \,\mathrm{d}s)$ *for some martingale* U *and* $\mathcal{E}(U)$ *its Doléans exponential.*

6.21 Here you prove a version of Abel's summation formula. Let $(a_k)_{k\geq 0}$ and $(b_k)_{k\geq 0}$ be two sequences of real numbers. Then for all positive integers t one has

$$a_t b_t - a_0 b_0 = \sum_{k=1}^{t} a_{k-1}\Delta b_k + \sum_{k=1}^{t} b_{k-1}\Delta a_k + \langle a, b\rangle_t,$$

where $\langle a, b\rangle_t = \sum_{k=1}^t \Delta a_k \Delta b_k$. [Recognize the similarity with (6.23) (you may want to think of partitions with vanishing mesh and of the definition of quadratic covariation), and with (6.14) for $f(x) = x^2$.]

6.22 Let X be a semimartingale with decomposition (6.10). Define $Y(t) = \exp(X(t) - \frac{1}{2}\langle X \rangle_t)$, $t \geq 0$. Show that Y satisfies (6.19) and that $Y = Y(0)\mathcal{E}(X)$.

7 Applications of stochastic integrals in finance

The purpose of this section is to show how we can use the theory of stochastic integration to say more on the models that we used in Section 5 to describe the evolution of the stock price. Furthermore we discuss again self-financing portfolios. We will present here another definition of a self-financing portfolio, which encompasses the definition in Section 5.2 as a special case. We will also return to hedging, and show that is possible to hedge certain composite claims, namely those whose pay-off depends on $S(T)$ and on the integral of S from the initial time to the time of maturity T.

7.1 The models

Consider the stock price process S and the representation under the equivalent martingale measure $\mathbb{Q}$ as given by Equation (5.1). Let $f(t,x) = s\exp(\sigma x + (r - \frac{1}{2}\sigma^2)t)$. Then $S(t) = f(t, W^{\mathbb{Q}}(t))$. Using Lemma 6.11, we obtain the following stochastic differential representation of S, of which we have seen a discrete time analogue in Equation (1.32),

$$\mathrm{d}S(t) = rS(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W^{\mathbb{Q}}(t). \tag{7.1}$$

We could do the same for the representation of S under the 'physical' probability measure $\mathbb{P}$ (the one that is supposed to model the 'true' dynamics of the stock price) with some $a \in \mathcal{P}$ (see Equation (5.3) with the process a there replaced by $a - \frac{1}{2}\sigma^2$) to obtain

$$\mathrm{d}S(t) = a(t)S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W^{\mathbb{P}}(t). \tag{7.2}$$

Equations (7.1) and (7.2) are examples of *stochastic differential equations*, this terminology should be clear.

The (somewhat artificial) measure $\mathbb{Q}$ will be used for pricing. If we let $\bar{S}(t)$ be the discounted price, $\bar{S}(t) = e^{-rt}S(t)$, we obtain from (7.1) by application of the (Itô) product rule

$$\bar{S}(t) = S(0) + \sigma \int_0^t \bar{S}(u)\,\mathrm{d}W^{\mathbb{Q}}(u). \tag{7.3}$$

Hence the process $\bar{S} = \{\bar{S}(t), t \in [0,T]\}$ is a martingale under $\mathbb{Q}$.

We use Girsanov's theorem 6.19 to see that we can deduce Equation (7.2) from Equation (7.1) and vice versa. In order to do so we have to find the Radon-Nikodym derivatives $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$ and $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$ (as always, $\mathbb{Q}$ and $\mathbb{P}$ restricted to $\mathcal{F}_T$). Indeed, with

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp(-\int_0^T \frac{a(t)-r}{\sigma}\,\mathrm{d}W^{\mathbb{P}}(t) - \frac{1}{2}\int_0^T (\frac{r-a(t)}{\sigma})^2\,\mathrm{d}t), \tag{7.4}$$

and

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} = \exp(-\int_0^T \frac{r-a(t)}{\sigma}\,\mathrm{d}W^{\mathbb{Q}}(t) - \frac{1}{2}\int_0^T (\frac{a(t)-r}{\sigma})^2\,\mathrm{d}t), \tag{7.5}$$

we have the relation $W^{\mathbb{P}}(t) = W^{\mathbb{Q}}(t) - \int_0^t \frac{a(s)-r}{\sigma}\,\mathrm{d}s$. Of course we take a such that the integrals in (7.4) and in (7.5) are well defined. Note that the probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent on $\mathcal{F}_T$. In Corollary 4.3 we have seen this for constant functions a. Note that a consequence of Girsanov's theorem is that the martingale parts in the equations for S, the terms with the two Brownian motions have the same coefficients, whereas the coefficients of the parts with $\mathrm{d}t$ differ.

Remark 7.1 The quantity $\frac{a(t)-r}{\sigma}$ in the expression for $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ is known as the *market price of risk*, a standardized excess rate of return over the risk-free rate r. We will see an analog of this in the next chapter for stochastic interest rate models.

In Section 1.2 we also considered the cumulative return process. We found in Proposition 1.10 a limit process R that was such that under the measure $\mathbb{Q}$ all increments $R(t) - R(s)$ have a normal $N(r(t-s), \sigma^2(t-s))$ distribution for $t > s$. This suggest to model R as a Brownian motion with drift, so

$$R(t) = R(0) + rt + \sigma W^{\mathbb{Q}}(t).$$

Likewise we can model R under the probability measure $\mathbb{P}$ as

$$R(t) = R(0) + \int_0^t a(s)\,\mathrm{d}s + \sigma W^{\mathbb{P}}(t).$$

Note that the Radon-Nikodym derivatives in (7.4) and in (7.5) transform the two representation for R in each other. Compare the equations for S and R to obtain both under the measures $\mathbb{P}$ and $\mathbb{Q}$ (like Equation (1.5) in discrete time)

$$\mathrm{d}S(t) = S(t)\,\mathrm{d}R(t). \tag{7.6}$$

Hence we can express S in terms of R using the Doléans exponential by the relation $S = S(0)\mathcal{E}(R)$, both under $\mathbb{P}$ and $\mathbb{Q}$. Note the similarity with the discrete time expression (1.7).

Recall that we took for the bond price process $B(t) = e^{rt}$, so

$$\mathrm{d}B(t) = rB(t)\,\mathrm{d}t. \tag{7.7}$$

We observe the similarity of the $\mathrm{d}t$ terms in Equations (7.1) and (7.7). Calling $a(t)$ in Equation (7.2) the rate of return of S under $\mathbb{P}$ and similarly r the rate of return of S under $\mathbb{Q}$, then we see that the rate of return of S under $\mathbb{Q}$ is just equal to the interest rate, the rate of return of the riskless asset, the bond. So we can characterize $\mathbb{Q}$ as the probability measure that gives the stock the same rate of return as the riskless asset. Because of this, $\mathbb{Q}$ is also called the risk neutral probability measure. The alternative name equivalent martingale measure is explained by noting that $\mathbb{Q}$ and $\mathbb{P}$ have been seen to be equivalent probability measures and that the discounted price process $\bar{S}$ is a martingale under $\mathbb{Q}$.

7.2 Self-financing portfolios and hedging

Consider a portfolio process (x_t, y_t) with value process V, so $V(t) = x_tS(t) + y_tB(t)$. Let us assume that V is a continuous semimartingale under the probability measure $\mathbb{P}$, so that we can speak of the stochastic differential of V. This allows us to formulate the definition of self-financing portfolios in a way that is the natural counterpart in continuous time of Equations (1.2) and (1.3), unlike the somewhat artificial Definition 5.1.

Definition 7.2 A portfolio process is called *self-financing* if it satisfies the integrability condition $\mathbb{E}_{\mathbb{Q}}\int_0^T x(u)^2 S(u)^2 \,\mathrm{d}u < \infty$ and if we have for all $t \in [0,T]$ $\mathbb{P}$-almost surely (and hence $\mathbb{Q}$-almost surely)

$$V(t) = V(0) + \int_0^t x_u \,\mathrm{d}S(u) + \int_0^t y_u \,\mathrm{d}B(u).$$

Like in discrete time, it is possible to show that this definition can equivalently be stated in terms of discounted processes. Let $\bar{V}(t) = e^{-rt}V(t)$, $t \in [0,T]$ and recall from Section 7.1 $\bar{S}(t) = e^{-rt}S(t)$.

Proposition 7.3 *A portfolio process is self-financing iff we have for all $t \in [0,T]$ almost surely $\bar{V}(t) = V(0) + \int_0^t x_u \,\mathrm{d}\bar{S}(u)$. A self-financing portfolio process is therefore such that the process $\bar{V}$ is a martingale under $\mathbb{Q}$.*

Proof Exercise 7.1. □

Remark 7.4 Note that for Definition 7.2 and Proposition 7.3 we need the stochastic integrals $\int_0^t x_u \,\mathrm{d}S(u)$ and $\int_0^t x_u \,\mathrm{d}\bar{S}(u)$. Compare this to Definition 5.1 where we needed a special form of the value process V, because stochastic integrals had not been introduced yet. Proposition 7.3 also tells us that the discounted value process of a self-financing portfolio is a martingale under $\mathbb{Q}$ and $\mathbb{E}_{\mathbb{Q}}\bar{V}(T)^2 < \infty$.

Let us turn to *arbitrage portfolios*. By definition these are self-financing and such that the corresponding value process V satisfies $\mathbb{P}(V(0) = 0) = 1$, $\mathbb{P}(V(T) \geq 0) = 1$ and $\mathbb{P}(V(T) > 0) > 0$. Clearly, we can equivalently rephrase the latter three conditions in terms of the discounted valued process $\bar{V}$, $\mathbb{P}(\bar{V}(0) = 0) = 1$, $\mathbb{P}(\bar{V}(T) \geq 0) = 1$ and $\mathbb{P}(\bar{V}(T) > 0) > 0$. As before, a market is arbitrage free, if no arbitrage portfolios exist. We have the following corollary to Proposition 7.3, whose proof now relies on Definition 7.2. This corollary we have already encountered in Section 5 as Proposition 5.2, albeit in a different setting since stochastic integrals had not yet been defined. We repeat it, for convenience, slightly differently formulated.

Corollary 7.5 *Consider the market as described by Equation* (7.2)*. Assume that there exists an equivalent martingale measure $\mathbb{Q}$. Then this market is arbitrage free.*

Proof Consider a self-financing portfolio. According to Proposition 7.3, its discounted value process is a martingale under $\mathbb{Q}$ and hence $\mathbb{E}_{\mathbb{Q}}\bar{V}(T) = \mathbb{E}_{\mathbb{Q}}\bar{V}(0)$. If the portfolio would be an arbitrage portfolio, then we would have by equivalence of $\mathbb{P}$ and $\mathbb{Q}$ that $\mathbb{Q}(\bar{V}(0) = 0) = 1$. But then we would have $\mathbb{E}_{\mathbb{Q}}\bar{V}(0) = 0$ and then also $\mathbb{E}_{\mathbb{Q}}\bar{V}(T) = 0$, which is impossible since we also have $\mathbb{Q}(\bar{V}(T) \geq 0) = 1$ and $\mathbb{Q}(\bar{V}(T) > 0) > 0$ as these two imply $\mathbb{E}_{\mathbb{Q}}\bar{V}(T) > 0$. □

The next proposition says that any process that is a martingale can be seen as the value process of a self-financing portfolio in the sense of Definition 7.2, if we restrict the class of martingales under consideration.

Proposition 7.6 *Let a process $\bar{V}$ be a martingale under $\mathbb{Q}$ and suppose that there exists a function $\bar{v} \in C^{1,2}$ such that $\bar{V}(t) = \bar{v}(t, \bar{S}(t))$. Then the portfolio* *defined by* $x_t = \bar v_x(t,\bar S(t))$ *and* $y_t = \bar V(t) - x_t\bar S(t)$ *is self-financing and its discounted value process is just* $\bar V$. *Moreover, for all* $t < T$ *and* $x > 0$ *the function* $\bar v$ *satisfies*

$$\bar v_t(t,x) + \frac{1}{2}\sigma^2 x^2 \bar v_{xx}(t,x) = 0. \tag{7.8}$$

Proof We exploit again the martingale property of $\bar V$. Apply Theorem 6.13 to $\bar V$ and use Equation (7.3). We obtain

$$\begin{aligned}\mathrm{d}\bar V(t) &= \bar v_t(t,\bar S(t))\,\mathrm{d}t + \bar v_x(t,\bar S(t))\,\mathrm{d}\bar S(t) + \frac{1}{2}\bar v_{xx}(t,\bar S(t))\,\mathrm{d}\langle \bar S\rangle_t\\ &= \bar v_t(t,\bar S(t))\mathrm{d}t + \bar v_x(t,\bar S(t))\sigma\bar S(t)\,\mathrm{d}W^{\mathbb{Q}}(t) + \frac{1}{2}\bar v_{xx}(t,\bar S(t))\sigma^2\bar S(t)^2\mathrm{d}t.\end{aligned}$$

In this equation the collected terms ending with $\mathrm{d}t$ vanish, because $\bar V$ is a martingale. Hence (7.8) must be satisfied for positive x (all values of $S(t)$ are positive) and we are left with $\mathrm{d}\bar V(t) = \bar v_x(t,\bar S(t))\,\mathrm{d}\bar S(t)$. So, choosing x_t as $\bar v_x(t,\bar S(t))$ and y_t as $y_t = \bar V(t) - x_t\bar S(t)$, we get that $\bar V$ is the discounted value process associated with this portfolio and that the portfolio process is self financing in view of Proposition 7.3. □

Proposition 7.7 *Let* $\bar v \in C^{1,2}$, *assume that it satisfies* (7.8) *and put* $\bar v(T,x) =: \bar F(x)$, *where* $\bar F$ *is such that* $\mathbb{E}_{\mathbb{Q}}|\bar F(\bar S(T))| < \infty$. *Then* $\bar V(t) := \bar v(t,\bar S(t))$ *defines a martingale under* $\mathbb{Q}$ *and the self-financing portfolio of Proposition 7.6 hedges the discounted claim* $\bar F(\bar S(T))$.

Proof Looking at the proof of Proposition 7.6, one sees that $\bar V$ is a martingale. Apply the assertion of that proposition and use that $\bar V(T) = \bar F(\bar S(T))$ to conclude. □

Proposition 7.6 has as an immediate consequence that every simple claim can be hedged.

Corollary 7.8 *Every simple claim* $X = F(S(T))$ *with* $\mathbb{E}_{\mathbb{Q}}|X| < \infty$ *can be hedged by a self-financing portfolio.*

Proof Let $\bar X = e^{-rT}X$. By the Markov property of $\bar S$ (under $\mathbb{Q}$), there exists a function $\bar v$ such that the martingale $\mathbb{E}_{\mathbb{Q}}[\bar X|\mathcal{F}_t]$ can be written as $\bar v(t,\bar S(t))$. Since $\bar v(T,\bar S(T)) = \bar X$ the proof is completed by invoking Proposition 7.6. □

The results above are in terms of discounted values, which we shall now transform into undiscounted values. The partial differential equation (7.8) we already encountered as (3.11) in Exercise 3.6, with v playing the role of $\bar v$. Keeping the notation of the present section, we define $v(t,x) := e^{rt}\bar v(t,e^{-rt}x)$. Then v satisfies the partial differential equation (3.12) (replace w there with v). So, in the notation of the present section we again encounter the Black-Scholes partial differential equation in the form of (5.9),

$$v_t(t,x) + \frac{1}{2}\sigma^2x^2v_{xx}(t,x) + rxv_x(t,x) - rv(t,x) = 0. \tag{7.9}$$

With $V(t) := v(t,S(t))$ we have the value at time t of the self-financing portfolio that hedges the claim $F(S(T)) := e^{rT}\bar F(e^{-rT}S(T))$, where $\bar F$ is as in Proposition 7.7 and $F(x) := e^{rT}\bar F(e^{-rT}x)$. Hence we endow (7.9) with the terminal condition $v(T,x) = F(x)$. Moreover the x_t of the hedge portfolio of Proposition 7.6 also satisfies $x_t = v_x(t, S(t))$, a property we already encountered in Section 5.2.

The Itô calculus can also be used to find hedging strategies for certain composite claims. These claims are of the following special structure. They depend on $S(T)$ and a certain integral of transformations of S. An example of such a claim is the Asian call option with maturity time T and strike price K, which is the claim defined as the one whose pay-off is $(\frac{1}{T}\int_0^T S(t)\,\mathrm{d}t - K)^+$.

Let $U(t) = \int_0^t g(v, S(v))\,\mathrm{d}v$ with some given function $g : \mathbb{R}^2 \to \mathbb{R}$. Assume that $U(t)$ is well defined for all t and consider the claim $F(S(T), U(T))$. Suppose that it is possible to hedge this claim (with a self-financing portfolio). In that case the discounted value of the portfolio is equal to the discounted price process of the claim and so it must be a martingale under $\mathbb{Q}$. We thus have $\bar{V}(t) = \mathbb{E}_{\mathbb{Q}}[\bar{F}(S(T), U(T))|\mathcal{F}_t]$. It would be nice if we could write $\bar{V}(t) = e^{-rt}V(t)$ with $V(t) = v(t, S(t), U(t))$ for some function v, and indeed this is what happens. To understand this we argue as follows, just like we did for simple claims. Consider the computation of the conditional expectation of $F\big(S(T), \int_0^T g(v, S(v))\,\mathrm{d}v\big)$ given $\mathcal{F}_t$,

$$\mathbb{E}_{\mathbb{Q}}[F(S(T), \int_0^T g(v, S(v))\,\mathrm{d}v\big)|\mathcal{F}_t]. \tag{7.10}$$

We decompose as follows

$$S(T) = \frac{S(T)}{S(t)}S(t),$$

$$\int_0^T g(v, S(v))\,\mathrm{d}v = \int_0^t g(v, S(v))\,\mathrm{d}v + \int_t^T g(v, S(v))\,\mathrm{d}v.$$

Given $\mathcal{F}_t$ we completely know in the latter expression the first summand, which is just $U(t)$. The second summand we write as $\int_t^T g(v, \frac{S(v)}{S(t)}S(t))\,\mathrm{d}v$. Now we exploit the independence under $\mathbb{Q}$ of $\frac{S(v)}{S(t)}$ and the whole past of S up to t (a consequence of the independent increment property of a Brownian motion) and apply Proposition A.12 (v). Carrying out the conditional expectation in (7.10) with values $S(t) = s$ and $U(t) = u$ then leaves us with the unconditional expectation

$$\mathbb{E}_{\mathbb{Q}}F\big(\frac{S(T)}{S(t)}s, u + \int_t^T g(v, \frac{S(v)}{S(t)}s\big)\,\mathrm{d}v) =: v(t, s, u),$$

an expression in terms of s and u, just what we desired. Can we characterize this expression? Yes, like for simple claims we have that also this function v is the solution of a partial differential equation with a boundary condition. We now state the result.

Theorem 7.9 *Let v, a sufficiently smooth function of t, x, u (here $t \in [0,T]$, $x \geq 0$, $u \in \mathbb{R}$) be the solution of the partial differential equation*

$$v_t(t,x,u) + xrv_x(t,x,u)$$

$$+\frac{1}{2}\sigma^2 x^2 v_{xx}(t,x,u) + g(t,x)v_u(t,x,u) - rv(t,x,u) = 0, \tag{7.11}$$

with boundary condition $v(T,x,u) = F(x,u)$. *The portfolio consisting of* $x_t = v_x(t,S(t),U(t))$ *and* $y_t = (v(t,S(t),U(t)) - x_t S(t))e^{-rt}$ *is self-financing and hedges the claim* $F(S(T),U(T))$. *The fair price of the claim at time* t *is* $V(t) = v(t,S(t),U(t))$.

