

# Introduction to stochastic finance in continuous time

P.J.C. Spreij

this version: January 31, 2012



## 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 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 be 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. Suggestions are always welcome.

Amsterdam, November 2001

Peter Spreij



# Contents

<b>1</b>	<b>From discrete to continuous time</b>	<b>1</b>
1.1	A summary of discrete time results . . . . .	1
1.2	Limits in the CRR model . . . . .	9
1.3	Exercises . . . . .	13
<b>2</b>	<b>Brownian motion</b>	<b>15</b>
2.1	Interpolation of continuous functions . . . . .	15
2.2	Existence of Brownian motion . . . . .	16
2.3	Properties of Brownian motion . . . . .	20
2.4	Exercises . . . . .	22
<b>3</b>	<b>The heat equation</b>	<b>24</b>
3.1	Some theory . . . . .	24
3.2	Exercises . . . . .	27
<b>4</b>	<b>Girsanov's theorem in a simple case</b>	<b>29</b>
4.1	Measure transformations and Girsanov's theorem . . . . .	29
4.2	Exercises . . . . .	33
<b>5</b>	<b>Black Scholes market</b>	<b>35</b>
5.1	Models with equivalent measures . . . . .	35
5.2	Arbitrage . . . . .	36
5.3	Hedging . . . . .	39
5.4	Exercises . . . . .	41
<b>6</b>	<b>Elementary Itô calculus</b>	<b>42</b>
6.1	The Itô integral, an informal introduction . . . . .	42
6.2	The Itô rule . . . . .	48
6.3	Girsanov's theorem revisited . . . . .	51
6.4	Exercises . . . . .	53
<b>7</b>	<b>Applications of stochastic integrals in finance</b>	<b>55</b>
7.1	The models . . . . .	55
7.2	Self-financing portfolios and hedging . . . . .	56
7.3	American style claims . . . . .	59
7.4	The Greeks . . . . .	60
7.5	Exercises . . . . .	62
<b>A</b>	<b>Some results in Probability and in Analysis</b>	<b>64</b>
A.1	Normal random variables . . . . .	64
A.2	Characteristic functions . . . . .	64
A.3	Central limit theorem . . . . .	65
A.4	Conditional expectations . . . . .	68
A.5	Filtrations and Martingales . . . . .	69
A.6	Modes of convergence . . . . .	70
A.7	Uniqueness and the heat equation . . . . .	71
A.8	Exercises . . . . .	73



# 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 trading takes place at time instants  $t \in \{0, \dots, N\}$ . Traded are a bond with corresponding prices at time  $t$  equal to  $B_t$  and a stock 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. 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 some prescribed 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 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  $t$ .

We will often look at *portfolios*. Formally, a portfolio (in our context) is a sequence of 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 by  $V_t$  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 of  $S$ , 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  $S_{t-1}$ .

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  $\{0, \dots, N\}$ ) if 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 lose 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 say that a portfolio is an arbitrage opportunity if  $\mathbb{P}(V_0 = 0) = 1$ ,  $\mathbb{P}(V_N \geq 0) = 1$  and  $\mathbb{P}(V_N > 0) > 0$ . Note that arbitrage is now defined in probabilistic terms.

Another important concept is that of *self financing* portfolios. 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: A portfolio is called self financing if for all  $t \geq 0$  we have

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

We can give this an equivalent expression after having introduced the difference  $\Delta$  operator. For any process  $Y$  the process  $\Delta Y$  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

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

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

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

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

$$Z_t = \frac{S_t}{S_{t-1}}. \quad (1.4)$$

In the CRR model it is assumed that for all  $t$  the ratio  $Z_t$  takes on one of only two values  $u$  and  $d$ . 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  $\{sd^t, sud^{t-1}, \dots, su^t\}$ . Along with the process  $Z$  we introduce the (*cumulative*) *return process*  $R = (R_1, \dots, R_N)$ . It is defined by the equations

$$\Delta R_t = \frac{\Delta S_t}{S_{t-1}}, \text{ for } t \in \{1, \dots, N\}. \quad (1.5)$$

Trivially, we get from (1.5) the following equivalent relations  $\Delta S_t = S_{t-1} \Delta R_t$  and  $S_n = S_0 \prod_{t=1}^n (1 + \Delta R_t)$ . All these relations are summarized in the *notation*

$$S = S_0 \mathcal{E}(R), \quad (1.6)$$

with  $\mathcal{E}(R)_n = \prod_{t=1}^n (1 + \Delta R_t)$ .

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 (Exercise 1.1).

**Proposition 1.1** *The CRR market with  $N = 1$  is free of arbitrage iff  $d < 1 + r < u$ .*

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  $X = (S_1 - K)^+$ . 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  $t = 0$ . The portfolio that has this property 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 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 solution 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

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

Under the no arbitrage condition the numbers

$$q_u := \frac{1+r-d}{u-d} \quad \text{and} \quad q_d := \frac{u-(1+r)}{u-d} \tag{1.7}$$

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 claim 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. 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. Clearly, this is 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, here we 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  $x_2, x_1 = x_0$  and  $y_2, y_1 = y_0$ . Together 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  $X = F(s, Z_1, \dots, Z_N)$ . 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  $S_{N-1}$ . 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,  $x_N$  will be the same, hence we have the two equations

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

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

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

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 in

$$V_{N-1} = \frac{B_{N-1}}{B_N} (F(s, Z_1, \dots, Z_{N-1}, u)q_u + F(s, Z_1, \dots, Z_{N-1}, d)q_d), \quad (1.8)$$

with  $q_u = \frac{1+r-d}{u-d}$  and  $q_d = \frac{u-(1+r)}{u-d}$ . If we fix the values of  $Z_1, \dots, Z_{N-1}$  at  $z_1, \dots, z_{N-1}$ , then we can write the right hand side of (1.8) as

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

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  $y_{N-1}$ . Like in the previous time step we consider the two cases for the stock price movement separately. We get from Equation (1.8) the two relations

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

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 (do the computation yourself!)

$$\begin{aligned} V_{N-2} &= \frac{1}{(1+r)^2} (F(s, Z_1, \dots, Z_{N-2}, u, u)q_u^2 \\ &\quad + F(s, Z_1, \dots, Z_{N-2}, u, d)q_uq_d \\ &\quad + F(s, Z_1, \dots, Z_{N-2}, d, u)q_uq_d \\ &\quad + F(s, Z_1, \dots, Z_{N-2}, d, d)q_d^2). \end{aligned} \tag{1.9}$$

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

$$\begin{aligned} V_{N-2} &=: v_{N-2}(z_1, \dots, z_{N-2}) \\ &= \frac{1}{(1+r)^2} \mathbb{E}_{\mathbb{Q}} F(s, z_1, \dots, z_{N-2}, Z_{N-1}, Z_N), \end{aligned} \tag{1.10}$$

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 (1.10) it becomes clear how the compact form should look like. At any time  $n$ , given that the values of  $Z_1, \dots, Z_n$  are  $z_1, \dots, z_n$ , we have

$$\begin{aligned} V_n &=: v_n(z_1, \dots, z_n) \\ &= \frac{1}{(1+r)^{N-n}} \mathbb{E}_{\mathbb{Q}} F(s, z_1, \dots, z_n, Z_{n+1}, \dots, Z_N). \end{aligned} \tag{1.11}$$

In this expression  $\mathbb{E}_{\mathbb{Q}}$  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  $n = 1, \dots, N$  and, moreover, it makes the  $Z_k$  independent random variables. We notice 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  $\mathbb{E}_{\mathbb{Q}}[\bar{S}_n | Z_1, \dots, Z_{n-1}] = \bar{S}_{n-1}$ . Equivalently we have  $\mathbb{E}_{\mathbb{Q}}[Z_n | Z_1, \dots, Z_{n-1}] = 1+r$ .

Let us show the asserted uniqueness of  $\mathbb{Q}$ . Computing the last conditional expectation as  $u\mathbb{Q}(Z_n = u | Z_1, \dots, Z_{n-1}) + d\mathbb{Q}(Z_n = d | Z_1, \dots, Z_{n-1}) = 1+r$ ,

we see that the conditional distribution of  $Z_n$  given  $Z_1, \dots, Z_{n-1}$  is determined by  $\mathbb{Q}(Z_n = u | Z_1, \dots, Z_{n-1}) = \frac{1+r-d}{u-d}$ , and that this conditional probability is then also equal to the unconditional probability. Hence it follows that the distribution under  $\mathbb{Q}$  of the vector  $(Z_1, \dots, Z_n)$  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.4 or A.6), we can write (1.11) in the following equivalent form (with  $X = F(s, Z_1, \dots, Z_N)$ ).