Proof We apply the multi-dimensional version of the Itô formula to the process $V(t) = v(t,S(t),U(t))$. We get, omitting the arguments $t, S(t)$ and $U(t)$ in v and g, in shorthand notation

$$\begin{aligned}
\mathrm{d}V(t) &= v_t\,\mathrm{d}t + v_x\,\mathrm{d}S(t) + v_u\,\mathrm{d}U(t) + \frac{1}{2}v_{xx}\,\mathrm{d}\langle S\rangle_t \\
&= v_t\,\mathrm{d}t + v_x\,rS(t)\,\mathrm{d}t + v_x\,\sigma S(t)\,\mathrm{d}W^{\mathbb{Q}}(t) + v_u\,g\,\mathrm{d}t + \frac{1}{2}v_{xx}\,\sigma^2 S(t)^2\,\mathrm{d}t \\
&= rV(t)\,\mathrm{d}t + v_x\,\sigma S(t)\,\mathrm{d}W^{\mathbb{Q}}(t),
\end{aligned}$$

where we used in the last equality that v was a solution of (7.11). Hence the discounted value process $\bar{V}$, recall that $\bar{V}(t) = e^{-rt}V(t)$, satisfies

$$\mathrm{d}\bar{V}(t) = -r\bar{V}(t)\,\mathrm{d}t + e^{-rt}\mathrm{d}V(t) \tag{7.12}$$

$$= e^{-rt}v_x\,\sigma S(t)\,\mathrm{d}W^{\mathbb{Q}}(t) \tag{7.13}$$

$$= v_x\,\mathrm{d}\bar{S}(t). \tag{7.14}$$

So if we define x and y as in the assertion of the theorem, we get $\bar{V}(t) = x_t\bar{S}(t)+y_t$, the discounted value process of the portfolio, and $\mathrm{d}\bar{V}(t) = x_t\mathrm{d}\bar{S}(t)$, so that the portfolio is self-financing. From the boundary condition on v we obtain $v(T,S(T),U(T)) = F(S(T),U(T))$, this portfolio is thus a hedging portfolio for the claim $F(S(T),U(T))$. □

As a corollary to Theorem 7.9 we get

Corollary 7.10 *Every claim* $F(S(T),U(T))$ *that satisfies an appropriate integrability condition can be hedged. The hedging portfolio is given in Theorem 7.9.*

Proof Similar to that of Corollary 7.8, using Theorem 7.9. □

The above results extends previous results, and, not surprisingly, collapse to those results when $g = 0$, for example (7.11) reduces to (7.9).

7.3 More general claims

We have seen in Theorem 7.9 that we can hedge certain composite claims, but they had a special structure. Recall from Section 1 that we called a market complete if every claim can be hedged. We ask ourselves whether the Black-Scholes market is complete, and if it is, how for a given claim the hedge portfolio would look like. The answer is affirmative, although one has to put some mild technical restrictions on the claim under consideration. Recall that $W = W^{\mathbb{P}}$ is a Brownian motion under $\mathbb{P}$. A most effective one is to demand that a claim X

has finite second moment under $\mathbb{Q}$. Under that condition, as we shall see below, any $\mathcal{F}_T = \mathcal{F}_T^W$-measurable claim, not only simple ones, can be hedged.

This fact is a consequence of the *martingale representation theorem* (MRT), a non-trivial result in probability theory, that we present without proof, as that would need results that are beyond what we treat in these notes.

Theorem 7.11 *Let M be a martingale (under $\mathbb{P}$!) that is adapted to the Brownian filtration $\mathbb{F}^W = \{\mathcal{F}_t^W, t \in [0,T]\}$, where $\mathcal{F}_t^W = \sigma(W_s, s \leq t)$, and that satisfies $\mathbb{E}M(T)^2 < \infty$. Then there exists an adapted process ϕ satisfying $\mathbb{E}\int_0^T \phi_u^2 \,\mathrm{d}u$ such that*

$$M(t) = M(0) + \int_0^t \phi_u \,\mathrm{d}W(u), \quad t \in [0,T]. \tag{7.15}$$

As we have seen $\tilde{M}(t) := \mathbb{E}_{\mathbb{Q}}[e^{-rT}X|\mathcal{F}_t]$ gives a martingale under $\mathbb{Q}$. We'd like to apply the MRT, but the first problem is that here we don't have a martingale under $\mathbb{P}$, but under $\mathbb{Q}$. This could be circumvented by a change of notation (replace $\mathbb{P}$ with $\mathbb{Q}$), but then the second problem shows up: the $\mathcal{F}_t$ are generated by W, and although $W^{\mathbb{Q}}$ is adapted to the $\mathcal{F}_t$, it is in general not true that $\mathcal{F}_t$ equals $\sigma(W^{\mathbb{Q}}(s), s \leq t)$, but only $\mathcal{F}_t \supset \sigma(W^{\mathbb{Q}}(s), s \leq t)$. Yet, the two filtrations coincide, and the problem disappears, in the important case where $W(t)$ and $W^{\mathbb{Q}}(t)$ differ by a nonrandom quantity, like in (4.7). See Exercise 7.9 for a way out in the general case. This leads to

Proposition 7.12 *Let X be an $\mathcal{F}_T = \mathcal{F}_T^W$-measurable claim with $\mathbb{E}_{\mathbb{Q}}X^2 < \infty$. Then there exists an adapted process $\bar{\phi}$ such that*

$$\bar{V}(t) := \mathbb{E}_{\mathbb{Q}}[e^{-rT}X|\mathcal{F}_t] = \bar{V}(0) + \int_0^t \bar{\phi}_u \,\mathrm{d}\bar{S}(u), \quad t \in [0,T].$$

Moreover, choosing $x_t = \bar{\phi}_t$ and $y_t = \bar{V}(t) - x_t\bar{S}_t$ yields a self-financing hedging strategy of X and $\bar{V}$ is the discounted value process of the claim X. Hence the portfolio process (x_t, y_t) hedges the claim X.

Proof Apply the result of Exercise 7.9 to get an adapted process $\tilde{\phi}$ such that

$$\bar{V}(t) = \bar{V}(0) + \int_0^t \tilde{\phi}_u \,\mathrm{d}W^{\mathbb{Q}}(u)$$

Recall that $\mathrm{d}\bar{S}(t) = \sigma\bar{S}(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$, so that one can write, as $\sigma, \bar{S(u)} > 0$,

$$\bar{V}(t) = \bar{V}(0) + \int_0^t \frac{\tilde{\phi}_u}{\sigma\bar{S}(u)} \,\mathrm{d}\bar{S}(u).$$

Choose then $\bar{\phi}_t = \frac{\tilde{\phi}_t}{\sigma\bar{S}(t)}$. □

In general, the process $\tilde{\phi}$ (like ϕ) is hard to characterize. There is usually no partial differential equation that has the value process of the portfolio as a solution. There is, however a theoretical description of x_t in terms of what is called a Malliavin derivative. This again touches upon a very advanced mathematical theory. However in discrete time we also encountered this derivative, when we described for the CRR market x_n as $\frac{DV_n}{DS_n}$, see (5.6), which also had an interpretation as some sort of derivative. In the continuous time case, as we said, the situation is much more complex, although it is indeed possible to give a meaning to the formal expression $x_t = \frac{DV}{DS}(t)$.

7.4 American style claims

In this section we briefly treat one example of an American claim, the American call option with maturity T and strike price K. The pay-off of this claim at a by the investor chosen time $\tau \leq T$ is $(S(\tau) - K)^+$. The (random) time τ may assume any value between 0 and T and may depend on the random evolution of the stock price, τ thus becomes a random variable, but has to be chosen such that only past information of the stock price is used. So, we require that for any $t \leq T$ the event $\{\tau \leq t\}$ depends on the values of S at times $s \leq t$ only. In technical terms, given a filtration, we require for all $t \in [0, T]$ that the event $\{\tau \leq t\}$ is $\mathcal{F}_t$-measurable, and when this is true, τ is called a *stopping time*. Note that any deterministic τ is a stopping time, because then, for any t, the event $\{\tau \leq t\}$ is either $\emptyset$ or Ω; both sets belong to any σ-algebra.

What is 'American' in this case is that the owner of the claim doesn't have to wait to maturity to exercise it, but may do it at any time before it, when it seems profitable. The problem here becomes to find a strategy that tells an investor when to exercise the claim. For an arbitrary claim X (not only a call option), technically speaking, one wants to maximize $\mathbb{E}_{\mathbb{Q}} e^{-r\tau} X$, which is by definition the value at time $t = 0$ of the claim when exercised at τ, where τ runs through the set of stopping times bounded by T. So the problem is finding $\sup_\tau \mathbb{E}_{\mathbb{Q}} e^{-r\tau} X$, and a maximizer τ^* (suppose it exists) is then called the optimal stopping time, or optimal exercise time. This is an example of an *optimal stopping problem*.

Problems of this type are hard to solve. But in the *special case* of a call option we can, and the answer is at first glance rather surprising: the optimal strategy is to wait until maturity! As a result, the value of the American claim is thus the same as that of its European version with same maturity time and same strike price, although the owner of the European claim has no freedom to choose himself an appropriate moment to exercise it. Let us see why this equivalence takes place.

Let us denote the price of the American option at t by $C_A(t)$ and the price of the corresponding European option by $C_E(t)$. Clearly one should have $C_A(t) \geq C_E(t)$ for all t. For simplicity we show this for $t = 0$. The value of the American claim is $C_A(0) = \sup_\tau \mathbb{E}_{\mathbb{Q}} e^{-r\tau} X$, which is greater than $C_E(0) = \mathbb{E}_{\mathbb{Q}} e^{-rT} X$, as the deterministic time T is also a stopping time. Hence $C_A(t) \geq C_E(t)$ for $t = 0$.

Consider a market with interest rate $r > 0$. We compare two financial products traded at time t. The first one is a portfolio that consists of 1 share of the stock and a loan with value K from the bank that has to be paid back at time T. At time t the loan thus has value $e^{-r(T-t)}K$. The total value of this portfolio at time t is thus $S(t) - e^{-r(T-t)}K$. The other product is one European call option with payoff $(S(T) - K)^+$, whose value at t is thus $C_E(t)$. We compare the values of the two products at time T. They are $S(T) - K$ and $C_E(T) = (S(T) - K)^+$, respectively and we have the obvious inequality $S(T) - K \leq C_E(T)$. But then this inequality between the two values is preserved at any time $t < T$ and we get $S(t) - Ke^{-r(T-t)} \leq C_E(t)$. Hence we also have the strict inequality $S(t) - K < C_E(t) \leq C_A(t)$. Now suppose the owner of an American call wants to exercise his option at the chosen time $\tau = t < T$, he will only do this when $S(t) > K$, and will then be paid $S(t) - K$ and thus gets less than the value of the American call. This means that it is not profitable for him to exercise the option before maturity. So the optimal time to exercise the American claim is $\tau = T$ and hence the value $C_A(t)$ of the American claim is equal to the value $C_E(t)$ of the European claim, $C_A(t) = C_E(t)$.

This surprising result has no counterpart for the American *put option*. It is possible to show that the value of an American put is strictly larger than the value of the European put. It is interesting to see why an argument similar to the one we used for a call option breaks down in this case (Exercise 7.10).

7.5 The Greeks

Starting point of this section is the Black-Scholes formula (1.30) for the price of a European call option. Clearly, it depends on a number of parameters. We shall investigate the sensitivity of the price w.r.t. these parameters. By this we mean that we shall quantify how small changes in the parameter result in the price of a European call. Not surprisingly, we can express these in terms of partial derivatives and these are known as the 'Greeks'. We write $C = C(t, s, r, \sigma)$ (here s is the value of $S(t)$) for the expression (1.30) and abbreviate the $d_i(t, x)$ by d_i. Note that these two quantities depend on t, s, r, σ too. Now we introduce the following *sensitivity parameters*, or *sensitivity measures*.

$$
\begin{aligned}
\Delta &= \frac{\partial C}{\partial s} \; (\textit{delta}),\\
\Gamma &= \frac{\partial^2 C}{\partial s^2} \; (\textit{gamma}),\\
\rho &= \frac{\partial C}{\partial r} \; (\textit{rho}),\\
\Theta &= \frac{\partial C}{\partial t} \; (\textit{theta}),\\
\mathcal{V} &= \frac{\partial C}{\partial \sigma} \; (\textit{vega}).
\end{aligned}
$$

Of course, the 'Greeks' in turn depend on all parameters as well, see Proposition 7.13, but for reasons of clarity we suppress this in the notation. In principle it is possible to define similar sensitivity measures for prices of other derivatives as well. Below we confine ourselves to formulas for call options, because we can give explicit expressions for them. The result is the following proposition.

Proposition 7.13 *Let ϕ be the density of the standard normal distribution and Φ its distribution function. The following hold.*

$$
\begin{aligned}
\Delta &= \Phi(d_1)\\
\Gamma &= \frac{\phi(d_1)}{s\sigma\sqrt{T-t}}\\
\rho &= K(T-t)e^{-r(T-t)}\Phi(d_2)\\
\Theta &= -\frac{s\phi(d_1)\sigma}{2\sqrt{T-t}} - rKe^{-r(T-t)}\Phi(d_2)\\
\mathcal{V} &= s\phi(d_1)\sqrt{T-t}.
\end{aligned}
$$

Proof Exercise 7.11. □

For a number of reasons it may be attractive to manage portfolios that are not sensitive, also called *neutral*, to one or more quantities like stock price or volatility. The corresponding partial derivatives then have to be (close to) zero. In general, for a given portfolio this will not be the case of course, but sometimes an extension of the portfolio with an additional derivative may accomplish it. Suppose we have a portfolio with value V (depending on some relevant quantities), not necessarily a hedge portfolio of a certain claim, and we wish to make it Δ-neutral by addition of z units of a product having value D. The resulting new value is delta-neutral if

$$0 = \frac{\partial V}{\partial s} + z\frac{\partial D}{\partial s}.$$

It follows that we should choose $z = -\frac{\Delta_V}{\Delta_D}$, in self-evident notation. In practice, adopting this as a basis for a trading strategy, is not always attractive, since one has to re-balance the portfolio at every time to keep it Δ-neutral. Moreover, if the delta of the portfolio itself is subject to big changes (a big value for Γ, Γ is always positive), the result at any time will be major adjustments of the portfolio (z depends on t). This may lead to high transactions costs (which we have completely ignored before) in practice. In such a situation, one may extend the portfolio with yet another product in order to accomplish Γ-neutrality. It turns out a good idea to take the underlying stock as this extra product. Supposing we buy an extra w shares, we require the two following equations to hold.

$$\begin{aligned} \frac{\partial V}{\partial s} + z\frac{\partial D}{\partial s} + w\frac{\partial s}{\partial s} &= 0 \\ \frac{\partial^2 V}{\partial s^2} + z\frac{\partial^2 D}{\partial s^2} + w\frac{\partial^2 s}{\partial s^2} &= 0. \end{aligned}$$

Solving this system of equations (using that $\frac{\partial s}{\partial s} = 1$ and $\frac{\partial^2 s}{\partial s^2} = 0$) results in (again in self-evident notation)

$$\begin{aligned} z &= -\frac{\Gamma_V}{\Gamma_D} \\ w &= \frac{\Gamma_V}{\Gamma_D}\Delta_D - \Delta_V. \end{aligned}$$

And the story doesn't end here ...

7.6 Exercises

7.1 Prove Proposition 7.3.

7.2 Consider the Black-Scholes market with claim $\int_0^T S(u)\,\mathrm{d}u$.

(a) Show that the fair price of this claim at time t is given by $V(t)$ with $V(t) = e^{r(t-T)}U(t) + S(t)\frac{1-e^{r(t-T)}}{r}$, where $U(t) = \int_0^t S(u)\,\mathrm{d}u$ and $r \neq 0$ the interest rate. What is $V(t)$ if $r = 0$?

(b) Write $V(t)$ as $v(t, S(t), U(t))$ and determine a partial differential equation for v.

(c) Find the hedge strategy for this claim.

7.3 Consider in a Black-Scholes market a European call and a European put option. Denote by C_t the price at time $t \leq T$ of the call and by P_t the price of the put. Derive the *put-call parity* $C_t - P_t = S_t - e^{r(t-T)}K$.

7.4 Consider the price of a European call option in the Black Scholes market. One has $V(t) = C(t, S(t))$ with C as in Equation (1.30). Show by direct application of the Itô formula that V satisfies $\mathrm{d}V(t) = rV(t)\,\mathrm{d}t + g(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$, where you also specify what $g(t)$ is.

7.5 Consider the Black-Scholes market. Let V be the value process of a simple claim. Show that there exists a process g such that V satisfies

$$\mathrm{d}V(t) = rV(t)\mathrm{d}t + g(t)\,\mathrm{d}W^{\mathbb{Q}}(t).$$

7.6 Consider in a Black-Scholes market the claim $F(S(T)) = K\mathbf{1}_{(a,b)}(S(T))$ (this is called a binary spread). Determine for each time $t \leq T$ the fair price of this claim. Give also a PDE with boundary condition that is satisfied by the price function (as a function of t and $S(t)$).

7.7 Consider a *straddle* in the Black-Scholes market. A straddle is a claim with pay-off at maturity equal to $X = |S(T) - K|$. Find the price of this claim at any time $t \leq T$. This claim can also be hedged with a *constant* portfolio that not only consists of shares and bonds, but contains as a third component European call options as well. Give this portfolio. Find also a PDE with boundary condition that is satisfied by the price function (as a function of t and $S(t)$).

7.8 Consider in the Black-Scholes market a *bull-spread*. This is a claim with pay-off $\min\{\max\{S(T), A\}, B\}$, where $B > A > 0$. Like a straddle (Exercise 7.7), also this claim can be hedged with a constant portfolio consisting of stocks, bonds and European call options. Find this portfolio and the price process of the bull-spread. Give also a PDE with boundary condition that is satisfied by the price function (as a function of t and $S(t)$).

7.9 The setting is as in Section 7.3. Let V be a martingale under $\mathbb{Q}$, and Z the density process. The latter means that $Z(t)$ can be written, a bit similar too (7.4), as $Z = \mathcal{E}(-\int_0^\cdot b(t)\,\mathrm{d}W^{\mathbb{P}}(t))$ for an appropriate process b. Note that $\mathrm{d}Z(t) = -Z(t)b(t)\,\mathrm{d}W^{\mathbb{P}}(t)$.

(a) Show that M defined by $M(t) = V(t)Z(t)$ is a martingale under $\mathbb{P}$. [Proposition 4.5 is useful here.]

(b) Apply the MRT to write $M(t) = V(0) + \int_0^t \phi_u\,\mathrm{d}W^{\mathbb{P}}(u)$, and show that one can also write $V(t) = V(0) + \int_0^t \tilde{\phi}_u\,\mathrm{d}W(u)$. To do that you need the Itô product rule (6.23) and rules for computing quadratic covariation. Express $\tilde{\phi}_u$ in terms of ϕ_u and $Z(u)$.

7.10 Show that an argument similar to the proof of the equivalence of American and European call options (comparing a simple portfolio with a European call option) breaks down if we apply it to American and European put options.

7.11 First show (ϕ denotes the density of the standard normal distribution)

$$s\phi(d_1) - Ke^{-r(T-t)}\phi(d_2) = 0,$$

and use this to derive the expressions for the Greeks in Proposition 7.13.

7.12 Compute the Greeks for the forward contract with payoff $S(T) - K$ in the Black Scholes model.

7.13 Compute the Greeks for a European put option (payoff $(K - S(T))^+$) in the Black Scholes model. You may use the results of Proposition 7.13.