$$V_n = \frac{1}{(1+r)^{N-n}} \mathbb{E}_{\mathbb{Q}}[F(s, Z_1, \dots, Z_N) | Z_1, \dots, Z_n] \quad (1.12)$$

$$= \frac{1}{(1+r)^{N-n}} \mathbb{E}_{\mathbb{Q}}[X | S_1, \dots, S_n], \quad (1.13)$$

where we used in the last equation that conditioning on  $Z_1, \dots, Z_n$  is equivalent to conditioning on  $S_1, \dots, S_n$ . Switching to the discounted value process  $\bar{V}$  we get from expression (1.13) that

$$\bar{V}_n = \mathbb{E}_{\mathbb{Q}}[\bar{X} | S_1, \dots, S_n]. \quad (1.14)$$

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

$$\bar{V}_n = \mathbb{E}_{\mathbb{Q}}[\bar{V}_{n+1} | S_1, \dots, S_n], \text{ for } n = 0, \dots, N-1. \quad (1.15)$$

Both equations (1.14) and (1.15) express the fact that also the discounted value process  $\bar{V}$  is a martingale under the measure  $\mathbb{Q}$ .

For simple claims expression (1.11) 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

$$\bar{v}_n(s) = \mathbb{E}_{\mathbb{Q}} \bar{f}(s \frac{S_N}{S_n}). \quad (1.16)$$

Since under the probability measure  $\mathbb{Q}$  the ratio  $S_N/S_n$  and  $S_n$  are independent (why?), property (iv) of Proposition A.6 says that  $\bar{v}_n(s) = \mathbb{E}_{\mathbb{Q}}[\bar{f}(S_N) | S_n = s]$  and hence  $\bar{V}_n = \mathbb{E}_{\mathbb{Q}}[\bar{f}(S_N) | S_n]$ .

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 a, possibly empty, interval in  $\{0, \dots, N-n\}$ . So, with  $a_n = \inf E(n)$  we get

$$\begin{aligned}
V_n &= (1+r)^{-N+n} \sum_{i=a_n}^{N-n} (S_n u^i d^{N-n-i} - K) \binom{N-n}{i} q_u^i q_d^{N-n-i} \\
&= (1+r)^{-N+n} \sum_{i=a_n}^{N-n} (S_n u^i d^{N-n-i} - K) \binom{N-n}{i} q_u^i q_d^{N-n-i} \\
&= S_n \sum_{i=a_n}^{N-n} \binom{N-n}{i} \left( \frac{uq_u}{1+r} \right)^i \left( \frac{dq_d}{1+r} \right)^{N-n-i} \\
&\quad - K(1+r)^{n-N} \sum_{i=a_n}^{N-n} \binom{N-n}{i} q_u^i q_d^{N-n-i}. \tag{1.17}
\end{aligned}$$

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 (1.17) as

$$S_n \pi(N-n, p, a_n) - K(1+r)^{-N+n} \pi(N-n, q_u, a_n). \tag{1.18}$$

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 for any contingent claim (under the proviso that  $u > d$ ). One says thus that every contingent claim in the CRR market can be *hedged* and, by definition, that the CRR market is *complete*. The sequence  $(x_n, y_n), n = 1, \dots, N$  is called a *hedging portfolio* or a *hedging strategy*. Not only is this market complete, we also have

**Proposition 1.2** *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.

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).  $\square$

The above construction of a hedging strategy shows an interesting aspect. Suppose that two persons have different views 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 .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!

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  $n$  by  $C_n$  and that of the put option by  $P_n$ . Then we easily obtain

$$C_n - P_n = S_n - (1 + r)^{n-N}K. \quad (1.19)$$

## 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.3). 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  $\sigma > 0$  we put

$$r_N = \exp(r\Delta_N) - 1 \quad (1.20)$$

$$u_N = \exp(\sigma\sqrt{\Delta_N}) \quad (1.21)$$

$$d_N = \exp(-\sigma\sqrt{\Delta_N}). \quad (1.22)$$

The consequences of these 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  $B^N(t) = B_k^N = (1 + r_N)^k = \exp(rt)$ .

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  $t_k^N \rightarrow t$  as  $N \rightarrow \infty$ , we get  $B^N(t) \rightarrow \exp(rt)$ .

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 have from Equation (1.7)

$$q_u(N) = \frac{e^{r\Delta_N} - \exp(-\sigma\sqrt{\Delta_N})}{\exp(\sigma\sqrt{\Delta_N}) - \exp(-\sigma\sqrt{\Delta_N})}. \quad (1.23)$$

We will consider what happens if  $N \rightarrow \infty$ . Using the Taylor expansion of the exponential function we get for  $N \rightarrow \infty$

$$q_u(N) = \frac{1}{2} + (r - \frac{1}{2}\sigma^2)\frac{\sqrt{\Delta_N}}{2\sigma} + O(\Delta_N), \quad (1.24)$$

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

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 (1.18) 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 we have the inequalities

$$\frac{\log \frac{K}{S^N(t)d_N^{N-n}}}{\log \frac{u_N}{d_N}} < a_N(t) \leq \frac{\log \frac{K}{S^N(t)d_N^{N-n}}}{\log \frac{u_N}{d_N}} + 1. \quad (1.25)$$

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  $N \rightarrow \infty$  and  $\frac{n}{N} \sim \frac{t}{T}$  so that  $t_n^N \rightarrow t$ . Let us introduce the auxiliary random variable  $Y_N$  which has a  $\text{Bin}(N-n, p_N)$  distribution. Note that  $(N-n)\Delta_N \rightarrow T-t$ . With the aid of  $Y_N$  we have  $\pi(N-n, p_N, a_N(t)) = \mathbb{P}(Y_N > a_N(t))$ . To apply the Central limit theorem, we have to compute

$$\mathbb{P}(Y_N > a_N(t)) = \mathbb{P}\left(\frac{Y_N - \mathbb{E}Y_N}{\sqrt{\text{Var}Y_N}} > \alpha_N(t)\right),$$

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

$$p_N = \frac{u_N q_u(N)}{1+r_N} = \frac{1}{2} + (r + \frac{1}{2}\sigma^2)\frac{\sqrt{\Delta_N}}{2\sigma} + O(\Delta_N). \quad (1.26)$$

Hence we get (using (1.25))

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

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.7)

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

The convergence of the probabilities  $\pi(N - n, q_u(N), a_N(t))$  can be treated similarly (this is Exercise 1.7). The computations above are now summarized in

**Theorem 1.3** *Under the assumptions of this section at time  $t$  when the stock price has the value  $S^N(t) = s$ , the fair price of a European call option with payoff  $(S^N(T) - K)^+$  has the limiting expression*

$$s\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2), \quad (1.27)$$

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

Equation (1.27) is the famous *Black-Scholes formula*, to which we will return later.

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  $k = 1, \dots, N$ . Define for each  $t \in [0, T]$  the random variable  $W^N(t) = \sum_{k \leq \frac{t}{T}N} \log Z_k^N$ .

**Proposition 1.4** *Let for each  $N$  the distribution of the  $S_0^N, \dots, S_N^N$  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  $k = 0, \dots, N$  and that makes the  $Z_1^N, \dots, Z_N^N$  independent. Let  $t_1, \dots, t_n$  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  $N \rightarrow \infty$  (under the probability measures  $\mathbb{Q}^N$ ) the random variables  $\Delta W_k^N$  converge in distribution to a random variable  $\Delta W_k$  that has a  $N((t_k - t_{k-1})(r - \frac{1}{2}\sigma^2), (t_k - t_{k-1})\sigma^2)$  distribution. Moreover, the distribution of the random  $n$ -vector  $(\Delta W_1, \dots, \Delta W_n)$  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  $k = 1, \dots, n$ ) 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  $N((r - \frac{1}{2}\sigma^2)(t_k - t_{k-1}), \sigma^2(t_k - t_{k-1}))$  distribution. Moreover, the (limit) distribution of this vector is such that its components are independent.