7.14 Here you prove the converse of Theorem 7.9. Suppose (the context of the theorem applies) that a claim $F(S(T), Z(T))$ can be hedged by a self-financing portfolio with value process V if the type $V(t) = v(t, S(t), Z(t))$, where v is sufficiently differentiable. Show that v satisfies Equation (7.11).

7.15 Consider Proposition 7.6. Show that the function $\bar{v}$ satisfies the PDE (3.11) for $x > 0$ and $t \geq 0$. If $\bar{V}$ is the discounted price process of a (hedgeable) claim $F(S_T)$, what is the boundary condition for $\bar{v}(T, x)$?

7.16 The result of Equation (7.15) is not constructive, it is not told how to construct the process ϕ from the given Brownian martingale M. In the following cases you have to find an explicit expression for ϕ_u for $u \in [0, T]$.

(a) $M_t = W_t^3 - c \int_0^t W_s \, \mathrm{d}s$ for a suitable constant c (which one?).

(b) For some fixed time T we have $M_t = \mathbb{E}[e^{W_T} | \mathcal{F}_t]$.

(c) For some fixed time T we take $M_t = \mathbb{E}[\int_0^T W_s \, \mathrm{d}s | \mathcal{F}_t]$.

(d) If v is a solution to the *backward heat equation*

$$\frac{\partial v}{\partial t}(t, x) + \frac{1}{2} \frac{\partial v}{\partial x^2}(t, x) = 0,$$

then $M_t = v(t, W_t)$ is a martingale. Show this to be true under a to be specified integrability condition. Give also two examples of a martingale M that can be written in this form.

(e) Suppose a square integrable martingale M is of the form $M_t = v(t, W_t)$, where W is a standard Brownian motion and $v : \mathbb{R}^2 \to \mathbb{R}$ is continuously differentiable in the first variable and twice continuously differentiable in the second variable. Show that v satisfies the backward heat equation.

7.17 Consider the situation of Theorem 7.9. We consider the discounted fair price $\bar{V}(t)$ of the underlying claim. This price can be written as a function $\bar{v}$ of the underlying discounted processes, $\bar{V}(t) = \bar{v}(t, \bar{S}(t), \bar{U}(t))$. Give a partial differential equation that is satisfied by $\bar{v}$.

8 Interest rate models, a swift introduction

In this section we focus on stochastic interest rate models in continuous time. Until now we always assumed that the interest rate was a constant r, also called short rate. This assumption will be relaxed and we will study some consequences. We also present a set of different interest rate notions, of which the short rate is one example. The basic setting is that all processes are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, endowed with a filtration $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ to which all processes are adapted. Brownian motions W are then always adapted to $\mathbb{F}$ and it is always assumed that $W_{t+h} - W_t$ is independent of the past at time t, i.e. independent of $\mathcal{F}_t$, for all $t, h \geq 0$. Usually we will have a finite time horizon $T > 0$.

8.1 Some general theory

To start, we consider a *zero coupon bond* with maturity time $T > 0$. The value of this product at maturity is fixed at 1 (euro). For $t \in [0, T]$ we denote by $P(t, T)$ the (fair) price of this bond at time t. Note that $P(T, T) = 1$ for any $T > 0$. As before we denote by $B(t)$ the bank account at time t. The fair price of the bond at time t can then be computed just as before when considering the pricing of claims. Throughout, to rule out arbitrage opportunities, it is assumed that an equivalent martingale measure $\mathbb{Q}$ exists, in the sense that $P(t,T) = B(t)\mathbb{E}_{\mathbb{Q}}[\frac{1}{B(T)}|\mathcal{F}_t]$ for $t \leq T$ and all $T > 0$. This assumption is equivalent to $t \mapsto \frac{P(t,T)}{B(t)}$ being a $\mathbb{Q}$-martingale for for $t \leq T$ and all $T > 0$.

In the situation that we treated before, r is a constant, we simply get $P(t,T) = \exp(-r(T-t))$. This simple expression will radically change if we make r time dependent and random. As a consequence $\{P(t,T),\, t \in [0,T]\}$ will become an adapted random process. For fixed t the random mapping $T \mapsto P(t,T)$ is called the *term structure* of the bond, also called the *discount curve*. Typically this curve turns out to be a smooth function of T, whereas the dependence of $P(t,T)$ on t is that of a diffusion process.

The *continuously compounded short rate* for $[t, T]$ is

$$R(t,T) = -\frac{\log P(t,T)}{T-t}.$$

The function $T \mapsto R(t,T)$, for $T \geq t$, is also known as the *yield curve* at time t.

The *instantaneous forward rate* with maturity T prevailing at time t is defined as (the derivative is assumed to exist)

$$f(t,T) = -\frac{\partial}{\partial T} \log P(t,T), \tag{8.1}$$

where the derivative is assumed to exist. The function $T \mapsto f(t,T)$, for $T \geq t$, is called the *forward curve* at time t. Note that it holds that

$$P(t,T) = \exp(-\int_t^T f(t,s)\,\mathrm{d}s), \tag{8.2}$$

since $P(T,T) = 1$.

The instantaneous *short rate* at time t is defined as

$$r(t) := f(t,t) = \lim_{T \downarrow t} R(t,T) = -\frac{\partial}{\partial T} \log P(t,T)|_{T=t}.$$

This short rate is used to define the bank account $B(t)$ as the solution to the differential equation

$$\mathrm{d}B(t) = r(t)B(t)\,\mathrm{d}t,\ B(0) = 1.$$

Equivalently, one has

$$B(t) = \exp(\int_0^t r(u)\,\mathrm{d}u).$$

Note that the forward curve f determines the bond price, Equation (8.2) and the short rate r. But knowing the short rate alone is in general not sufficient to know the bond price. Nevertheless we will see later situations where this is indeed possible.

Note that for constant r one gets $R(t,T) = r$, $f(t,T) = r$ and $r(t) = r$. In this case all introduced rates coincide. Moreover one obtains in this case $B(t) = \exp(rt)$.

8.2 Short rate models and pricing

We will model the short rate as a diffusion process, so for some adapted processes b and σ we assume that

$$\mathrm{d}r(t) = b(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \tag{8.3}$$

where W is a Brownian motion under the physical measure $\mathbb{P}$, b and σ are certain stochastic processes that will be specified later on. Furthermore we assume that there exists an adapted process λ such that $Z(T) := \mathcal{E}(-\int_0^\cdot \lambda(t)\,\mathrm{d}W(t))_T$ has expectation one. Then $Z(T)$ determines a probability measure $\mathbb{Q}$ on $\mathcal{F}_T$, $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z(T)$. Moreover, we assume that $\mathbb{Q}$ is such that $\frac{P(t,T)}{B(t)}$ is a $\mathbb{Q}$-martingale, seen as a function of t. Consequently, under these assumptions we have, as an alternative to (8.2),

$$P(t,T) = \mathbb{E}_{\mathbb{Q}}[\frac{B(t)}{B(T)}|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[\exp(-\int_t^T r(u)\,\mathrm{d}u)|\mathcal{F}_t]. \tag{8.4}$$

By Girsanov's theorem 6.19 we can now also write a stochastic differential equation for r under $\mathbb{Q}$. With $W^{\mathbb{Q}}$ the $\mathbb{Q}$-Brownian motion defined by

$$W^{\mathbb{Q}}(t) = W(t) + \int_0^T \lambda(s)\,\mathrm{d}s, \tag{8.5}$$

one obtains from (8.3)

$$\mathrm{d}r(t) = (b(t) - \lambda(t)\sigma(t))\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W^{\mathbb{Q}}(t).$$

There are many choices possible for $\lambda(t)$ and hence as many choices for $\mathbb{Q}$, as long as $Z(T)$ has expectation one. This has everything to do with non-completeness of the market with the product $B(t)$ alone. It is a custom to specify a model for r under (some) $\mathbb{Q}$. This custom will be followed below, by directly specifying the coefficients appearing in Equation (8.7).

If we specify b, σ and λ further by assuming that $b(t)$, $\sigma(t)$ and $\lambda(t)$ are functions of t and $r(t)$, this SDE takes the form (with some ambiguity of notation)

$$\mathrm{d}r(t) = \big(b(t,r(t)) - \lambda(t,r(t))\sigma(t,r(t))\big)\,\mathrm{d}t + \sigma(t,r(t))\,\mathrm{d}W^{\mathbb{Q}}(t). \tag{8.6}$$

Introducing $b^{\mathbb{Q}}(t,r) = b(t,r) - \lambda(t,r)\sigma(t,r)$, we can rewrite (8.6) as

$$\mathrm{d}r(t) = b^{\mathbb{Q}}(t,r(t))\,\mathrm{d}t + \sigma(t,r(t))\,\mathrm{d}W^{\mathbb{Q}}(t). \tag{8.7}$$

Under mild assumptions it results that r is a Markov process under $\mathbb{Q}$ and consequently $P(t,T) = \mathbb{E}_{\mathbb{Q}}[\exp(-\int_t^T r(s)\,\mathrm{d}s|\mathcal{F}_t]$, as a conditional expectation of a functional of $r(s)$ with $t \le s \le T$ given the past becomes a function of t, $r(t)$ and T. To express this, we write

$$P(t,T) = F(t,r(t);T). \tag{8.8}$$

Let us apply the Itô formula to $F(t,r(t);T)$ for $t < T$ with T fixed, assuming that (8.7) holds and that F is sufficiently differentiable. We obtain, omitting all arguments,

$$\begin{aligned}\mathrm{d}F &= F_t\mathrm{d}t + F_r\mathrm{d}r + \frac{1}{2}F_{rr}\mathrm{d}\langle r\rangle \\ &= (F_t + F_r b^{\mathbb{Q}} + \frac{1}{2}F_{rr}\sigma^2)\mathrm{d}t + F_r\sigma\mathrm{d}W^{\mathbb{Q}}.\end{aligned}$$

Recall that $B(t) = \exp(\int_0^t r(s)\,\mathrm{d}s)$ and apply the product rule to $M(t) = \frac{1}{B(t)}P(t,T)$ to get

$$\begin{aligned}\mathrm{d}M &= -\frac{1}{B^2}rBF\,\mathrm{d}t + \frac{1}{B}\left((F_t + F_r b^{\mathbb{Q}} + \frac{1}{2}F_{rr}\sigma^2)\mathrm{d}t + F_r\sigma\mathrm{d}W^{\mathbb{Q}}\right) \\ &= \frac{1}{B}\left((-rF + F_t + F_r b^{\mathbb{Q}} + \frac{1}{2}F_{rr}\sigma^2)\,\mathrm{d}t + F_r\sigma\,\mathrm{d}W^{\mathbb{Q}}\right).\end{aligned}$$

It follows from (8.4) that $M(t) = \mathbb{E}_{\mathbb{Q}}[\frac{1}{B(T)}|\mathcal{F}_t]$, so M is a martingale under $\mathbb{Q}$. Hence the dt-terms in the above display vanish. In full, writing the arguments again, one thus obtains

$$\begin{aligned}&-r(t)F(t,r(t);T) + F_t(t,r(t);T) \\ &\quad + F_r(t,r(t);T)b^{\mathbb{Q}}(t,r(t)) + \frac{1}{2}F_{rr}(t,r(t);T)\sigma^2(t,r(t)) = 0.\end{aligned}$$

This equation should hold for any $t < T$ and any possible value of $r(t)$. This implies that $F(t,r;T)$ as a function of t and r should solve the partial differential equation

$$-rF(t,r;T) + F_t(t,r;T) + F_r(t,r;T)b^{\mathbb{Q}}(t,r) + \frac{1}{2}F_{rr}(t,r;T)\sigma^2(t,r) = 0, \tag{8.9}$$

together with the boundary condition $F(T,r;T) = 1$. Equation (8.9), with the boundary condition, is called the *term structure equation* of the bond. Should one impose a different boundary condition of the type $F(T, r; T) = \Phi(r)$ for some appropriate function Φ, representing a claim on the value $r(T)$, then $F(t, r(t); T)$ is the fair price of this claim at time t. This can then be used to price interest rate derivatives in a fashion analogous to claims in a Black-Scholes market. We can turn the situation around.

Proposition 8.1 *Suppose $F(t, r; T)$ is a solution to (8.9) with boundary condition $F(T, r; T) = \Phi(r)$. Suppose that it also satisfies the integrability condition $\mathbb{E}_{\mathbb{Q}}\big(F_r(t, r(t); T)\sigma(t, r(t)) \exp(-\int_0^t r(s)\,\mathrm{d}s)\big)^2 < \infty$ for $t \le T$. Then M defined by $M(t) := \exp(-\int_0^t r(s)\,\mathrm{d}s)F(t, r(t); T)$, $t \le T$, is a martingale under $\mathbb{Q}$ and the price of the claim $\Phi(r(T))$ at time t is given by $F(t, r(t); T) = \mathbb{E}_{\mathbb{Q}}[\exp\big(-\int_t^T r(s)\,\mathrm{d}s\big)\Phi(r(T))|\mathcal{F}_t]$.*

Proof Apply the Itô product rule to $M(t)$ and use the fact F solves (8.9). A computation then shows that (in abbreviated notation) $\mathrm{d}M = \frac{F_r}{B}\sigma \mathrm{d}W^{\mathbb{Q}}$. By the condition on the expectation in the statement of the proposition, it follows that M is indeed a martingale. Hence

$$M(t) = \mathbb{E}_{\mathbb{Q}}[M(T)|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[\exp\big(-\int_0^T r(s)\,\mathrm{d}s\big)\Phi(r(T))|\mathcal{F}_t],$$

where the conditional expectation is now well defined. Multiplication with $B(t) = \exp(\int_0^t r(s)\,\mathrm{d}s)$ gives the price of the claim. □

Corollary 8.2 *Under the assumptions of Proposition 8.1 the bond price satisfies under the probability measures $\mathbb{Q}$ and $\mathbb{P}$ the stochastic differential equations*

$$\mathrm{d}P(t,T) = r(t)P(t,T)\,\mathrm{d}t + P(t,T)\sigma(t;T)\,\mathrm{d}W^{\mathbb{Q}}, \tag{8.10}$$

$$\mathrm{d}P(t,T) = b(t;T)P(t,T)\,\mathrm{d}t + P(t,T)\sigma(t;T)\,\mathrm{d}W, \tag{8.11}$$

where $\sigma(t;T) = \frac{\sigma(t,r(t))F_r(t,r(t);T)}{F(t,r(t);T)}$ is the volatility of $P(t,T)$, and $b(t;T) = r(t) + \lambda(t, r(t))\sigma(t;T)$.

Proof Apply the Itô rule to $F(t, r(t); T)$ and Equation (8.9) to obtain (8.11) and then (8.5) to obtain (8.10). □

Remark 8.3 The relation $\lambda(t, r(t)) = \frac{b(t;T)-r(t)}{\sigma(t;T)}$ can be interpreted as being the market price of risk of the bond. Compare to Remark 7.1 in Section 7.1 for the Black-Scholes market analogue. Note that *this market price of risk is the same for all bonds in the market*, irrespective of their maturity time T.

Looking again at Proposition 8.1 and the martingale property of the $M(t)$ in the proof, we infer that for any claim X that is $\mathcal{F}_T$ measurable and satisfies the appropriate integrability conditions, its fair, arbitrage free, price $V(t)$ at time $t \le T$ is given by

$$V(t) = \mathbb{E}_{\mathbb{Q}}[\exp(-\int_t^T r(s)\,\mathrm{d}s)\,X|\mathcal{F}_t], \tag{8.12}$$

but in general we cannot use (8.9) to find a formula for $V(t)$ as C will not be a function of $r(T)$ only. See Section 8.5 for an alternative approach.

8.3 Affine term structures

One speaks of an *affine term structure* if the function F of Equation (8.8) has a special dependence on t, T and r, namely

$$F(t,r;T) = \exp(-A(t,T) - B(t,T)r), \tag{8.13}$$

for some appropriate functions A and B. The term in the exponential of (8.13) is (for every t, T) an *affine* function of r. Note that necessary in this case is $A(T,T) = 0$ and $B(T,T) = 0$ as $F(T,r;T) = 1$ for all r (and T). We will see that this situation occurs for some popular short rate models, i.e. for certain specific function $b^{\mathbb{Q}}$ and σ in (8.7), although there also exists a much more general theory. In the remainder of the present section we will work under the risk neutral measure $\mathbb{Q}$ and we simply write W and b instead of $W^{\mathbb{Q}}$ and $b^{\mathbb{Q}}$.

Proposition 8.4 *The short rate model* (8.7) *with* $b(t,r) = b_0(t) + b_1(t)r$ *and* $\sigma^2(t,r) = a_0(t) + a_1(t)r \geq 0$, *where* a_i *and* b_i *are continuous functions* ($i = 0, 1$) *yields an affine term structure as in Equation* (8.13) *if and only if the functions* A *and* B *satisfy the differential equations (a system of Riccati equations where the dot denotes differentiation w.r.t.* t*),*

$$\dot{A}(t,T) = \frac{1}{2}a_0(t)B(t,T)^2 - b_0(t)B(t,T),\ A(T,T) = 0, \tag{8.14}$$

$$\dot{B}(t,T) = \frac{1}{2}a_1(t)B(t,T)^2 - b_1(t)B(t,T) - 1,\ B(T,T) = 0. \tag{8.15}$$

Proof Let F be as in (8.13) with the specified functions b and σ. We first compute, omitting the arguments t and T, the partial derivatives $F_t = -F(\dot{A} + \dot{B}r)$, $F_r = -FB$ and $F_{rr} = FB^2$. Plugging all this into (8.9), we obtain, omitting arguments and after dividing by F,

$$-r - (\dot{A} + \dot{B}r) - B(b_0 + b_1 r) + \frac{1}{2}B^2(a_0 + a_1 r) = 0. \tag{8.16}$$

Under the assumption that (8.13) holds with the specified functions b and σ, the partial differential equations (8.9) and (8.16) are equivalent. In other words, under this assumption, F solves (8.9) iff A and B satisfy (8.16).

Suppose now that A and B satisfy (8.14), respectively (8.15). Then, by a simple substitution, also (8.16) is satisfied for all r and $t < T$. Conversely, assume that (8.16) holds, for all r (and $t < T$), in particular for $r = 0$. That case yields $-\dot{A} - Bb_0 + \frac{1}{2}B^2 a_0 = 0$ which gives Equation (8.14). Moreover, using this, one obtains from (8.16) the differential equation $-r - \dot{B}r - Bb_1 r + \frac{1}{2}B^2 a_1 r = 0$, which yields (8.15). □

With the functions b and σ as in Proposition 8.4, Equation (8.7) becomes

$$\mathrm{d}r(t) = (b_0(t) + b_1(t)r(t))\,\mathrm{d}t + \sqrt{a_0(t) + a_1(t)r(t)}\,\mathrm{d}W(t). \tag{8.17}$$

This SDE needs to have a solution and in particular the argument of the square root has to be nonnegative for all possible values of $r(t)$. We discern two cases.

In the first case we allow $r(t)$ to have any real value and so we have to impose (for all t) $a_1(t) = 0$ and $a_0(t) \geq 0$. In the second case we want $r(t)$ to be nonnegative and then we have to impose $a_0(t) = 0$, $a_1(t) \geq 0$ and $b_0(t) \geq$

0. To understand the latter conditions, let's see what happens if $r(t) = 0$ in (8.17) in a rather heuristic way. One then gets $\mathrm{d}r(t) = b_0(t)\,\mathrm{d}t + \sqrt{a_0(t)}\,\mathrm{d}W(t)$. To assure that $r(t)$ remains nonnegative one needs $b_0(t) \geq 0$ (gives then an upward push) and $\sqrt{a_0(t)}$ needs to be zero, otherwise the possibly negative Brownian increment $\mathrm{d}W(t)$ (although not much more than notation, think of it as a random variable) could push $r(t)$ to fall below zero.

So we have for each of these cases a model. The first one is

$$\mathrm{d}r(t) = (b_0(t) + b_1(t)r(t))\,\mathrm{d}t + \sqrt{a_0(t)}\,\mathrm{d}W(t),$$

with $a_0(t) \geq 0$. It is common to take the functions b_1 and a_0 as constants, and with $\sigma = \sqrt{a_0}$, we get

$$\mathrm{d}r(t) = (b_0(t) + b_1 r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t), \tag{8.18}$$

This model is called the Hull-White model. In the particular case that also $b_0(\cdot)$ is a constant function, denoted b_0, one has the Vasiček model,

$$\mathrm{d}r(t) = (b_0 + b_1 r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t). \tag{8.19}$$

The Vasiček model produces random variables $r(t)$ that have a normal distribution, see Exercise 8.1, if $r(0)$ has a normal distribution. As a consequence $r(t)$ can assume negative values with positive probability, which has been critized in the past. By now, this has become a 'normal' situation. To alleviate the criticism, one may switch to a different model in which $r(t)$ is guaranteed to stay nonnegative. This (second) model is

$$\mathrm{d}r(t) = (b_0(t) + b_1(t)r(t))\,\mathrm{d}t + \sqrt{a_1(t)r(t)}\,\mathrm{d}W(t),\ r(0) > 0,$$

with $a_1(t) \geq 0$ and $b_0(t) \geq 0$. Of particular interest is the case where $a_1(t)$, $b_0(t)$ and $b_1(t)$ are constant functions, $a_1 \geq 0$ and $b_0 \geq 0$. This is the Cox-Ingersoll-Ross (CIR) model, with $\sigma = \sqrt{a_1} > 0$,

$$\mathrm{d}r(t) = (b_0 + b_1 r(t))\,\mathrm{d}t + \sigma\sqrt{r(t)}\,\mathrm{d}W(t). \tag{8.20}$$

The $r(t)$ resulting from (8.20) doesn't have a particularly nice distribution, for which however formulas are available. For instance, conditional on $r(0)$, $r(t)$ has a scaled non-central chi-squared distribution. However, for $b_1 < 0$, $r(t)$ has for large t approximately a gamma distribution, $\Gamma(\frac{2b_0}{\sigma^2}, \frac{-2b_1}{\sigma^2})$.

The Riccati Equations (8.14) and (8.15) are in general difficult, if not impossible, to analytically solve in view of the time dependent coefficients. This holds especially for (8.15), whereas $A(t,T)$ is an integral that can hopefully computed once $B(t,T)$ is known. For the Vasiček and CIR models with constant coefficients there are explicit solutions, see Exercises 8.2 and 8.3.

8.4 Forward curve fitting

The Vasiček model has only a very few parameters (as well as the CIR model). In principle they can be obtained by using observed bond prices of $P(0,T)$, the term structure at time $t = 0$. But there are many of them, just as many maturity times (in principle infinitely many). Hence obtaining a *perfect fit* of the Vasiček model to all observed bond prices is hopeless. A least squares approach to fit parameters of the computed values of $P(0,T)$ for this model to their observed counterparts is a common way out, hoping it results in an acceptable fit. The Hull-White extension (8.18) of the Vasiček model (8.19) has the non-constant function $b_0(\cdot)$ and the resulting flexibility turns out to be helpful to fit the initial forward curve.

Recall from (8.2) that $P(0,T) = \exp(-\int_0^T f(0,u)\,\mathrm{d}u)$ and combine this with Equations (8.8) and (8.13). Write $'$ to denote differentiation w.r.t. T and obtain

$$f(0,T) = A'(0,T) + B'(0,T)r(0).$$

The Riccati equations (8.14) and (8.15) take the form

$$\dot{A}(t,T) = \frac{1}{2}\sigma^2 B(t,T)^2 - b_0(t)B(t,T),\ A(T,T) = 0,$$
$$\dot{B}(t,T) = -b_1 B(t,T) - 1,\ B(T,T) = 0.$$

A simple calculation shows that (b_1 is taken nonzero)

$$B(t,T) = \frac{1}{b_1}(e^{b_1(T-t)} - 1),$$
$$A(t,T) = -\frac{\sigma^2}{2}\int_t^T B(s,T)^2\,\mathrm{d}s + \int_t^T b_0(s)B(s,T)\,\mathrm{d}s.$$

Note that $\dot{B}(t,T) = -B'(t,T)$. These relations are useful in the proof of the next Proposition.

Proposition 8.5 *The initial forward curve $T \mapsto f(0,T)$ is in the Hull-White model related to the function b_0 by*

$$b_0(T) = f'(0,T) + \frac{\sigma^2}{2}\frac{\partial}{\partial T}B(0,T)^2 - b_1 f(0,T) - b_1\frac{\sigma^2}{2}B(0,T). \tag{8.21}$$

Proof Exercise 8.5. □

Proposition 8.5 can be used to fit the Hull-White model to an observed curve of initial forward rates $f^*(0,T)$ (assumed to be differentiable in T). Given b_1 and σ, one obtains a to the observations fitted function b_0^* by replacing $f(0,T)$ in (8.21) by $f^*(0,T)$. What then remains to be done for a full specification of the model is to choose b_1 and σ, in practice often done by matching theoretical prices with observed prices for a selected number of financial products, or by different considerations that are for some reason 'convenient'.

The proposition has a nice corollary. If one has observed bond prices $P^*(0,t)$ and corresponding forward rates $f^*(0,t)$ for all t in an interval $[0,T]$, and if one assumes the Hull-White model, then it is possible to determine also the prices $P(t,T)$ of bounds maturing at T at any time $t \leq T$.

Corollary 8.6 *Assume the Hull-White model. Then for any $t \leq T$ bond prices $P(t,T)$ can be computed from observed bond prices and forward rates at time $t = 0$ in a consistent way by*

$$P(t,T) = \frac{P^*(0,T)}{P^*(0,t)}\exp\big(B(t,T)(f^*(0,t) - r(t)) + \frac{\sigma^2}{4b_0}B^2(t,T)(1 - e^{2b_0 t})\big).$$

Proof There is not much more to do than using b_0 as given (8.21) in $A(t,T)$, (8.13) and performing very tedious calculations. These are omitted. □

8.5 Forward measures

The risk neutral EMM $\mathbb{Q}$ is such that by B discounted prices are $\mathbb{Q}$-martingales. As an alternative to this way of discounting, we can look at discounting prices by a bond price process $P(\cdot, T)$, the process associated to a bond maturing at T, and the question we'd like to answer is whether we can find a probability measure, denoted $\mathbb{Q}^T$ and called the T-*forward measure* or simply *forward measure*, such that in this way discounted processes become martingales under $\mathbb{Q}^T$. The answer is, not surprising, positive. Before stating the precise result, let us introduce some notation. Recall that $M(t) = \frac{P(t,T)}{B(t)}$ gives a $\mathbb{Q}$-martingale, and then in particular $P(0,T) = \mathbb{E}_{\mathbb{Q}} \frac{1}{B(T)}$. Put

$$L^T = \frac{1}{B(T)P(0,T)},$$

and observe that $\mathbb{E}_{\mathbb{Q}} L^T = 1$. Hence we can use L^T to define a new *probability* measure $\mathbb{Q}^T$ on $\mathcal{F}_T$ by $\frac{\mathrm{d}\mathbb{Q}^T}{\mathrm{d}\mathbb{Q}} = L^T$. Furthermore we can compute, using the just mentioned martingale property, for $t \leq T$,

$$\begin{aligned} L^T(t) &:= \mathbb{E}_{\mathbb{Q}}[L^T|\mathcal{F}_t] = \frac{1}{P(0,T)} \mathbb{E}_{\mathbb{Q}}[\frac{P(T,T)}{B(T)}|\mathcal{F}_t] \\ &= \frac{1}{P(0,T)} \frac{P(t,T)}{B(t)} = \frac{P(t,T)}{P(0,T)B(t)} \end{aligned} \tag{8.22}$$

and note that $L^T(0) = 1$ and $L^T(T) = L^T$.

Let's turn to the pricing formula (8.12). To evaluate the (conditional) expectation, one in principle needs the joint distribution of $(X, B(T))$ under $\mathbb{Q}$. Of course, if $B(T)$ is nonrandom, one can take it out of the expectation, which substantially simplifies the computation. The interesting phenomenon is that something similar happens if we consider pricing under the T-forward measure. Note that for deterministic interest rates nothing changes as then $\mathbb{Q}$ and $\mathbb{Q}^T$ are the same, as $L^T = 1$ in such a case.

Proposition 8.7 *Assume that X is $\mathcal{F}_T$-measurable and that the expectation $\mathbb{E}_{\mathbb{Q}} \frac{|X|}{B(T)}$ is finite. Then the price $V(t)$ at time t of X is given by*

$$V(t) = P(t,T)\mathbb{E}_{\mathbb{Q}^T}[X|\mathcal{F}_t].$$

Proof First we note that $\mathbb{E}_{\mathbb{Q}^T}|X| = \mathbb{E}_{\mathbb{Q}}|X|L^T = \frac{1}{P(0,T)}\mathbb{E}_{\mathbb{Q}}\frac{|X|}{B(T)} < \infty$. We use Equation (4.5) with the appropriate measures and Radon-Nikodym derivatives to develop

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}^T}[X|\mathcal{F}_t] &= \frac{1}{L^T(t)} \mathbb{E}_{\mathbb{Q}}[XL^T|\mathcal{F}_t] \\ &= \frac{1}{L^T(t)} \mathbb{E}_{\mathbb{Q}}[X \frac{1}{B(T)P(0,T)}|\mathcal{F}_t] \\ &= \frac{1}{L^T(t)P(0,T)} \mathbb{E}_{\mathbb{Q}}[\frac{X}{B(T)}|\mathcal{F}_t] \\ &= \frac{1}{L^T(t)P(0,T)} \frac{V(t)}{B(t)} \end{aligned}$$

$$= \frac{1}{P(t,T)} V(t),$$

where the last equality follows from (8.22). The assertion follows. □

The assertion of this proposition is one of the main reasons to work with forward measures. Note that the forward measure $\mathbb{Q}^T$ depends on the maturity T of the bond. Here is another nice property of forward measures.

Proposition 8.8 *The forward rates $f(t,T)$, $0 \le t \le T$, in Equation (8.1) form a martingale under the forward measure $\mathbb{Q}^T$. More precisely,*

$$f(t,T) = \mathbb{E}_{\mathbb{Q}^T}[r(T)|\mathcal{F}_t].$$

Proof Let the prime $'$ denote differentiation w.r.t. T. Then we claim

$$\mathbb{E}_{\mathbb{Q}}[\frac{r(T)}{B(T)}|\mathcal{F}_t] = -\frac{P'(t,T)}{B(t)}. \tag{8.23}$$

Accepting for a while that the claim holds true, we develop by analogy with Proposition 4.5

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}^T}[r(T)|\mathcal{F}_t] &= \frac{\mathbb{E}_{\mathbb{Q}}[r(T)L(T)|\mathcal{F}_t]}{L^T(t)} = \frac{\mathbb{E}_{\mathbb{Q}}[\frac{r(T)}{P(0,T)B(T)}|\mathcal{F}_t]}{\frac{P(t,T)}{P(0,T)B(t)}} \\
&= \frac{\mathbb{E}_{\mathbb{Q}}[\frac{r(T)}{B(T)}|\mathcal{F}_t]}{\frac{P(t,T)}{B(t)}} = \frac{-\frac{P'(t,T)}{B(t)}}{\frac{P(t,T)}{B(t)}} \\
&= -\frac{P'(t,T)}{P(t,T)} = f(t,T),
\end{aligned}$$

which is what the proposition asserts.

We proceed by proving the claim (8.23). To that end we consider, using Fubini's theorem for conditional expectations (take this for granted) in the first equality,

$$\begin{aligned}
\int_t^T \mathbb{E}_{\mathbb{Q}}[\frac{r(u)}{B(u)}|\mathcal{F}_t]\,\mathrm{d}u &= \mathbb{E}_{\mathbb{Q}}[\int_t^T \frac{r(u)}{B(u)}\,\mathrm{d}u|\mathcal{F}_t] \\
&= \mathbb{E}_{\mathbb{Q}}[-\frac{1}{B(T)} + \frac{1}{B(t)}|\mathcal{F}_t] \\
&= -\frac{p(t,T)}{B(t)} + \frac{1}{B(t)}.
\end{aligned}$$

By differentiation w.r.t. T one gets (8.23). □

Assume that the bond prices $P(t,T)$ are as in (8.8) and the short rates $r(t)$ satisfy (8.7). It then follows that $P(t,T)$ satisfies the stochastic differential equation

$$\mathrm{d}P(t,T) = r(t)P(t,T)\,\mathrm{d}t + P(t,T)\sigma(t;T)\,\mathrm{d}W^{\mathbb{Q}}(t),$$

for a certain process $\sigma(t;T)$; see Exercise 8.4. Let's look at the discounted bond price $\bar{P}(t,T) = \frac{P(t,T)}{B(t)}$, $0 \le t \le T$. It satisfies the equation

$$\mathrm{d}\bar{P}(t,T) = \bar{P}(t,T)\sigma(t;T)\,\mathrm{d}W^{\mathbb{Q}}(t).$$

From this, using that $L^T(t) = \frac{\bar{P}(t,T)}{P(0,T)}$ we obtain the dynamics for $L^T(t)$,

$$\mathrm{d}L^T(t) = L^T(t)\sigma(t;T)\,\mathrm{d}W^{\mathbb{Q}}(t),\ L^T(0) = 1,$$

and the process W^T defined by

$$W^T(t) = W^{\mathbb{Q}}(t) - \int_0^t \sigma(u;T)\,\mathrm{d}u \tag{8.24}$$

is a Brownian motion under $\mathbb{Q}^T$ for $t \leq T$, as follows from Girsanov's theorem 6.19.

The forward measure is such that a tradable asset with price process X becomes a martingale under $\mathbb{Q}^T$, when discounted with the bond price $P(\cdot, T)$. To see this we compute similar to the proof of Proposition 8.7, using Proposition 4.5, for $t \leq T$, and that discounting with B gives martingales under $\mathbb{Q}$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}^T}[\frac{X(T)}{P(T,T)}|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}^T}[X(T)|\mathcal{F}_t] &= \frac{\mathbb{E}_{\mathbb{Q}}[X(T)L^T|\mathcal{F}_t]}{L^T(t)} \\
&= \frac{\mathbb{E}_{\mathbb{Q}}[\frac{X(T)}{B(T)P(0,T)}|\mathcal{F}_t]}{\frac{P(t,T)}{P(0,T)B(t)}} = \frac{\mathbb{E}_{\mathbb{Q}}[\frac{X(T)}{B(T)}|\mathcal{F}_t]}{\frac{P(t,T)}{B(t)}} \\
&= \frac{\frac{X(t)}{B(t)}}{\frac{P(t,T)}{B(t)}} = \frac{X(t)}{P(t,T)}.
\end{aligned}$$

There are many other possibilities for measure changes that yield martingales under a new measure after discounting with an appropriate *numéraire*. One of the other options is discounting with the stock price S. To do that we introduce a measure $\mathbb{R}$ (the notation should not be confused with the same one for the real numbers) with a Radon-Nikodym derivative on $\mathcal{F}_T$ given by

$$R(T) := \frac{\mathrm{d}\mathbb{R}}{\mathrm{d}\mathbb{Q}} = \frac{S(T)}{B(T)}\frac{1}{S(0)}.$$

By the martingale property of the discounted stock prices $\frac{S(t)}{B(t)}$ under $\mathbb{Q}$ one has for $t \leq T$,

$$R(t) := \mathbb{E}_{\mathbb{Q}}[R(T)|\mathcal{F}_t] = \frac{S(t)}{B(t)}\frac{1}{S(0)}.$$

As above for working under the forward measure $\mathbb{Q}^T$, one now has that any price process $X(\cdot)$ becomes martingale under $\mathbb{R}$ after discounting with S. Indeed, one has (assuming all expectations make sense)

$$\mathbb{E}_{\mathbb{R}}[\frac{X(T)}{S(T)}|\mathcal{F}_t] = \frac{X(t)}{S(t)}, \tag{8.25}$$

which is the content of Exercise 8.7.

This property is convenient for determining European call option prices in the presence of stochastic interest rates. Recall the pay-off of a European call,

$X = (S(T) - K)^+$. This can conveniently be written as $X = S(T)\mathbf{1}_{\{S(T)\geq K\}} - K\mathbf{1}_{\{S(T)\geq K\}}$. From (8.12) we have that the fair price of X at time t is given by

$$\begin{aligned} V(t) &:= B(t)\mathbb{E}_{\mathbb{Q}}[\frac{X}{B(T)}|\mathcal{F}_t] \\ &= B(t)\left(\mathbb{E}_{\mathbb{Q}}[\frac{S(T)}{B(T)}\mathbf{1}_{\{S(T)\geq K\}}|\mathcal{F}_t] - K\mathbb{E}_{\mathbb{Q}}[\frac{1}{B(T)}\mathbf{1}_{\{S(T)\geq K\}}|\mathcal{F}_t]\right). \end{aligned}$$

For the first term in the big parentheses one has, using the intermediate result derived from Proposition 4.5 and valid for any Z for which the expectation exist,

$$\mathbb{E}_{\mathbb{Q}}[ZR(T)|\mathcal{F}_t] = R(t)\mathbb{E}_{\mathbb{R}}[Z|\mathcal{F}_t].$$

Applied to $Z = S(0)\mathbf{1}_{\{S(T)\geq K\}}$ this gives,

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[\frac{S(T)}{B(T)}\mathbf{1}_{\{S(T)\geq K\}}|\mathcal{F}_t] &= R(t)S(0)\mathbb{E}_{\mathbb{R}}[\mathbf{1}_{\{S(T)\geq K\}}|\mathcal{F}_t] \\ &= R(t)S(0)\mathbb{R}(S(T) \geq K|\mathcal{F}_t) \\ &= R(t)S(0)\mathbb{R}(\frac{1}{S(T)} \leq \frac{1}{K}|\mathcal{F}_t). \end{aligned}$$

For the second term in the big parentheses one similarly has, using now the arguments for the forward measure,

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[\frac{1}{B(T)}\mathbf{1}_{\{S(T)\geq K\}}|\mathcal{F}_t] &= L^T(t)P(0,T)\mathbb{E}_{\mathbb{Q}^T}[\mathbf{1}_{\{S(T)\geq K\}}|\mathcal{F}_t] \\ &= L^T(t)\mathbb{Q}^T(S(T) \geq K|\mathcal{F}_t). \end{aligned}$$

Summing up, we obtain

$$\begin{aligned} V(t) &= B(t)R(t)S(0)\mathbb{R}(\frac{1}{S(T)} \leq \frac{1}{K}|\mathcal{F}_t) \\ &\quad - KB(t)L^T(t)P(0,T)\mathbb{Q}^T(S(T) \geq K|\mathcal{F}_t) \\ &= S(t)\mathbb{R}(\frac{1}{S(T)} \leq \frac{1}{K}|\mathcal{F}_t) - KP(t,T)\mathbb{Q}^T(S(T) \geq K|\mathcal{F}_t). \end{aligned} \tag{8.26}$$

To get a clean explicit formula for $V(t)$, not only in terms of abstract conditional probabilities, further assumptions have to be imposed. Recall Exercise 8.4 and the equation given there for $P(t,T)$. In a Black-Scholes setting we have similar to (7.1)

$$\mathrm{d}S(t) = r(t)S(t)\,\mathrm{d}t + \sigma(t)S(t)\,\mathrm{d}W^{\mathbb{Q}}(t)$$

where we can take σ a suitable piecewise continuous process. Put $\tilde{S}(t) = \frac{S(t)}{P(t,T)}$. Then $\tilde{S}$ is a martingale under the forward measure $\mathbb{Q}^T$ if the $\mathbb{Q}^T$-expectations $\mathbb{E}_{\mathbb{Q}^T}\tilde{S}(t)$ are finite for all $t \leq T$. In fact one has, Exercise 8.8,

$$\mathrm{d}\tilde{S}(t) = \tilde{\sigma}(t)\tilde{S}(t)\,\mathrm{d}W^T(t), \tag{8.27}$$

for a suitable process $\tilde{\sigma}$ and with W^T as in (8.24).