**Proof** First we compute  $\mathbb{E} \Delta W_k^N = \sum_{\frac{t_{k-1}}{T}N < j \leq \frac{t_k}{T}N} \mathbb{E} \log Z_j^N$ . 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 (1.23) 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  $\Delta W_k^N$  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  $\text{Var} \Delta W_k^N = 4q_u(N)q_d(N)\sigma^2(t_k - t_{k-1}) \rightarrow \sigma^2(t_k - t_{k-1})$ . Introducing  $\Delta \tilde{W}_k^N = \Delta W_k^N - \mathbb{E} \Delta W_k^N$ , we apply Theorem A.3 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  $\Delta W_k^N$  then follows from Exercise A.7.  $\square$

In Proposition 1.4 we have found the limit distribution for the log price process of the stock, it was such that the limit random variable  $S(t)$  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 1.5** *Let  $t_1, t_2, \dots$  be an increasing sequence in  $[0, T]$ . Under the same assumptions as in Proposition 1.4 we have that the  $n$ -vector with elements  $R^N(t_k) - R^N(t_{k-1})$  ( $k = 1, \dots, n$ ) 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  $N(r(t_k - t_{k-1}), \sigma^2(t_k - t_{k-1}))$  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 the previous proposition. 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.3 and Exercise A.7 will complete the proof.  $\square$

Compare the limit distributions of Propositions 1.4 and 1.5. We see that they are both normal, with the same variance, but with different expectation. 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  $\Delta R_k^N$  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 distributions that we observed.

The assertion of Proposition 1.5 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  $N \rightarrow \infty$ )

$$\Delta S^N(t_k) \approx S^N(t_{k-1})(r\Delta t_k + \sigma\Delta\mathcal{B}(t_k)). \quad (1.28)$$

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

With the result of Proposition 1.4 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  $e^U$  for the ratio  $S^N(T)/S^N(t)$ . Then we have parallel to Equation (1.16)

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

From Proposition 1.4 we know that the limit distribution of  $U$  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.1) to conclude that  $\bar{v}_t^N(s)$  converges to

$$\bar{v}_t(s) := \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \bar{f}(se^{(r - \frac{1}{2}\sigma^2)(T-t) + \sigma\sqrt{T-t}z}) e^{-z^2/2} dz. \quad (1.29)$$

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 1.4 and 1.5). It is possible to go beyond this and show weak 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 1.1.

**1.2** (a) Consider the claim with final payoff  $S_N$ . Derive its fair price at time  $n \leq N$ .

(b) Derive the put-call parity Equation (1.19).

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

**1.4** Show the validity of (1.9).

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

**1.7** Show by using the Central limit theorem that with  $n \sim \frac{t}{T}N$  one gets  $\lim_{N \rightarrow \infty} \pi(N - n, q_u(N), a_N(t)) = d_2(t)$  (notation as in Theorem 1.3).

**1.8** Prove Proposition 1.5.

**1.9** What is the distribution of the vector  $(R(t_1), \dots, R(t_n))$  of Proposition 1.5?

**1.10** Use Equation (1.29) to compute explicitly the limit price of a European put option (with payoff  $(K - S(T))^+$ ).

**1.11** 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. Derive the limit put-call parity equation  $C(t) - P(t) = s - e^{-r(T-t)}K$ . Use the result of Exercise 1.10 to arrive at the Black-Scholes formula (1.27).

## 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  $f : [0, 1] \rightarrow \mathbb{R}$  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  $\{k2^{-n} : k = 0, \dots, 2^n\}$ . 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  $k = 0, \dots, 2^{n-1}$  and by linear interpolation between these points. Look now at the points  $t_k^n$  for  $k = 0, \dots, 2^n$ . 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  $k \in I(n) = \{1, \dots, 2^{n-1}\}$  by

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

where  $\sigma_n = 2^{-\frac{1}{2}(n+1)}$ . Next we put  $S_k^n(t) = \int_0^t H_k^n(u) du$ . Note that 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  $\sigma_n$  at  $t_{2k-1}^n$ .

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  $\eta_k^n$  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  $t$

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

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

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

**Theorem 2.1** *Let  $f$  be a continuous function on  $[0, 1]$ . With the  $f_n$  defined by (2.3) we have  $\|f - f_n\| \rightarrow 0$ , 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  $\|f - f_n\| \rightarrow 0$ .  $\square$

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

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

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  $\Omega$ ). One may alternatively view  $X$  as a function  $(\omega, t) \mapsto X(\omega, t)$  on  $\Omega \times \mathcal{T}$ , so that  $X(t)$  is defined by  $X(t) : \omega \rightarrow X(\omega, t)$ . For fixed  $\omega$  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  $0 \leq t_0 \leq \dots \leq t_n$  the random variables  $W(t_k) - W(t_{k-1})$  ( $k = 1, \dots, n$ ) are independent.