Proposition 8.9 *Assume that the $\tilde\sigma(t)$ in (8.27) are deterministic. Then the price $V(t)$ of the European call option as in (8.26) is given by the formula*

$$V(t) = S(t)\Phi(d_+(t)) - KP(t,T)\Phi(d_-(t)), \tag{8.28}$$

where

$$d_+(t) = \frac{\log\frac{\tilde S(t)}{K} + \frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}{\sqrt{\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}}, \tag{8.29}$$

$$d_-(t) = \frac{\log\frac{\tilde S(t)}{K} - \frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}{\sqrt{\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}}. \tag{8.30}$$

Proof From (8.27) one obtains

$$\tilde S(T) = \tilde S(t)\exp(\int_t^T \tilde\sigma(u)\,\mathrm{d}W^T(u) - \frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u) =: \tilde S(t)E(t,T),$$

where the term in the exponential has under $\mathbb{Q}^T$ a normal distribution with mean $-\frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u$ and variance $\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u$ and is also independent of $\mathcal{F}_t$. This has the following consequence.

$$\begin{aligned}
\mathbb{Q}^T(S(T) \geq K|\mathcal{F}_t) &= \mathbb{Q}^T(\tilde S(t)E(t,T) \geq K|\mathcal{F}_t)\\
&= \mathbb{Q}^T(\log E(t,T) \geq \log\frac{K}{\tilde S(t)}|\mathcal{F}_t)\\
&= 1 - \Phi(\frac{\log\frac{K}{\tilde S(t)} + \frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}{\sqrt{\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}})\\
&= \Phi(\frac{\log\frac{\tilde S(t)}{K} - \frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}{\sqrt{\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u}}),
\end{aligned}$$

which yields the second term in (8.28) because of (8.30).

We move on to the probability in the first term of (8.28). Note first that one can write, by yet another application of Girsanov's theorem,

$$\begin{aligned}
\frac{1}{\tilde S(T)} &= \frac{1}{\tilde S(t)}\exp\big(-\int_t^T \tilde\sigma(u)\,\mathrm{d}W^T(u) + \frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u\big)\\
&= \frac{1}{\tilde S(t)}\exp\big(-\int_t^T \tilde\sigma(u)\,\mathrm{d}W^{\mathbb{R}}(u) - \frac{1}{2}\int_t^T \tilde\sigma(u)^2\,\mathrm{d}u\big)\\
&=: \frac{1}{\tilde S(t)}\tilde E(t,T),
\end{aligned}$$

where

$$W^{\mathbb{R}}(t) = W^T(t) - \int_0^t \tilde\sigma(u)\,\mathrm{d}u$$

gives a Brownian motion under the probability measure $\mathbb{R}$. As a consequence, $\log \tilde{E}(t,T)$ has under $\mathbb{R}$ a normal distribution with mean $-\frac{1}{2}\int_t^T \tilde{\sigma}(u)^2\,\mathrm{d}u$ and variance $\int_t^T \tilde{\sigma}(u)^2\,\mathrm{d}u$ and is independent of $\mathcal{F}_t$. We now compute

$$\begin{aligned}
\mathbb{R}(\frac{1}{S(T)} \leq \frac{1}{K}|\mathcal{F}_t) &= \mathbb{R}(\frac{1}{\tilde{S}(T)} \leq \frac{1}{K}|\mathcal{F}_t)\\
&= \mathbb{R}(\frac{1}{\tilde{S}(t)}\tilde{E}(t,T) \leq \frac{1}{K}|\mathcal{F}_t)\\
&= \mathbb{R}(\tilde{E}(t,T) \leq \frac{\tilde{S}(t)}{K}|\mathcal{F}_t)\\
&= \mathbb{R}(\tilde{E}(t,T) \leq \frac{\tilde{S}(t)}{K})\\
&= \Phi(\frac{\frac{\tilde{S}(t)}{K} + \frac{1}{2}\int_t^T \tilde{\sigma}(u)^2\,\mathrm{d}u}{\sqrt{\int_t^T \tilde{\sigma}(u)^2\,\mathrm{d}u}}),
\end{aligned}$$

as desired in view of (8.29). □

Proposition 8.9 gives a formula for the price of a European call option on a stock in the presence of stochastic interest rates. The formula resembles the Black-Scholes formula for the price of a European call option on a stock with constant interest rate and constant parameters in the equation for $S(t)$. Indeed, in the latter case Equation (8.28) reduces to (1.30) for $V(t) = C(t,S(t))$.

A similar result exists for the price of European call option on a bond, also the techniques to arrive at a pricing formula are similar.

Recall that the forward measure $\mathbb{Q}^T$ depends on the maturity T of the bond. For another maturity, S say, there is a different forward measure, which can then be used for the pricing of an $\mathcal{F}_S$ measurable claim. Consider the forward measure $\mathbb{Q}^S$ for $S > T$. Then, on $\mathcal{F}_S$, the Radon-Nikodym derivative is $L^S := \frac{\mathrm{d}\mathbb{Q}^S}{\mathrm{d}\mathbb{Q}} = \frac{1}{B(S)P(0,S)}$ and, analogous to (8.22) with the approporiate substitutions, we have

$$L^S(T) := \mathbb{E}_{\mathbb{Q}}[L(S)|\mathcal{F}_T] = \frac{P(T,S)}{P(0,S)B(T)},$$

which is the Radon-Nikodym derivative of $\mathbb{Q}^S$ w.r.t. $\mathbb{Q}$ when restricted to $\mathcal{F}_T$. We will use this for pricing a special call option, introduced below, where we see that we use expectations and probabilities under the two forward measures $\mathbb{Q}^S$ and $\mathbb{Q}^T$.

Recall that, for $T < S$, $P(T,S)$ is the price at time T for a bond maturing at time S. We consider the European call option on the bond with pay-off $X := (P(T,S) - K)^+$ at time T. The arbitrage free price of X at time $t \leq T$ is $p(t) := B(t)\mathbb{E}_{\mathbb{Q}}[\exp(-\int_0^T r(s)\,\mathrm{d}s)(P(T,S) - K)^+|\mathcal{F}_t]$. We can write, similar to the pricing of the European call,

$$\begin{aligned}
\frac{p(t)}{B(t)} &= \mathbb{E}_{\mathbb{Q}}[\exp(-\int_0^T r(s)\,\mathrm{d}s)(P(T,S) - K)\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t]\\
&= \mathbb{E}_{\mathbb{Q}}[\frac{P(T,S)}{B(T)}\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}}[\frac{K}{B(T)}\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t].
\end{aligned}$$

We treat the two conditional expectations separately. First we have, using the forward measure $\mathbb{Q}^S$ and another application of Proposition 4.5,

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[\frac{P(T,S)}{B(T)}\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t] &= P(0,S)\mathbb{E}_{\mathbb{Q}}[L^S(T)\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t] \\ &= P(0,S)\mathbb{E}_{\mathbb{Q}^S}[\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t]\,L^S(t) \\ &= \frac{P(t,S)}{B(t)}\mathbb{Q}^S(P(T,S)\geq K|\mathcal{F}_t) \\ &= \frac{P(t,S)}{B(t)}\mathbb{Q}^S(\frac{P(T,T)}{P(T,S)}\leq\frac{1}{K}|\mathcal{F}_t).\end{aligned}$$

Second we have

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[\frac{K}{B(T)}\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t] &= KP(0,T)\mathbb{E}_{\mathbb{Q}}[L^T(T)\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t] \\ &= KP(0,T)\mathbb{E}_{\mathbb{Q}^T}[\mathbf{1}_{\{P(T,S)\geq K\}}|\mathcal{F}_t]L^T(t) \\ &= K\frac{P(t,T)}{B(t)}\mathbb{Q}^T(P(T,S)\geq K|\mathcal{F}_t) \\ &= K\frac{P(t,T)}{B(t)}\mathbb{Q}^T(\frac{P(T,S)}{P(T,T)}\geq K|\mathcal{F}_t).\end{aligned}$$

It follows that

$$p(t) = P(t,S)\mathbb{Q}^S(\frac{P(T,T)}{P(T,S)}\leq\frac{1}{K}|\mathcal{F}_t) - KP(t,T)\mathbb{Q}^T(\frac{P(T,S)}{P(T,T)}\geq K|\mathcal{F}_t). \quad (8.31)$$

The content of the next proposition is much like Proposition 8.9.

Proposition 8.10 *Consider the European option on the bond price $P(T,S)$ for $T<S$. Let $\sigma^{S,T}(u)=\sigma(u;S)-\sigma(u;T)$ and assume that the $\sigma^{S,T}(u)$ are deterministic. Then the price $p(t)$ of this option at time $t\leq T$ is given by*

$$p(t) = P(t,S)\Phi(d_+(t)) - KP(t,T)\Phi(d_-(t)), \quad (8.32)$$

where

$$d_+(t) = \frac{\log\frac{P(t,S)}{KP(t,T)} + \frac{1}{2}\int_t^T \sigma^{S,T}(u)^2\,\mathrm{d}u}{\sqrt{\int_t^T \sigma^{S,T}(u)^2\,\mathrm{d}u}}, \quad (8.33)$$

$$d_-(t) = \frac{\log\frac{P(t,S)}{KP(t,T)} - \frac{1}{2}\int_t^T \sigma^{S,T}(u)^2\,\mathrm{d}u}{\sqrt{\int_t^T \sigma^{S,T}(u)^2\,\mathrm{d}u}}. \quad (8.34)$$

Proof Now we will use the result of Exercise 8.6 and the notation there. That exercise implies that $\frac{P(t,S)}{P(t,T)}$ is a martingale under $\mathbb{Q}^T$ given by

$$\frac{P(t,S)}{P(t,T)} = \frac{P(0,S)}{P(0,T)}\exp\big(\int_0^t \sigma^{S,T}(u)\,\mathrm{d}W^T(u) - \frac{1}{2}\int_0^t \sigma^{S,T}(u)^2\,\mathrm{d}u\big).$$

From this one obtains

$$\frac{P(T,S)}{P(T,T)} = \frac{P(t,S)}{P(t,T)}\exp\big(\int_t^T \sigma^{S,T}(u)\,\mathrm{d}W^T(u) - \frac{1}{2}\int_t^T \sigma^{S,T}(u)^2\,\mathrm{d}u\big)$$

$$=: \frac{P(t,S)}{P(t,T)}\hat{E}(t,T).$$

We observe that $\log \hat{E}(t,T)$ has a normal distribution under $\mathbb{Q}^T$ with mean $-\frac{1}{2}\int_t^T \sigma^{S,T}(u)^2 \,\mathrm{d}u$ and variance $\int_t^T \sigma^{S,T}(u)^2 \,\mathrm{d}u$ and that it is independent of $\mathcal{F}_t$. Hence

$$\begin{aligned}
\mathbb{Q}^T(\frac{P(T,S)}{P(T,T)} \geq K|\mathcal{F}_t) &= \mathbb{Q}^T(\frac{P(t,S)}{P(t,T)}\hat{E}(t,T) \geq K|\mathcal{F}_t) \\
&= \mathbb{Q}^T(\hat{E}(t,T) \geq \frac{KP(t,T)}{P(t,S)}|\mathcal{F}_t) \\
&= 1 - \Phi(\frac{\log\frac{KP(t,T)}{P(t,S)} + \frac{1}{2}\int_t^T \sigma^{S,T}(u)^2 \,\mathrm{d}u}{\sqrt{\int_t^T \sigma^{S,T}(u)^2 \,\mathrm{d}u}}) \\
&= \Phi(\frac{\log\frac{P(t,S)}{KP(t,T)} - \frac{1}{2}\int_t^T \sigma^{S,T}(u)^2 \,\mathrm{d}u}{\sqrt{\int_t^T \sigma^{S,T}(u)^2 \,\mathrm{d}u}}) \\
&= \Phi(d_-(t)),
\end{aligned}$$

with $d_-(t)$ as in (8.33). The work on the conditional probability $\mathbb{Q}^S(\frac{P(T,T)}{P(T,S)} \leq \frac{1}{K}|\mathcal{F}_t)$ is similar and left as part of Exercise 8.6. □

Remark 8.11 The crucial assumptions in Proposition 8.9 and Proposition 8.10 are that the volatilities there are supposed to be deterministic. If the underlying model for $r(t)$ is the Vasiček model, the assumptions are satisfied and it is possible to obtain explicit formulas for the prices of the options on stok and bond. We refrain from the tedious computations.

8.6 Pricing of complexer products

Recall that Proposition 8.1 and its consequence Equation (8.12) give the price of a claim C. Complexer products may also have maturity times prior to T. Suppose that $T_1 < T$ and that C_1 is a claim that is $\mathcal{F}_{T_1}$-measurable of which the appropriate expectations (under $\mathbb{Q}$) exist. The claim expires at T_1 and it becomes worthless after that time. But the value of the claim at time $t \leq T_1$ is simply $V^1(t) = B(t)\mathbb{E}_{\mathbb{Q}}[\exp(-\int_0^{T_1} r(u)\,\mathrm{d}u)C_1|\mathcal{F}_t]$. More general, assume there are times $0 < T_1 < \cdots < T_n < T$ and claims C_i that are $\mathcal{F}_{T_i}$-measurable random variables for which the expectations below exist and which pay off at the times T_i. Then the value at time $t \leq T_1$ of the portfolio consisting of these n claims is

$$V(t) = B(t)\mathbb{E}_{\mathbb{Q}}[\sum_{i=1}^n \exp(-\int_0^{T_i} r(u)\,\mathrm{d}u)C_i|\mathcal{F}_t] = B(t)\mathbb{E}_{\mathbb{Q}}[\sum_{i=1}^n \frac{C_i}{B(T_i)}|\mathcal{F}_t]. \quad (8.35)$$

Let us consider bonds with a nominal value v_i that expire on the dates T_i. We can consider each of this products as claims with values at any time $t \leq T_i$ equal to $v_iP(t,T_i)$, where $P(t,T_i)$ is the time t value of a zero coupon bond as in Section 8.1. Hence the price at time $t \leq T_1$ of a portfolio of such claims, called *coupons*, is simply $\sum_{i=1}^n v_iP(t,T_i)$. We will give a number of examples of complex products. Recall Equation (8.4) which we will repeatedly use below in the form (for various values of $t \leq T$)

$$\mathbb{E}_{\mathbb{Q}}[\frac{1}{B(T)}|\mathcal{F}_t] = \frac{P(t,T)}{B(t)}, \tag{8.36}$$

as well as the martingale property under $\mathbb{Q}$ of processes discounted with $B(\cdot)$. Recall that Equation (8.36) is an instance of this martingale property as we have $P(T,T) = 1$.

Coupon bonds

A (fixed) coupon bond is a contract specified by a number of future dates $T_1 < \cdots < T_n$ (the coupon dates), where T_n is the maturity of the bond). At each coupon date T_i a deterministic amount (coupon) c_i is paid and an additional nominal value N at maturity. The price $p(t)$ at time $t \leq T_1$ of this coupon bond is given by the sum of discounted cash flows,

$$p(t) = \sum_{i=1}^{n} c_i P(t,T_i) + NP(t,T_n).$$

Often there is a fixed period $\delta = T_{i+1} - T_i$, and the coupons are given as a fixed fraction of the nominal value, $c_i = K\delta N$, for some fixed interest rate K. The above formula reduces to

$$p(t) = N(\sum_{i=1}^{n} K\delta P(t,T_i) + P(t,T_n)).$$

Floating rate bonds

A floating rate bond is specified by future dates $T_0 < T_1 < \cdots < T_n$ and a nominal value N, for simplicity one can take $N = 1$. The previous deterministic coupon payments c_i for the fixed coupon bond are now replaced with

$$C_i = (T_i - T_{i-1})F(T_{i-1},T_i)N,$$

where, more general, $F(t,T)$ is the average rate, the return per unit time over an interval $[t,T]$ of an investment at time t in a bond with maturity T,

$$F(t,T) = \frac{1}{T-t}(\frac{1}{P(t,T)} - 1).$$

Note that $F(T_{i-1},T_i)$ is determined already at time T_{i-1}. The value $p(t)$ of this note at time $t \leq T_0$ is obtained as follows. By definition of $F(T_{i-1},T_i)$ one has $C_i = \frac{1}{P(T_{i-1},T_i)} - 1$. To determine the price of this product we use Equation (8.35).

We start with computing the conditional expectation for a single term indexed by i,

$$\mathbb{E}_{\mathbb{Q}}[\frac{C_i}{B(T_i)}|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}[\frac{\frac{1}{P(T_{i-1},T_i)} - 1}{B(T_i)}|\mathcal{F}_t]$$

$$= \mathbb{E}_{\mathbb{Q}}[\mathbb{E}_{\mathbb{Q}}[\frac{\frac{1}{P(T_{i-1},T_i)} - 1}{B(T_i)}|\mathcal{F}_{T_{i-1}}]|\mathcal{F}_t]$$
$$= \mathbb{E}_{\mathbb{Q}}[(\frac{1}{P(T_{i-1},T_i)} - 1)\mathbb{E}_{\mathbb{Q}}[\frac{1}{B(T_i)}|\mathcal{F}_{T_{i-1}}]|\mathcal{F}_t]$$
$$= \mathbb{E}_{\mathbb{Q}}[(\frac{1}{P(T_{i-1},T_i)} - 1)\frac{P(T_{i-1},T_i)}{B(T_{i-1})}||\mathcal{F}_t]$$
$$= \mathbb{E}_{\mathbb{Q}}[(1 - P(T_{i-1},T_i))\frac{1}{B(T_{i-1})}||\mathcal{F}_t]$$
$$= \frac{P(t,T_{i-1})}{B(t)} - \frac{P(t,T_i)}{B(t)}.$$

Summing up over all i yields the simple formula

$$p(t) = P(t,T_n) + \sum_{i=1}^{n}(P(t,T_{i-1}) - P(t,T_i)) = P(t,T_0) \tag{8.37}$$

for the time t value of the floating rate note; nothing else than the price of a bond with maturity T_0.