Having posed the definition of Brownian motion, we ask whether it exists in a precise mathematical sense. Clearly, the underlying outcome space  $\Omega$  has to be big enough (on an  $\Omega$  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  $\Omega$  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  $\Omega$  that is such all the paths become continuous, so that we can identify this  $\Omega$  with  $C[0, \infty)$ . In the construction below, a different  $\Omega$  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 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  $\xi$  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 below.

Let, as in Section 2.1,  $I(0) = \{1\}$  and  $I(n)$  be the set  $\{1, \dots, 2^{n-1}\}$  for  $n \geq 1$ . Take a sequence of independent standard normally distributed random variables  $\xi_k^n$  that are all defined on some probability space  $\Omega$  with  $k \in I(n)$  and  $n \in \mathbb{N} \cup \{0\}$  (it is a result in probability theory that one can take for  $\Omega$  a countable product of copies of  $\mathbb{R}$ , endowed with a product  $\sigma$ -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

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

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

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

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

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

**Theorem 2.5** *For almost all  $\omega$  the functions  $t \mapsto W^n(\omega, t)$  converge uniformly to a continuous function  $t \mapsto W(\omega, t)$  and the process  $W : (\omega, t) \rightarrow W(\omega, t)$  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)

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

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 in virtue of the Borel-Cantelli lemma (Exercise 2.3)  $\mathbb{P}(\limsup\{\beta_n > n\}) = 0$ . Hence we can find a subset  $\tilde{\Omega}$  of  $\Omega$  with  $\mathbb{P}(\tilde{\Omega}) = 1$ , 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  $p > n \geq n(\omega)$  we have

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

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  $\omega$ 's not in  $\tilde{\Omega}$  we define  $W(\omega, \cdot) = 0$ . So we now have continuous functions  $W(\omega, \cdot)$  for all  $\omega$  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.2) we will prove that the characteristic function  $\mathbb{E} \exp(i \sum_j \lambda_j \Delta_j)$  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  $L^2[0, 1]$  (see Exercise 2.5). Hence every function  $f \in L^2[0, 1]$  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  $L^2[0, 1]$  and where the infinite sum is convergent in  $L^2[0, 1]$ . As a result we have for any two functions  $f$  and  $g$  in  $L^2[0, 1]$  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  $f = 1_{[0,t]}$  and  $g = 1_{[0,s]}$  results in  $\langle 1_{[0,t]}, H_k^n \rangle = S_k^n(t)$  and  $t \wedge s = \langle 1_{[0,t]}, 1_{[0,s]} \rangle = \sum_{n,k} S_k^n(t) S_k^n(s)$ .

We use this property to compute the limit of  $\text{Cov}(W^n(t), W^n(s))$ . We have

$$\begin{aligned} \text{Cov}(W^n(t), W^n(s)) &= \mathbb{E} \left( \sum_{m=0}^n \sum_{k \in I(m)} \xi_k^m S_k^m(t) \sum_{p=0}^n \sum_{l \in I(p)} \xi_l^p S_l^p(s) \right) \\ &= \sum_{m=0}^n \sum_{k \in I(m)} S_k^m(t) S_k^m(s), \end{aligned}$$

which converges to  $s \wedge t$  for  $n \rightarrow \infty$ . Since for all fixed  $t$  we have  $W^n(t) \rightarrow W(t)$  a.s., we have  $\mathbb{E} \exp(i \sum_j \lambda_j \Delta_j) = \lim_{n \rightarrow \infty} \mathbb{E} \exp(i \sum_j \lambda_j \Delta_j^n)$  with  $\Delta_j^n = W^n(t_j) - W^n(t_{j-1})$ . We compute

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

Working out the double product, we get in the exponential the sum over the four variables  $i, j, m \leq n, k$  of  $-\frac{1}{2} \lambda_i \lambda_j$  times

$$S_k^m(t_j) S_k^m(t_i) - S_k^m(t_{j-1}) S_k^m(t_i) - S_k^m(t_j) S_k^m(t_{i-1}) + S_k^m(t_{j-1}) S_k^m(t_{i-1})$$

and this sum converges to  $\exp(-\frac{1}{2} \sum_j \lambda_j^2 (t_j - t_{j-1}))$  as  $n \rightarrow \infty$ . This completes the proof.  $\square$

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  $\Omega = \prod_n \Omega_n$ ,  $\mathcal{F} = \prod_n \mathcal{F}_n$  and  $\mathbb{P} = \prod_n \mathbb{P}_n$ . Note that these Brownian motions are independent by construction. Let  $\omega = (\omega_1, \omega_2, \dots)$  and define then

$$W(\omega, t) = \sum_{n \geq 0} 1_{[n, n+1)}(t) \left( \sum_{k=1}^n W_k(\omega_k, 1) + W_{n+1}(\omega_{n+1}, t - n) \right). \quad (2.10)$$

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 see that the thus defined process has independent increments. It is immediate that  $\mathbb{E} W(t) = 0$  and that  $\text{Var} W(t) = t$ .

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  $\sigma^2$  is the process  $\sigma W$ . Brownian motion with linear drift is a process  $X$  with  $X(t) = bt + \sigma W(t)$ , where  $b$  and

$\sigma$  are real constants. Note that also  $X$  has independent increments and that  $\mathbb{E}X(t) = bt$  and  $\text{Var } X(t) = \sigma^2 t$ .

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)

$$\mathbb{E}S(t) = \exp(bt + \frac{1}{2}\sigma^2 t) \quad (2.11)$$

$$\text{Var } S(t) = \exp(2bt + \sigma^2 t)(\exp(\sigma^2 t) - 1). \quad (2.12)$$

### 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 of Proposition 2.6. The content of this proposition is that Brownian motion is of bounded *quadratic variation*. Let us introduce some notation. Consider an interval  $[0, t]$  and let  $\Pi_n = \{t_0^n, \dots, t_{k_n}^n\}$  be a partition of it with  $0 = t_0^n < \dots < t_{k_n}^n = t$  and denote by  $\mu_n$  its mesh:  $\mu_n = \max\{t_j^n - t_{j-1}^n : j = 1, \dots, k_n\}$ . Let  $V_n^2 = \sum_{j=1}^{k_n} (W(t_j^n) - W(t_{j-1}^n))^2$ .

**Proposition 2.6** *If  $\mu_n \rightarrow 0$ , then  $\mathbb{E}(V_n^2 - t)^2 \rightarrow 0$ . If moreover  $\sum_{n=1}^{\infty} \mu_n < \infty$ , then also  $V_n^2 \rightarrow t$  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  $\mathbb{E}V_n^2 = t$ . Moreover  $\text{Var } V_n^2 = \sum_{j=1}^{k_n} \text{Var} (W(t_j^n) - W(t_{j-1}^n))^2 = 2 \sum_{j=1}^{k_n} (t_j^n - t_{j-1}^n)^2$  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  $V_n^2 \rightarrow t$  in probability. Indeed  $\mathbb{P}(|V_n^2 - t| > \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var } V_n^2 \rightarrow 0$ .

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  $\omega \in \liminf E_n^c$  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.  $\square$

**Corollary 2.7** *With  $V_n^1 = \sum_{j=1}^{k_n} |W(t_j^n) - W(t_{j-1}^n)|$  we have  $V_n^1 \rightarrow \infty$  a.s., if  $\mu_n \rightarrow 0$ .*

**Proof** This follows from Exercise 2.10.  $\square$

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

**Proposition 2.8** *The paths of Brownian motion are a.s. Hölder continuous of any order  $\gamma$  with  $\gamma < \frac{1}{2}$ , i.e. for almost every  $\omega$  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). 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).  $\square$

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 2.9 says that they are nowhere differentiable.

**Proposition 2.9** *Put  $D = \{\omega : t \mapsto W(\omega, t) \text{ is differentiable at some } s \in (0, 1)\}$ . Then  $D$  is contained in a set of zero probability.*

**Proof** Let  $A_{jk}(s)$  be the set

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

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  $n \geq 4/k$ , choose  $\omega \in A_{jk}(s)$  and choose  $i$  such that  $s \in (\frac{i-1}{n}, \frac{i}{n}]$ . Note first for  $l = 0, 1, 2, 3$  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

$$A_{jk} \subset \bigcup_{i=1}^n \bigcap_{l=1,2,3} \{\omega : |W(\omega, \frac{i+l}{n}) - W(\omega, \frac{i+l-1}{n})| \leq \frac{2l+1}{n}j\}. \quad (2.13)$$