Interest rate swaps

An *interest rate swap* is a scheme where a payment stream at a fixed rate of interest is exchanged for a payment stream at a floating rate (or vice versa). The first product of this kind we look at is the *payer interest rate swap* settled in arrears. This product is specified by future dates $T_0 < T_1 < \cdots < T_n$ with fixed $T_i - T_{i-1} = \delta$, (T_n is the maturity of the swap), a fixed rate K, a nominal value N. Cash flows (pay offs) take place only at the coupon dates $T_1, \ldots, T_n$. At each T_i, the holder of the contract: pays a fixed amount $K\delta N$, and receives the floating amount $F(T_{i-1},T_i)\delta N$. The resulting net cash flow at T_i is thus $(F(T_{i-1},T_i) - K)\delta N$. The value at $t \le T_0$ of this cash flow can be computed by the same reasoning that led to (8.37) and becomes

$$N\big(P(t,T_{i-1}) - P(t,T_i) - K\delta P(t,T_i)\big).$$

The total value $\Pi_p(t)$ of the swap at time $t \le T_0$ is thus the sum of these quantities,

$$\Pi_p(t) = N\big(P(t,T_0) - P(t,T_n)\big) - K\delta N \sum_{i=1}^{n} P(t,T_i). \tag{8.38}$$

A *receiver interest rate swap* settled in arrears is obtained by changing the sign of the cash flows at times $T_1, \ldots, T_n$. Its value at time $t \le T_0$ is thus $\Pi_r(t) = -\Pi_p(t)$. There should be a 'fair' fixed rate K for which this swap can be traded. The forward swap rate $R_{\text{swap}}(t)$ at time $t \le T_0$ is the fixed rate K above which gives $\Pi_r(t) = \Pi_p(t)$, and hence this value has to be zero. $R_{\text{swap}}(t)$ is then easily computed by solving the equation $\Pi_p(t) = 0$, resulting in

$$R_{\text{swap}}(t) = \frac{P(t,T_0) - P(t,T_n)}{\delta \sum_{i=1}^{n} P(t,T_i)}.$$

Caps and Floors

A *caplet* with reset date T and settlement date $T+\delta$ pays the holder the difference between a simple market rate $F(T,T+\delta)$ and the strike rate K, if positive. It is thus like a European option and the cash flow of such a caplet at time $T+\delta$ is thus $\delta(F(T,T+\delta)-K)^+$.

A *cap* is a strip of caplets. It thus consists of a number of future dates $T_0 < T_1 < \cdots < T_n$ with all $T_i - T_{i-1} = \delta$ (T_n is the maturity of the cap), and a cap rate K. Cash flows take place at the dates $T_1,\ldots,T_n$. At T_i the holder of the cap receives $\delta(F(T_{i-1},T_i)-K)^+$. Let $t \leq T_0$. We write $\mathrm{Cpl}(t;T_{i-1},T_i)$, $i=1,\ldots,n$, for the time t price of the i-th caplet with reset date T_{i-1} and settlement date T_i, and $\mathrm{Cp}(t) = \sum_{i=1}^n \mathrm{Cpl}(t;T_{i-1},T_i)$ for the time t price of the cap. A cap gives the holder a protection against rising interest rates. It guarantees that the interest to be paid on a floating rate loan never exceeds the predetermined cap rate K.

A *floor* is the converse to a cap. It protects against low rates. A floor is a strip of *floorlets*, the cash flow of which is – with the same notation as above – at time T_i equal to $\delta(K - F(T_{i-1},T_i))^+$. Write $\mathrm{Fll}(t;T_{i-1},T_i)$ for the price of the i-th floorlet and $\mathrm{Fl}(t) = \sum_{i=1}^n \mathrm{Fll}(t;T_{i-1},T_i)$ for the price of the floor. It is market practice to price a cap/floor according to *Black's formula*, which is of the same type as pricing formulas as e.g. in Proposition 8.10. We don't present the result, nor a theoretical justification of it as this would involve models for the forward rates that we don't treat.

We close this section by mentioning one final related derivative. It is the European payer (or receiver) *swaption* with strike rate K, an option giving the right to enter a payer (or receiver) swap with fixed rate K at a given future date, the swaption maturity. Pricing formulas can be derived, but we chose not to do so.

8.7 Exercises

8.1 Consider the Vasiček model (8.19), let $t > s \geq 0$.

(a) Show that

$$r(t) = e^{b_1(t-s)}r(s) + \frac{b_0}{b_1}(e^{b_1(t-s)}-1) + \sigma e^{b_1 t}\int_s^t e^{-b_1 u}\,\mathrm{d}W(u).$$

(b) Show that conditional on $\mathcal{F}_s$, the distribution of $r(t)$ is normal with mean $e^{b_1(t-s)}r(s) + \frac{b_0}{b_1}(e^{b_1(t-s)}-1)$ and variance $\frac{\sigma^2}{2b_1}(e^{2b_1(t-s)}-1)$.

(c) Suppose that $b_1 < 0$ and that $r(0)$ has a normal distribution with mean $-\frac{b_0}{b_1}$ and variance $-\frac{\sigma^2}{2b_1}$. What is the distribution of $r(t)$?

8.2 Consider the Vasiček model (8.19) for the short rate.

(a) Show that the Riccati equations take the following form.

$$\dot{A}(t,T) = \frac{1}{2}\sigma^2 B(t,T)^2 - b_0 B(t,T),\ A(T,T) = 0,$$
$$\dot{B}(t,T) = -b_1 B(t,T) - 1,\ B(T,T) = 0.$$

(b) Solve the Riccati equations and show that the solutions are given by

$$A(t,T) = \frac{\sigma^2}{4b_1^3}(4e^{b_1(T-t)} - e^{2b_1(T-t)} - 2b_1(T-t) - 3)$$

$$
\begin{aligned}
&\quad + \frac{b_0}{b_1^2}(e^{b_1(T-t)} - 1 - b_1(T-t)), \\
B(t,T) &= \frac{1}{b_1}(e^{b_1(T-t)} - 1).
\end{aligned}
$$

8.3 Consider the CIR model (8.20) for the short rate.

(a) Show that the Riccati equations take the following form.

$$
\begin{aligned}
\dot{A}(t,T) &= -b_0 B(t,T),\ A(T,T) = 0, \\
\dot{B}(t,T) &= \frac{1}{2}\sigma^2 B(t,T)^2 - b_1 B(t,T) - 1,\ B(T,T) = 0.
\end{aligned}
$$

(b) Solve the Riccati equations and show that the solutions are given by

$$
\begin{aligned}
A(t,T) &= -\frac{2b_0}{\sigma^2}\log\left(\frac{2\gamma e^{(\gamma-b_1)(T-t)/2}}{(\gamma-b_1)(e^{\gamma(T-t)}-1)+2\gamma}\right) \\
&\quad + \frac{b_0}{b_1^2}(e^{b_1(T-t)} - 1 - b_1(T-t)), \\
B(t,T) &= \frac{2(e^{\gamma(T-t)}-1)}{(\gamma-b_1)(e^{\gamma(T-t)}-1)+2\gamma},
\end{aligned}
$$

where $\gamma = \sqrt{b_1^2 + 2\sigma^2}$.

8.4 Assume that the bond prices $P(t,T)$ are as in (8.8) and r satisfies (8.7).

(a) Show that $P(t,T)$ satisfies

$$
\mathrm{d}P(t,T) = P(t,T)r(t)\,\mathrm{d}t + P(t,T)\sigma(t;T)\,\mathrm{d}W^{\mathbb{Q}}(t),
$$

and give an expression for $\sigma(t;T)$. Specialize to the case where F is affine, so it satisfies (8.13), to give a more explicit expression for $\sigma(t;T)$.

(b) Let $\bar{P}(t,T) = B(t)^{-1}P(t,T)$. Show that

$$
\mathrm{d}\bar{P}(t,T) = \bar{P}(t,T)\sigma(t;T)\,\mathrm{d}W^{\mathbb{Q}}(t).
$$

There is no drift (no $\mathrm{d}t$ term) in the equation for $\bar{P}(t,T)$, which is not surprising. Explain.

8.5 Prove Proposition 8.5.

8.6 Recall that $\mathbb{Q}^T$ and $\mathbb{Q}^S$ are two forward measures connected to bonds maturing at times T and S respectively.

(a) Show that for $t \le T \le S$ one has

$$
\frac{\mathrm{d}\mathbb{Q}^S}{\mathrm{d}\mathbb{Q}^T}\Big|_{\mathcal{F}_t} = \exp\Big(\int_0^t \sigma^{S,T}(u)\,\mathrm{d}W^T(u) - \frac{1}{2}\int_0^t \sigma^{S,T}(u)^2\,\mathrm{d}u\Big),
$$

where $\sigma^{S,T}(u) = \sigma(u;S) - \sigma(u;T)$ and W^T the usual Brownian motion under the forward measure $\mathbb{Q}^T$ as in (8.24). *Hint:* $\frac{\mathrm{d}\mathbb{Q}^S}{\mathrm{d}\mathbb{Q}^T} = \frac{\mathrm{d}\mathbb{Q}^S}{\mathrm{d}\mathbb{Q}} / \frac{\mathrm{d}\mathbb{Q}^T}{\mathrm{d}\mathbb{Q}}$.

(b) Prove the formula for $p(t)$ in Equation (8.32).

8.7 Show that $\mathbb{E}_{\mathbb{Q}}R(T) = 1$ so that $\mathbb{R}$ is indeed a probability measure and that Equation (8.25) holds.

8.8 Show that Equation (8.27) holds true and express $\tilde{\sigma}(t)$ in $\sigma(t)$ and $\sigma(t;T)$.

8.9 Show the parity relation $\mathrm{Cp}(t) - \mathrm{Fl}(t) = \Pi_p(t)$, where $\Pi_p(t)$ is the value at t of a payer swap with rate K, nominal value one and the same tenor structure as the cap and floor. See Equation (8.38) with $N = 1$.

A Some results in Probability and in Analysis

A.1 Bare essentials of probability

The reader is supposed to be familiar with the concept of probability space. Here is a brief recap. We mostly denote such as space by $(\Omega, \mathcal{F}, \mathbb{P})$. Here Ω is a non-empty set, $\mathcal{F}$ a σ-algebra on it (elements of σ are called events), and $\mathbb{P} : \mathcal{F} \to [0, 1]$ a probability measure. An event F happens almost surely, abbreviated a.s., if $\mathbb{P}(F) = 1$. On $\mathbb{R}$ we usually work with the Borel σ-algebra, the smallest σ-algebra that contains the open sets in $\mathbb{R}$. The notion of Borel σ-algebra extends to arbitrary topological spaces.

A function $X : \Omega \to \mathbb{R}$ is called a random variable (or $\mathcal{F}$-measurable) if the inverse images $X^{-1}[B]$ belong to $\mathcal{F}$ for all $B \in \mathcal{B}$. Commonly we use the notation $\{X \in B\}$ for $X^{-1}[B]$. Let $\sigma(X) = \{X^{-1}[B] : B \in \mathcal{B}\}$. It follows that X is a random variable iff $\sigma(X) \subset \mathcal{F}$. The distribution of a random variable X, denoted $\mathbb{P}^X$, is the probability measure on $\mathcal{B}$ defined by $\mathbb{P}^X(B) = \mathbb{P}(X \in B)$, for $B \in \mathcal{B}$.

A function $f : \mathbb{R} \to \mathbb{R}$ is called Borel-measurable if $f^{-1}[B]$ belong to $\mathcal{B}$ for all $B \in \mathcal{B}$. Note that the concept of measurability depends on the σ-algebras. A continuous function f is Borel-measurable. A number of operations on random variables yield new random variables. See Exercise A.22 for some standard examples.

Expectation of random variables, denoted $\mathbb{E}X$, are by definition Lebesgue integrals $\int_\Omega X \, \mathrm{d}\mathbb{P}$, well defined if $\int_\Omega |X| \, \mathrm{d}\mathbb{P} < \infty$. If X has a density f, this reduces to $\int_\mathbb{R} x f(x) \, \mathrm{d}x$. For discrete random variables the expectation is a sum. In general, for Borel-measurable h, the expectation $\mathbb{E}h(X)$, if well defined, is equal to the Lebesgue integral $\int_\mathbb{R} h \, \mathrm{d}\mathbb{P}^X$. In the density case the latter integral equals $\int_\mathbb{R} h(x) f(x) \, \mathrm{d}x$.

A.2 Normal random variables and vectors

If X is a normally distributed real random variable with mean μ and variance $\sigma^2 > 0$, then by definition it has density p given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}).$$

An extension of this formula holds for the case where $X = (X_1, \ldots, X_n)^\top$ is a random vector in $\mathbb{R}^n$. If $\mathbb{E}X = \mu \in \mathbb{R}^n$ and X has covariance matrix $\mathbb{C}\mathrm{ov}(X) = \Sigma$, which is assumed to be strictly positive definite, then by definition X has a density on $\mathbb{R}^n$ given by

$$p(x) = \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{1/2}} \exp(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu)). \tag{A.1}$$

A random vector X with such a density is said to have a nondegenerate multivariate normal distribution. If the matrix Σ is singular, then it is still possible to speak of multivariate normal distributions, although a density in this case does not exist.

An important property of multivariate normal random vectors is that affine transformations are again multivariate normal. If $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ and if A has full row rank, then $Y = AX + b$ is again multivariate normal with $\mathbb{E}Y = A\mu + b$ and $\mathbb{C}\text{ov}(Y) = A\Sigma A^\top$. Note that $\mathbb{C}\text{ov}(Y)$ is again strictly positive definite, so that Y also possesses a density (on $\mathbb{R}^m$). It now follows that every subvector of X is nondegenerate normal.

Another important property of multivariate normally distributed random vectors is that *uncorrelated components are independent.* Specifically, if X^1 and X^2 are subvectors of X such that $\mathbb{C}\text{ov}(X^1, X^2) = 0$ (zero matrix), then X^1 and X^2 are independent random vectors.

Here is a consequence of this last property. Let X^1 be a subvector of X with covariance matrix Σ_1. Then we have $Z := X^2 - \mathbb{C}\text{ov}(X^2, X^1)\Sigma_1^{-1}X^1$ is *independent* (see Exercise A.7) of X_1. Hence the conditional expectation (see Section A.6 for a definition and some properties) $\mathbb{E}[X^2|X^1]$ is equal to $\mathbb{C}\text{ov}(X^2, X^1)\Sigma_1^{-1}X^1 + \mathbb{E}Z = \mathbb{E}X^2 + \mathbb{C}\text{ov}(X^2, X^1)\Sigma_1^{-1}(X^1 - \mathbb{E}X^1)$.

If $\mathbb{E}X^1$ and $\mathbb{E}X^2$ are zero, then $\mathbb{E}[X^2|X^1] = \mathbb{C}\text{ov}(X^2, X^1)\Sigma_1^{-1}X^1$ and hence we have the decomposition $X^2 = \mathbb{E}[X^2|X^1] + Z$, with $\mathbb{E}[X^2|X^1]$ and Z independent random vectors.

A.3 Characteristic functions

Let X be a random vector in $\mathbb{R}^n$ ($n \geq 1$). The characteristic function of X is $\lambda \mapsto \mathbb{E}\exp(\mathrm{i}\lambda^\top X)$. We immediately see that characteristic functions only depend on the distribution of X: if X and Y have the same distribution, then their characteristic functions coincide. But for different distributions, the characteristic functions are different as well. So characteristic functions correspond uniquely to the underlying distributions and the terminology 'characteristic' is completely justified. Characteristic functions are very useful in probability. The best known example of their use is the Central limit theorem, see Section A.5.

Consider as an example a random variable X that has a standard normal distribution, so it has density $p(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$. Let ϕ be its characteristic function. Then

$$\phi(\lambda) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(\mathrm{i}\lambda x - \frac{1}{2}x^2)\,\mathrm{d}x.$$

With methods from complex integration it is not difficult to show that $\phi(\lambda) = \exp(-\frac{1}{2}\lambda^2)$. An alternative method is presented in Exercise A.4. Knowing this, it is easy to show that the characteristic function of $N(\mu, \sigma^2)$ distributed random variable is given by $\phi(\lambda) = \exp(\mathrm{i}\mu\lambda - \frac{1}{2}\sigma^2\lambda^2)$. But then it is almost immediate (Exercise A.6) that for this case

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathrm{i}\lambda x)\phi(\lambda)\,\mathrm{d}\lambda. \tag{A.2}$$

It can be shown that relation (A.2) holds more general. The density can be found back from the characteristic function by an integral transformation (Fourier transform) if the characteristic function belongs to $\mathcal{L}^1(\mathbb{R})$.

Other properties that we use are the following. X and Y are independent random vectors iff $\mathbb{E}\exp(\mathrm{i}\lambda^\top X + \mathrm{i}\mu^\top Y) = \mathbb{E}\exp(\mathrm{i}\lambda^\top X)\mathbb{E}\exp(\mathrm{i}\mu^\top Y)$. Here is an example: if X has a $N(0, \sigma^2)$ distribution and Y has a $N(0, \tau^2)$ distribution, then X and Y are independent iff $\mathbb{E}\exp(\mathrm{i}\lambda X + \mathrm{i}\mu Y) = \exp(-\frac{1}{2}(\sigma^2\lambda^2 + \tau^2\mu^2))$.

Another property concerns derivatives of characteristic functions. Let X be a real random variable and let ϕ be its characteristic function. If $\mathbb{E}|X|^k < \infty$, then ϕ is k times differentiable and $\phi^{(k)}(0) = \mathrm{i}^k \mathbb{E}X^k$.

A.4 Modes of convergence

In this section we briefly treat various convergence concepts for random variables. Let us start by defining them. First a notational convention. If X is a random variable, then F (or sometimes F_X) denotes its distribution function. If there are indexed random variables like X_n, then their distribution functions are denoted F_n (or sometimes F_{X_n}).

Definition A.1 Consider a sequence of random variables $X_1, X_2, \dots$ and another random variable X. We say that

(i) X_n converges in probability to X if for all $\varepsilon > 0$ one has $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

(ii) X_n converges to X almost surely (a.s.), if $\mathbb{P}(\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$.

(iii) X_n converges to X in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ if $\lim_{n\to\infty} \mathbb{E}|X_n - X| = 0$.

(iv) X_n converges to X in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ if $\lim_{n\to\infty} \mathbb{E}|X_n - X|^2 = 0$.

(v) X_n converges to X in distribution if $\lim_{n\to\infty} F_n(x) = F(x)$ in all $x \in \mathbb{R}$ at which F is continuous.

Here are some relations between the different modes of convergence.

Proposition A.2 *We have the following implications.*

(i) If X_n converges to X almost surely, then also in probability.

(ii) If X_n converges to X in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, then also in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

(iii) If X_n converges to X in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, then also $\mathbb{E}X_n \to \mathbb{E}X$.

(iv) If X_n converges to X in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, then also in probability.

(v) If the sequence X_n is bounded and X_n converges to X in probability, then also in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

(vi) If X_n converges to X in probability, then there is a (X_{n_k}) such that X_{n_k} converges to X almost surely.

(vii) If X_n converges to X in probability, then X_n also converges to X in distribution.

(viii) Denote by $\stackrel{*}{\to}$ any of the modes of convergence in (i)-(iv) of Definition A.1. Then $X_n \stackrel{*}{\to} X$ and $Y_n \stackrel{*}{\to} Y$ imply $X_n + Y_n \stackrel{*}{\to} X + Y$.

Proof We only prove (vi), the other assertions we leave as an exercise. From the definition of convergence in probability we have that for each k the probability $\mathbb{P}(|X_n - X| > \frac{1}{k})$ tends to zero, so eventually it will become less than 2^{-k}. Hence there is n_k such that $\mathbb{P}(E_k) < 2^{-k}$, where $E_k = \{|X_{n_k} - X| > \frac{1}{k}$. The Borel-Cantelli lemma implies that $\mathbb{P}(\liminf_{k\to\infty} E_k^c) = 1$. But for $\omega \in \liminf_{k\to\infty} E_k^c$ we can find some $N \in \mathbb{N}$ such that for all $k > N$ we have $|X_{n_k}(\omega) - X(\omega)| < \frac{1}{k}$, hence the $X_{n_k}(\omega)$ converge to $X(\omega)$. □

We also need some results that say, when random variables X_n that converge to a limit X almost surely, also converge in $\mathcal{L}^1$. Here we have some answers, the basic convergence theorems in measure theory.

Theorem A.3 *Let $X_n \to X$ almost surely.*

(i) (Monotone convergence) If the X_n form a nonnegative increasing sequence with limit X, then $\mathbb{E}X_n \uparrow \mathbb{E}X \leq \infty$.

(ii) If the X_n are bounded by a random variable Y, that is such that $\mathbb{E}|Y| < \infty$, then X_n converges to X in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. A special case arises when Y can be taken as a (nonrandom) constant.

(iii) (Fatou's lemma) Without any of the two above conditions we still have $\liminf \mathbb{E}X_n \geq \mathbb{E}X$, if the X_n are nonnegative.

Remark A.4 Since expectations are by definition integrals w.r.t. the underlying probability measure, we have similar results for integrals of functions w.r.t. some measure.

A.5 Central limit theorem

The central limit theorem below tells us that the distribution of a standardized sum of *iid* random variables with finite second moments converges to the standard normal distribution function. For random variables with the corresponding distributions one speaks of convergence in distribution. Let us make these notions precise.

Definition A.5 Consider a sequence $Z_1, Z_2, \ldots$ of (real valued) random variables, another random variable Z and denote by F_n the distribution function of Z_n and by F the distribution function of Z. We say that the F_n *weakly converge* to F, if $F_n(x)$ converges to $F(x)$ in all x where F is continuous. In this case we also say that the Z_n *converge to Z in distribution.*

In probability theory one also considers *weak convergence* of probability measures. For random variables this concept takes the following form. We say that the distributions of the Z_n *weakly converge* to the distribution of a random variable Z if for all (real or complex valued) bounded continuous functions f on $\mathbb{R}$ one has the convergence $\mathbb{E}f(Z_n) \to \mathbb{E}f(Z)$. One of the versions of the *portmanteau theorem* states that convergence in distribution of the Z_n to Z is equivalent to the weak convergence of the distributions of Z_n to that of Z. We use it in the following form.

Theorem A.6 *A sequence $Z_1, Z_2, \ldots$ of random variables converges in distribution to a random variable Z iff for all bounded continuous functions h the convergence $\mathbb{E}h(Z_n) \to \mathbb{E}h(Z)$ takes place.*

Proof Assume convergence in distribution of the Z_n. We have to show convergence of the distribution functions of the Z_n to that of Z in all point x where the latter is continuous. So, let $x \in \mathbb{R}$. First note that $\mathbb{P}(Z_n \leq x) = \mathbb{E}\mathbf{1}_{(-\infty,x]}(Z_n)$. Let $\varepsilon > 0$ be given and let h^ε be a bounded continuous function that has value one on $(-\infty, x]$, values in $(0,1)$ on $(x, x+\varepsilon)$ and zero on $[x+\varepsilon, \infty)$. Let $h_\varepsilon(y) = h^\varepsilon(y+\varepsilon)$. Note that $h_\varepsilon \leq \mathbf{1}_{(-\infty,x]} \leq h^\varepsilon$ and hence $\mathbb{E}h_\varepsilon(Z_n) \leq \mathbb{P}(Z_n \leq x) \leq \mathbb{E}h^\varepsilon(Z_n)$. The two extreme members of this inequality converge to $\mathbb{E}h_\varepsilon(Z)$

and $\mathbb{E}h^\varepsilon(Z)$. Use that $\mathbb{P}(Z \leq x - \varepsilon) \leq \mathbb{E}h_\varepsilon(Z)$ and $\mathbb{E}h_\varepsilon(Z) \leq \mathbb{P}(Z \leq x + \varepsilon)$ to see that with $\varepsilon \to 0$ we obtain $\limsup_{n\to\infty} \mathbb{P}(Z_n \leq x) \leq \mathbb{P}(Z \leq x)$ and $\liminf_{n\to\infty} \mathbb{P}(Z_n \leq x) \geq \mathbb{P}(Z < x)$. The result follows.

The proof of the converse statement is as follows. Suppose first that h is a bounded continuous function with compact support contained in an interval $(-A, A]$ (with $A > 0$). Let $\varepsilon > 0$ be given. Since h is uniformly continuous on $(-A, A]$ there exists $N \in \mathbb{N}$ such that $|h(x) - h(y)| < \varepsilon$ as soon as $|x - y| < 1/N$. Divide $(-A, A]$ into intervals $I_k^N = (L_k^N, R_k^N]$ of equal length $1/N$. Let $\bar{h}_k^N = \sup\{h(x) : x \in I_k^N\}$ and observe that $h(x) \geq \bar{h}_k^N - \varepsilon$ for $x \in I_k^N$. Then we have $\mathbb{E}h(Z_n) = \sum_k \mathbb{E}h(Z_n)\mathbf{1}_{I_k^N}(Z_n) \leq \sum_k \bar{h}_k^N \mathbb{P}(L_k^N < Z_n \leq R_k^N) \to \sum_k \bar{h}_k^N \mathbb{P}(L_k^N < Z \leq R_k^N)$ as $n \to \infty$. Hence we have $\limsup_n \mathbb{E}h(Z_n) \leq \sum_k \bar{h}_k^N \mathbb{P}(L_k^N < Z \leq R_k^N)$.

Since we also have $\mathbb{E}h(Z) \geq \sum_k \bar{h}_k^N \mathbb{P}(L_k^N < Z \leq R_k^N) - \varepsilon$, we arrive at $\limsup_n \mathbb{E}h(Z_n) \leq \mathbb{E}h(Z) + \varepsilon$. Replacing h with $-h$ we get the companion inequality $\liminf_n \mathbb{E}h(Z_n) \geq \sum_k \bar{h}_k^N \mathbb{P}(L_k^N < Z \leq R_k^N) - \varepsilon$. Let $\varepsilon \to 0$ to finish the proof for continuous functions with compact support. Use this result to prove the assertion for arbitrary bounded continuous h (Exercise A.11). □

Here is a useful result, that show that convergence in distribution is weaker than convergence in probability (and then also weaker than almost sure convergence).

Proposition A.7 *Suppose X_n converges to X in probability. Then we also have convergence of X_n to X in distribution.*

Proof Let $\varepsilon > 0$ and consider

$$\begin{aligned}\mathbb{P}(X_n \leq x) &= \mathbb{P}(X_n \leq x, |X_n - X| \leq \varepsilon) + \mathbb{P}(X_n \leq x, |X_n - X| > \varepsilon) \\ &\leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon).\end{aligned}$$

It follows that $\limsup \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \varepsilon)$ for all $\varepsilon > 0$. By right continuity of $x \mapsto \mathbb{P}(X \leq x)$ we obtain $\limsup \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x)$.

Similarly, exchange X and X_n and replace x with $x - \varepsilon$, one has

$$\mathbb{P}(X \leq x - \varepsilon) \leq \mathbb{P}(X_n \leq x) + \mathbb{P}(|X_n - X| > \varepsilon).$$

It then follows that $\mathbb{P}(X \leq x - \varepsilon) \leq \liminf \mathbb{P}(X_n \leq x)$. If x is a continuity point of $x \mapsto \mathbb{P}(X \leq x)$, then we obtain for $\varepsilon \to 0$ the inequality $\mathbb{P}(X \leq x) \leq \liminf \mathbb{P}(X_n \leq x)$. Together with the inequality for the limsup, this proves the assertion. □

The most important example of weak convergence is the content of the Central limit theorem:

Theorem A.8 *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with mean μ and variance $\sigma^2 < \infty$. Then for all $x \in \mathbb{R}$ we have*

$$\mathbb{P}\Big(\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} \leq x\Big) \to \Phi(x),$$

as n tends to ∞, where Φ is the distribution function of the standard normal distribution.

To connect the statement of Theorem A.8 with the introduction of this section it is sufficient to take $Z_n = \sum_{i=1}^n (X_i - \mu)/\sigma\sqrt{n}$, since Φ is continuous in any element of $\mathbb{R}$.

The standard proof of the Central limit theorem uses characteristic functions. In view of the property that they correspond uniquely to distributions, this sounds reasonable. However one also needs the *theorem* that distribution functions F_n converge weakly to a distribution function F iff the characteristic functions of the F_n converge pointwise to the characteristic function of F. (Note that one implication of this theorem is easy in view of the portmanteau theorem, the other implication is harder to prove). Let us take this theorem for granted and sketch the proof of Theorem A.8.

Proof of Theorem A.8. Without loss of generality we assume that $\mathbb{E}X_k = 0$. We start off with some facts from analysis.

Recall that for complex $z_n \to z$ we have the convergence $(1 + \frac{z_n}{n})^n \to e^z$. Let $R_n(x) = e^{\mathrm{i}x} - \sum_{k=0}^n (\mathrm{i}x)^k/k!$ for $x \in \mathbb{R}$. Then (this is part of Exercise A.9)

$$|R_n(x)| \leq \min\{\frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!}\}. \tag{A.3}$$

Let ϕ be the characteristic function of the X_k. It then follows (this is another part of Exercise A.9) that

$$|\phi(\lambda) - 1 + \frac{1}{2}\sigma^2\lambda^2| \leq \mathbb{E}|R_2(\lambda X)| \leq \lambda^2\mathbb{E}\left(\min\{|X|^2, \frac{1}{6}|\lambda||X|^3\}\right). \tag{A.4}$$

Letting $\lambda \to 0$, we obtain from (A.4) by application of the dominated convergence theorem that $\phi(\lambda) = 1 - \frac{1}{2}\sigma^2\lambda^2(1 + o(1))$ for $\lambda \to 0$.

Now we are ready to prove the assertion of the theorem. Let $Z_n = \frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}}$. It has characteristic function $\phi_{Z_n}(\lambda) = \phi(\frac{\lambda}{\sigma\sqrt{n}})^n$. Using the expansion above for ϕ, we can write $\phi_{Z_n}(\lambda) = \left(1 - \frac{\lambda^2}{2n}(1 + o(1))\right)^n$ which converges to $\exp(-\frac{1}{2}\lambda^2)$, the characteristic function of the $N(0,1)$ distribution. In view of the remarks preceding the theorem, this is exactly what we had to prove. □

Sometimes one has to work with a central limit theorem for *arrays*. Let us state what this means. We consider for each $n \in \mathbb{N}$ integers k_n and random variables $X_{n,k}$, with $k = 1, \ldots, k_n$. The family $\{X_{n,k} : n \in \mathbb{N}, k = 1, \ldots, k_n\}$ is called a triangular array, a reasonable name if k_n is increasing in n.

Theorem A.9 *Consider a triangular array of random variables $\{X_{n,k} : n \in \mathbb{N}, k = 1, \ldots, k_n\}$ with $\lim_{n\to\infty} k_n = \infty$. Assume that for each n the random variables $X_{n,k}$ for $k = 1, \ldots, k_n$ are independent and identically distributed. Define $S_n = \sum_{k=1}^{k_n} X_{n,k}$. Assume moreover that $\mathbb{E}X_{n,k} = 0$ for all n, k and that $\lim_{n\to\infty} \mathbb{E}S_n^2 = \sigma^2$ with $\sigma^2 < \infty$. Then we have weak convergence of the distributions of the S_n to the $N(0, \sigma^2)$ distribution.*

Proof Exercise A.12. □

We close this section that all notions and results above can be completely taken over to the case where one deals with finite dimensional random vectors. The situation drastically changes if one considers random vectors in infinite dimensional spaces, but this is of no concern for this course.

A.6 Conditional expectations

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given. Consider two random variables or vectors X and Y that both assume finitely many values in sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Assume that $\mathbb{P}(Y = y) > 0$ for all $y \in \mathcal{Y}$. Then the conditional probabilities $\mathbb{P}(X = x|Y = y)$ are all well defined as well as for any function f on $\mathcal{X}$ the *conditional expectation* $\mathbb{E}[f(X)|Y = y] := \sum_{x\in\mathcal{X}} f(x)\mathbb{P}(X = x|Y = y)$.

Consider the function $\hat{f}$ defined by $\hat{f}(y) = \mathbb{E}[f(X)|Y = y]$. With the aid of $\hat{f}$ we define the *conditional expectation of* $f(X)$ *given* Y, denoted by $\mathbb{E}[f(X)|Y]$, as $\hat{f}(Y)$. A simple calculation suffices to check the relation

$$\mathbb{E}[f(X)|Y] = \sum_y \hat{f}(y)\mathbf{1}_{\{Y=y\}} = \sum_y \frac{\mathbb{E}(f(X)\mathbf{1}_{\{Y=y\}})}{\mathbb{P}(Y = y)}\mathbf{1}_{\{Y=y\}}. \tag{A.5}$$

Elementary properties of conditional expectation like linearity are in this case easy to prove. Other properties are equally fundamental and easy to prove. They are listed in Proposition A.10. Replacing X above by (X, Y) we can also define conditional expectations like $\mathbb{E}[f(X, Y)|Y]$.

Proposition A.10 *The following properties hold for conditional expectations.*

(i) If X and Y are independent, then $\mathbb{E}[f(X)|Y] = \mathbb{E}[f(X)]$, the unconditional expectation.

(ii) Let $f(X, Y)$ be the product $f_1(X)f_2(Y)$. Then

$$\mathbb{E}[f_1(X)f_2(Y)|Y] = f_2(Y)\mathbb{E}[f_1(X)|Y].$$

(iii) If X and Y are independent, then

$$\mathbb{E}[f(X, Y)|Y] = \sum_{x\in\mathcal{X}} f(x, Y)\mathbb{P}(X = x).$$

Proof Exercise A.13. □

The random variable Y induces a sub-σ-algebra of $\mathcal{F}$ on Ω, call it $\mathcal{G}$, which is generated by the sets $\{Y = y\}$ and that $\mathcal{G} = \sigma(Y)$. Note that the sets $\{Y = y\}$ constitute a partition of Ω and hence every set in $\mathcal{G}$ is a finite union of some $\{Y = y\}$.

Consider again the function $\hat{f}$ above. It defines another function on Ω, F say, according to $F(\omega) = \hat{f}(y)$ on the set $\{\omega : Y(\omega) = y\}$, hence $F(\omega) = \hat{f}(Y(\omega))$, in short $F = \hat{f}(Y)$. In this way we can identify the conditional expectation $\hat{f}(Y)$ with the random variable F. Note that F is $\mathcal{G}$-measurable, it is constant on the sets $\{Y = y\}$. Moreover, one easily verifies (Exercise A.14) that

$$\mathbb{E}(F\mathbf{1}_{\{Y=y\}}) = \mathbb{E}(f(X)\mathbf{1}_{\{Y=y\}}). \tag{A.6}$$

In words, the expectation of $f(X)$ and its conditional expectation over the sets $\{Y = y\}$ are the same, and therefore we will also have

$$\mathbb{E}(F\mathbf{1}_G) = \mathbb{E}(f(X)\mathbf{1}_G) \text{ for every set } G \in \mathcal{G}. \tag{A.7}$$

Conditional expectation can also be defined for random variables that are not discrete. To that end we proceed directly to the general definition as it is used in modern probability theory. That it is possible to define conditional expectation as we will do below is a consequence of the Radon-Nikodym theorem in measure theory. The definition is motivated by (A.7).

Definition A.11 If X is a random variable with $\mathbb{E}|X| < \infty$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$, then the conditional expectation of X given $\mathcal{G}$ is any $\mathcal{G}$-measurable random variable $\hat{X}$ with the property that $\mathbb{E}X\mathbf{1}_G = \mathbb{E}\hat{X}\mathbf{1}_G$ for all $G \in \mathcal{G}$. We will use the notation $\mathbb{E}[X|\mathcal{G}]$ for any of the $\hat{X}$ above.

Note that the conditional expectation is not uniquely defined, but only up to almost sure equivalence, if $\hat{X}$ and $\hat{X}'$ both satisfy the requirements of Definition A.11 then $\mathbb{P}(\hat{X} = \hat{X}') = 1$, see Exercise A.23. Different $\hat{X}$ are called versions of the conditional expectation. Conditioning with respect to a random variable (vector) Y is obtained by taking $\mathcal{G}$ equal to the σ-algebra that Y induces on Ω. It is a theorem in probability theory that any version of the conditional expectation given Y can be represented by a (measurable) function of Y. We have seen this to be true at the beginning of this section. The most important properties of conditional expectation that we use in this course are collected in the following proposition.

Proposition A.12 *Let X be a random variable with $\mathbb{E}|X| < \infty$.*

(i) $\mathbb{E}(\mathbb{E}[X|\mathcal{G}]) = \mathbb{E}X$.

(ii) *If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$.*

(iii) *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ (iterated conditioning).*

(iv) *If Y is $\mathcal{G}$-measurable and $\mathbb{E}|XY| < \infty$, then $\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y$.*

(v) *If X is independent of $\mathcal{G}$, Y is $\mathcal{G}$-measurable and f is a measurable function on $\mathbb{R}^2$ with $\mathbb{E}|f(X,Y)| < \infty$, then $\mathbb{E}[f(X,Y)|\mathcal{G}] = \int f(x,Y)\,\mathbb{P}^X(dx)$, with $\mathbb{P}^X$ the distribution of X. Alternatively written, $\mathbb{E}[f(X,Y)|\mathcal{G}] = \hat{f}(Y)$, where $\hat{f}(y) = \mathbb{E}f(X,y)$.*

(vi) *The conditional expectation is a linear operator on the space of random variables with finite expectation.*

A special occurs of conditional expectation occurs if $\mathcal{G}$ is generated by a partition $\{G_1, \ldots, G_m\}$ of Ω with all $\mathbb{P}(G_i) > 0$. Then, completely analogous to (A.5),

$$\mathbb{E}[X|\mathcal{G}] = \sum_i \frac{\mathbb{E}(X\mathbf{1}_{G_i})}{\mathbb{P}(G_i)}\mathbf{1}_{G_i}. \tag{A.8}$$

Indeed, (A.5) is a special case of (A.8). Take $\mathcal{G} = \sigma(Y)$ and $G_i = \{Y = y_i\}$, where the y_i are the different values that Y assumes.

If the random variable X in Definition A.11 has the stronger integrability property $\mathbb{E}X^2 < \infty$, then one has (this is Exercise A.25)

$$\mathbb{E}(X - \hat{X})^2 \leq \mathbb{E}(X - Y)^2, \text{ for all } \mathcal{G}\text{-measurable } Y \text{ with } \mathbb{E}Y^2 < \infty. \tag{A.9}$$

Equation (A.9) offers a nice interpretation of conditional expectation as a projection, think of this! And with this interpretation in mind, property (iii) of Proposition A.12 should look familiar: it is analogous to repeated, iterated projection, first on a subspace, than on a smaller subspace, being equivalent to immediate projection on the smaller subspace.

A.7 Filtrations and Martingales

In probabilistic terms the loose term 'information' can be expressed by means of σ-algebras. One is given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A *filtration* in discrete time is by definition an increasing sequence of sub-σ-algebras of $\mathcal{F}$. So we have a family $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ where the $\mathcal{F}_n$ are sub-σ-algebras of $\mathcal{F}$ that satisfy $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \geq 0$.

Information often comes to us in the form of an observed sequence of random variables $X = (X_1, X_2, \ldots)$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$. For each n we then put $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, the smallest σ-algebra that makes $X_1, \ldots, X_n$ measurable functions on Ω. In this case one often speaks of the filtration generated by the sequence $X_1, X_2, \ldots$ and one sometimes write, to emphasize this relation, $\mathbb{F}^X$ and $\mathcal{F}_n^X$. In this situation the $\mathcal{F}_n$ are invariant under many transformations of the observations. For instance if we take $S_n = \sum_{k=1}^n X_k$ $(n \geq 0)$, then the filtrations $\mathbb{F}^X$ and $\mathbb{F}^S$ are the same: one knows all $S_1, \ldots, S_n$ iff one knows all $X_1, \ldots, X_n$.

Let $\mathbb{F}$ be a given filtration. A process X that is such that X_n is $\mathcal{F}_n$-measurable for all n is called *adapted* (to $\mathbb{F}$). Obviously, any process X is adapted to $\mathbb{F}^X$.

Martingales form one of the corner stones in modern probability. Let us first give a formal definition.

Definition A.13 A sequence M of random variables is said to be a martingale w.r.t. a filtration $\mathbb{F}$ if the following conditions are satisfied.

(i) For every $n \geq 0$ the random variable M_n is $\mathcal{F}_n$-measurable.

(ii) $\mathbb{E}|M_n| < \infty$ for all n.

(iii) $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ for all $n \geq 0$.

Property (iii) in Definition A.13 is equivalent to (this is Exercise A.26)

$$\mathbb{E}[M_m|\mathcal{F}_n] = M_n, \text{ for all } m \geq n. \tag{A.10}$$

Sometimes this conditional expectation property is explicitly formulated, using the definition of a conditional expectation as in Definition A.11, as

$$\mathbb{E}[M_m \mathbf{1}_F] = \mathbb{E}[M_n \mathbf{1}_F], \text{ for all } F \in \mathcal{F}_n.$$

Another equivalent formulation of property (iii) is

$$\mathbb{E}[\Delta M_{n+1}|\mathcal{F}_n] = 0 \text{ for all } n \geq 0.$$

The standard and easiest example of a martingale is the sum of independent random variables with zero expectation (this is Exercise A.15). But there is also a method to generate new martingales starting from a given one. To explain this method we have to introduce *predictable* sequences of random variables. A process $H = (H_1, H_2, \ldots)$ is said to be predictable w.r.t. a filtration $\mathbb{F}$ if for all $n \geq 1$ the random variable H_n is $\mathcal{F}_{n-1}$-measurable. If we want to include a variable H_0, we take $\mathcal{F}_{-1}$ the trivial σ-algebra, which implies that H_0 is a non-random constant. Suppose that M is martingale w.r.t. to a given filtration $\mathbb{F}$ and that H is a predictable process (w.r.t. the same filtration). Assume that the products $H_n(M_n - M_{n-1})$ have finite expectation (this surely happens if H is bounded). Define the process V by

$$V_n = \sum_{k=1}^{n} H_k \Delta M_k = \sum_{k=1}^{n} H_k (M_k - M_{k-1}), \tag{A.11}$$

(with $M_{-1} = 0$ if we also need $k = 0$). Then also V is a martingale w.r.t. $\mathbb{F}$ (this is Exercise A.17).

The relevant notions in continuous time are similarly defined. A filtration in continuous time is an increasing collection of sub-σ-algebras of $\mathcal{F}$. So we have a family $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,\infty)}$ where the $\mathcal{F}_t$ are sub-σ-algebras of $\mathcal{F}$ that satisfy $\mathcal{F}_s \subset \mathcal{F}_t$ for all $t \geq s \geq 0$. A process X in continuous time, so $X = (X(t),\, t \geq 0)$, is adapted to the filtration $\mathbb{F}$ if $X(t)$ is $\mathcal{F}_t$-measurable for all $t \geq 0$.

A martingale in continuous time is an adapted process M that satisfies the conditions of Definition A.13 '*with n replaced by t*'. More precisely, we have for the martingale property (iii) in Definition A.13 the analog of (A.10),

$$\mathbb{E}[M_t|\mathcal{F}_s] = M_s, \text{ for all } t > s,$$

equivalent to

$$\mathbb{E}[M_t \mathbf{1}_F] = \mathbb{E}[M_s \mathbf{1}_F], \text{ for all } t > s \text{ and } F \in \mathcal{F}_s. \tag{A.12}$$

Occasionally we need a convenient property of martingales that have the property that $\mathbb{E}M_t^2 < \infty$ for all t. In that case one has (Exercise A.27)

$$\mathbb{E}[(M_t - M_s)^2|\mathcal{F}_s] = \mathbb{E}[M_t^2 - M_s^2|\mathcal{F}_s], \text{ for all } t \geq s. \tag{A.13}$$

If any continuous time process X is given, then we will again use the notation $\mathbb{F}^X$ for the filtration it generates. So $\mathbb{F}^X = \{\mathcal{F}_t^X : t \in [0,\infty)\}$, where $\mathcal{F}_t^X$ is the smallest σ-algebra that makes all the $X(s)$ with $s \leq t$ random variables. The theory of continuous time filtrations and stochastic processes is much more subtle and difficult than the discrete time theory, but in this course we don't need these subtleties.

A.8 The heat equation: uniqueness of solutions

In Section 3 we announced that under regularity conditions the heat equation has a unique solution, see Theorem 3.3. In this section we state and prove a theorem on this. In the proof of this theorem we use a maximum principle that will be discussed first. For given $T > 0$ and real constants A and B, let $D \subset \mathbb{R}^2$ be the open rectangular domain $D = \{(t,x) : 0 < t < T, A < x < B\}$. We consider a function u that belongs to $C^{1,2}(D)$ and that is continuous on the boundary ∂D of D. We will also need the *parabolic boundary* $\partial_0 D = \{(t,x) \in \partial D : t = 0, \text{ or } x \in \{A,B\}\}$. Here is the parabolic maximum principle.

Proposition A.14 *If the function u above satisfies the heat inequality*

$$u_t(t,x) \leq \frac{1}{2} u_{xx}(t,x),\ (t,x) \in D, \tag{A.14}$$

then

$$\max_{(t,x)\in\bar{D}} u(t,x) = \max_{(t,x)\in\partial_0 D} u(t,x).$$

Proof The proof is easier if we would have a strict inequality in (A.14). So let us assume this for the time being. Let $\varepsilon \in (0,T)$ and consider in what follows the restricted domain $D_\varepsilon = \{(t,x) \in D : t < T - \varepsilon\}$. Suppose that u has a maximum on the closure $\bar{D}_\varepsilon$ of D_ε that is attained at an interior point (t_0, x_0) of D_ε. Then the function $t \mapsto u(t, x_0)$ is maximal at t_0 and hence $u_t(t_0, x_0) = 0$, whereas the function $x \mapsto u(t_0, x)$ is maximal in x_0 and hence $u_{xx}(t_0, x_0) \leq 0$. Hence, in (t_0, x_0) the strict inequality (A.14) would be violated. We conclude that the maximum of u on $\bar{D}_\varepsilon$ is attained on the boundary ∂D_ε.

Suppose now that u is maximal at $(T-\varepsilon, x_0) \in \partial D_\varepsilon \backslash \partial_0 D_\varepsilon$. As above we then have $u_{xx}(T-\varepsilon, x_0) \leq 0$, whereas we also conclude that $u_t(T-\varepsilon, x_0) \geq 0$ (would this be negative, than $t \mapsto u(t, x_0)$ would be decreasing in a neighborhood of $T-\varepsilon$ and thus we could find $t' < T-\varepsilon$ with $u(t', x_0) > u(T-\varepsilon, x_0)$ contradicting maximality at $(T-\varepsilon, x_0)$). This again violates (A.14). Hence we conclude

$$\max_{(t,x)\in\bar{D}_\varepsilon} u(t,x) = \max_{(t,x)\in\partial_0 D_\varepsilon} u(t,x). \tag{A.15}$$

Now we let $\varepsilon \downarrow 0$ and use uniform continuity of u on $\bar{D}$ to conclude that the left hand side of (A.15) increases to $\max_{(t,x)\in\bar{D}} u(t,x)$ and the right hand side to $\max_{(t,x)\in\partial_0 D} u(t,x)$. We proved the assertion for functions u for which we have strict inequality in (A.14).

The general case is as follows. Instead of considering u directly we consider u' defined by $u'(t,x) = u(t,x) - \delta t$, with $\delta > 0$. Then u' satisfies the strict inequality and we conclude in view of the forgoing that the assertion of the proposition is valid for u'. But we also have the two inequalities $\max_{(t,x)\in\bar{D}} u'(t,x) \geq \max_{(t,x)\in\bar{D}} u(t,x) - \delta T$ and $\max_{(t,x)\in\partial_0 D} u'(t,x) \leq \max_{(t,x)\in\partial_0 D} u(t,x)$. Hence, by $\delta \downarrow 0$, $\max_{(t,x)\in\bar{D}} u(t,x) \leq \max_{(t,x)\in\partial_0 D} u(t,x)$ and the result follows. □

The uniqueness theorem for solutions to the heat equation is a consequence of the following

Theorem A.15 *Let $u \in C^{1,2}((0,\infty)\times\mathbb{R})$ and continuous on $[0,\infty)\times\mathbb{R}$. Let there be nonnegative constants A and B such that $|u(t,x)| \leq A\exp(Bx^2)$ for all $(t,x) \in [0,\infty)\times\mathbb{R}$. If u satisfies the heat inequality $u_t \leq \frac{1}{2}u_{xx}$ on $(0,\infty)\times\mathbb{R}$ and if $u(x,0) \leq 0$ for all $x \in \mathbb{R}$, then $u(t,x) \leq 0$ for all $(t,x) \in [0,\infty)\times\mathbb{R}$.*

Proof Fix $T \in (0, \frac{1}{4B})$. We will first show that the assertion of the theorem is valid on $[0,T]\times\mathbb{R}$. Let $w(t,x) = (2T-t)^{-1/2}\exp(x^2/2(2T-t))$ for $(t,x) \in (0,T)\times\mathbb{R}$ and note that w satisfies the heat equation. Let $\delta > 0$ and define $v = v_\delta$ by $v(t,x) = u(t,x) - \delta w(t,x)$, then v satisfies the heat inequality as well. Consider for $h > 0$ the domain $D = D_h = \{(t,x) : t \in (0,T), |x| < h\}$. We apply the parabolic maximum principle to v with respect to the domain D and we estimate the function v on $\partial_0 D$. Letting $x = \pm h$, one obtains

$$\begin{aligned} v(t,\pm h) &\leq A\exp(Bh^2) - \delta w(t,h) \\ &\leq A\exp(Bh^2) - \delta(2T)^{-1/2}\exp(h^2/4T), \end{aligned} \tag{A.16}$$

where the last inequality follows from the previous one by taking $t = 0$, since $t \mapsto w(t,h)$ is increasing. By taking h big enough ($h^2 > \log\frac{A\sqrt{2T}}{\delta}/(\frac{1}{4T} - B)$), we can make (A.16) negative. Since we trivially have $v(0,x) \leq u(0,x) \leq 0$ we conclude that $v(t,x) \leq 0$ on all of $\partial_0 D$ and hence on $\bar{D}$ by Proposition A.14.

Note that it now follows that $v(t,x) \leq 0$ on $[0,T] \times \mathbb{R}$ by letting $h \to \infty$. This means that $u(t,x) \leq \delta w(t,x)$ and we conclude that thus (by letting $\delta \downarrow 0$) $u(t,x) \leq 0$ for all $(t,x) \in [0,T] \times \mathbb{R}$. Play now the same game with the function u' defined by $u'(t,x) = u(t+T,x)$ for $t \in [0,T]$. Clearly also u' satisfies the conditions of the theorem, so we conclude that $u'(t,x) \leq 0$ and hence $u(t,x) \leq 0$ for $(t,x) \in [T,2T] \times \mathbb{R}$. Iteration of this procedure concludes the proof. □

Corollary A.16 *Under the conditions of Theorem A.15 the heat equation with initial condition $u(0,\cdot) = f$ admits at most one solution.*

Proof Suppose that u and u' are two solutions, then their difference $v = u - u'$ is a solution to the heat equation with initial condition $v(0,x) = 0$. A double application of Theorem A.15 yields the result. □

A.9 Exercises

A.1 Let X be a random variable with $\mathbb{P}(X \geq 0) = 1$ and $\mathbb{P}(X > 0) > 0$.

(a) Show that there exists $n \in \mathbb{N}$ such that $\mathbb{P}(X > 1/n) > 0$. (*Reason by contradiction, assume that* $\mathbb{P}(X > 1/n) = 0$ *for all* $n \in \mathbb{N}$.)
(b) Show that $\mathbb{E}X > 0$.
(c) Suppose X is such that $\mathbb{P}(X \geq 0) = 1$ and $\mathbb{E}X = 0$. Show that it follows that $\mathbb{P}(X > 0) = 0$, equivalently $\mathbb{P}(X = 0) = 1$.

A.2 Show by computation that the density of a non-degenerate multivariate normal random vector has integral equal to 1.

A.3 Let Z be a standard normal random variable. Show by integration that $\mathbb{E}\exp(uZ) = \exp(\frac{1}{2}u^2)$ for $u \in \mathbb{R}$. If X has a normal $N(\mu, \sigma^2)$ distribution, what is $\mathbb{E}\exp(uX)$?

A.4 Let X have the standard normal distribution and $\phi(\lambda) = \mathbb{E}\exp(\mathrm{i}\lambda X)$, for $\lambda \in \mathbb{R}$.

(a) Argue that $\phi'(\lambda) = \mathrm{i}\mathbb{E}(X\exp(\mathrm{i}\lambda X))$.
(b) Show (use integration by parts) that $\phi'(\lambda) = -\lambda\phi(\lambda)$.
(c) Conclude that $\phi(\lambda) = \exp(-\frac{1}{2}\lambda^2)$.
(d) Let $X \sim N(\mu, \sigma^2)$. Show that $\phi(\lambda) = \exp(\mathrm{i}\mu\lambda - \frac{1}{2}\sigma^2\lambda^2)$.

A.5 Let Z be a random variable with $N(\mu, \sigma^2)$ distribution. Determine $m(u) = \mathbb{E}e^{uZ}$ (m is called the *moment generating function* of Z). *Hint*: Write the expectation as an integral with integration variable z and apply the substitution $y = \frac{z-\mu-u\sigma^2}{\sigma}$.

A positive random variable Z is said to have a *log-normal* distribution with parameters μ and σ^2 if $\log Z$ has a $N(\mu, \sigma^2)$ distribution. What is the expectation of a log-normally distributed random variable with parameters μ and σ^2?

A.6 Show by computation of the integral that Equation (A.2) holds for standard normal random variables.

A.7 Use characteristic functions to show that for a multivariate normal random vector X with $X^\top = (X_1^\top, X_2^\top)$ the components are independent iff they are uncorrelated.

A.8 Let for each $n \in \mathbb{N}$ the random variable X_n be degenerate in $1/n$. Show that $\mathbb{E}f(X_n) \to f(0)$ for every (bounded) continuous function f. Determine the limit random variable X to which the X_n converge weakly. Show also that the distribution functions F_n of X_n converge in all $x \in \mathbb{R}$. Is the limit function the distribution function X?

A.9 Define for $x \in \mathbb{R}$ the functions R_n by $R_n(x) = e^{\mathrm{i}x} - \sum_{k=0}^{n}(\mathrm{i}x)^k/k!$. Show that $|R_0(x)| \leq \min\{2, |x|\}$, $R_n(x) = \mathrm{i}\int_0^x R_{n-1}(t)\,\mathrm{d}t$ and finally that the estimates (A.3) and (A.4) are valid.

A.10 Suppose that $X_1, X_2, \ldots$ converges in distribution to the standard normal distribution. Let (a_n) and (b_n) be convergent sequences in $\mathbb{R}$ with limits a and b. Show that $\lim_{n\to\infty} \mathbb{P}(a_n < X_n \leq b_n) = \Phi(b) - \Phi(a)$.

A.11 Finish the proof of Theorem A.6.

A.12 Prove Theorem A.9.

A.13 Prove Proposition A.10.

A.14 Let X and Y be finite valued random variables and f a function defined on the range of X, and let $\hat{f}(Y) := \mathbb{E}[f(X)|Y]$.

(a) Write $F = \hat{f}(Y)$. Show that (A.6) holds, with (A.7) as its consequence.

(b) Conversely, if $\hat{f}$ is a function defined on the range of Y with the property of Equation (A.6). Show that $\hat{f}(y) = \mathbb{E}[f(X)|Y = y]$ for all y with $\mathbb{P}(Y = y) > 0$.

A.15 Let $X_1, X_2, \ldots$ be an independent sequence with $\mathbb{E}X_n = 0$ for all n. Let for each n the random variable M_n be defined by $M_n = \sum_{k=1}^{n} X_k$. Show that M is martingale w.r.t. the filtration $\mathbb{F}^X$.

A.16 Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{E}X_i = 1$ for all i. Define $P_n = \prod_{i=1}^{n} X_i$ and let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ $(n \geq 1)$. Show that the P_n form a martingale sequence.

A.17 Show that the process V defined by Equation (A.11) is a martingale.

A.18 Let X_n be multivariate normal random vectors, with expectations μ_n and covariance matrices Σ_n. Suppose that the X_n converge in $\mathcal{L}^2$-sense to a random vector X. Show that also X has a multivariate normal distribution. What are the expectation and covariance matrix of X?

A.19 Let $u : \mathbb{R}^2 \to \mathbb{R}$ be given by $u(t, x) = \alpha t + x^2$. Note that u is a solution to the heat equation $u_t = \frac{1}{2}\alpha u_{xx}$. Consider the region $D = \{(t, x) : t > 0, |x| < h\}$, where $h > 0$. Find (for different values of α) the points of $\bar{D}$ where u attains a maximum. Make a sketch of the level sets of u.

A.20 Here is an example of nonunicity of solutions to the heat equation. Let the initial condition be $u(0, \cdot) = 0$. One solution is $u(t, x) \equiv 0$. Let ϕ be an infinitely many times differentiable function. Show that $u(t, x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \frac{\mathrm{d}^n}{\mathrm{d}t^n}\phi(t)$ is a solution to the heat equation. Next we make the special choice $\phi(t) = \exp(-1/t^2)\mathbf{1}_{\{t>0\}}$. This function belongs to $C^\infty(\mathbb{R})$ and is 'flat' in $t = 0$, all derivatives in $t = 0$ are zero. Note that now $u(0, \cdot) = 0$, whereas $u(t, x)$ is not identically zero for $t > 0$. Why can't we apply Theorem A.15?

A.21 Let $u \in C^{1,2}((0,T) \times \mathbb{R})$ and continuous on $[0,T] \times \mathbb{R}$. Assume that u satisfies $|u(t,x)| \leq A \exp(B \log |x|^2)$ for some $A, B > 0$ and that $u(T, \cdot) = 0$. If u is a solution to the Black-Scholes partial differential equation, then u is identically zero on $[0,T] \times \mathbb{R}$.

A.22 Let X and Y be random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) Show that $X + Y$ and aX are random variables too, where $a \in \mathbb{R}$ is a constant. The space of random variables is thus a vector space over $\mathbb{R}$.

(b) Show that XY is a random variable.

(c) Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel-measurable function. Show that $f \circ X$ (usually denoted $f(X)$) is a random variable.

A.23 Show that two versions of the conditional expectation are equal almost surely: if $\hat{X}$ and $\hat{X}'$ both satisfy the requirements of Definition A.11, then $\mathbb{P}(\hat{X} = \hat{X}') = 1$. *Hint: take* $G = \{\hat{X} > \hat{X}'\}$.

A.24 Prove Proposition A.12(iii) from the general Definition A.11 of conditional expectation.

A.25 Show the validity of the inequality in (A.9)

A.26 Show that the martingale property (iii) of Definition A.13 and Equation A.10 are equivalent.

A.27 Prove the validity of Equation (A.13).

Index