We proceed by showing that the right hand side of this inclusion has probability 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 (2.13) is the product of the probabilities of each of the terms and this product is less than  $105j^3n^{-3/2}$ . Hence  $\mathbb{P}(A_{jk}) \leq 105j^3n^{-1/2}$ , which tends to zero for  $n \rightarrow \infty$ . The conclusion now follows from  $D \subset \bigcup_{jk} A_{jk}$ .  $\square$

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 2.10** *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*

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

**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  $\Phi$ ) we can compute, exploiting the mentioned independence and using Proposition A.6 (iv),

$$\begin{aligned} \mathbb{P}(W(t+h) \leq x | W(s), s \leq t) &= \\ \mathbb{P}(W(t+h) - W(t) \leq x - W(t) | W(s), s \leq t) &= \Phi\left(\frac{x - W(t)}{\sqrt{h}}\right). \end{aligned}$$

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

Proposition 2.10 has as a consequence that all conditional expectations of functions of  $W(t+h)$  given  $W(s), s \leq t$  reduce to conditional expectations given  $W(t)$ .

## 2.4 Exercises

**2.1** Show that the random variable  $Z$  on page 17 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  $\mathbb{P}(\limsup A_k) = 0$ . *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  $L^2[0, 1]$ . 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  $F = \int_0^1 f(u) du$ . Show that  $F$  is zero in all  $t = k2^{-n}$ , and therefore zero on the whole interval. Conclude that  $f = 0$ .

**2.6** Let  $(X_n)$  be a sequence of random  $k$ -vectors that in  $L^2$ -sense converge to a random vector  $X$ . Show that  $\mathbb{E}X_n \rightarrow \mathbb{E}X$  and  $\text{Cov}(X_n) \rightarrow \text{Cov}(X)$ .

Assume then that all  $X_n$  are (multivariate) normal. Show that also  $X$  is (multivariate) normal. What are the expectation and the covariance matrix of  $X$ ?

**2.7** Consider the processes  $W^n$  of Section 2.2. Let  $t_1, \dots, t_k \in [0, 1]$ . Show that the sequence of random vectors  $(W_{t_1}^n, \dots, W_{t_k}^n)$  in  $L^2$ -sense converges to  $(W_{t_1}, \dots, W_{t_k})$ . (*Hint*: this sequence is Cauchy in  $L^2$ . Identify the limit.)

**2.8** Show that the increments of the process  $W$  defined by Equation (2.10) are independent.

**2.9** Show the validity of equations (2.11) and (2.12).

**2.10** The  $p$ -th order variation of a function  $f : [0, 1] \rightarrow \mathbb{R}$  over a partition  $\Pi = \{t_0, \dots, t_n\}$  of  $[0, 1]$  with  $0 = t_0 < t_1 < \dots < t_n = 1$  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.

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  $0 < V^p(f) < \infty$  (cf Exercise 2.10) for all  $p > 0$ .

**2.12** Prove Proposition 2.8.

**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,  $X(t) = bt + \sigma W(t)$ . 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  $\sigma^2 T = \sigma^2 \langle W \rangle_T$ .

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

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

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

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

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

Let  $f$  be a sufficiently well behaving function (below we 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

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

Now suppose that it is allowed to interchange integration and (partial) differentiation in (3.2). Then we obtain  $\frac{\partial u}{\partial t}(t, x) = \int_{-\infty}^{\infty} f(z) \frac{\partial p}{\partial t}(t, x - z) dz$  and  $\frac{\partial^2 u}{\partial x^2}(t, x) = \int_{-\infty}^{\infty} f(z) \frac{\partial^2 p}{\partial x^2}(t, x - z) dz$ . 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  $\int_{-\infty}^{\infty} \exp(-ax^2)|f(x)| dx$  is finite.

**Proposition 3.2** Let  $f$  satisfy Condition 3.1 and let  $u : [0, \frac{1}{2a}) \times \mathbb{R} \rightarrow \mathbb{R}$  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

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

Note that there exists a constant  $C = C(a, \varepsilon, t_0)$ , such that 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

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

Let's concentrate on the last factor in (3.3). We will show that it is bounded for all  $y \in \mathbb{R}$  and  $(t, x) \in E$ .

First we have  $ay^2 - \frac{(x-y)^2}{2t} \leq ay^2 - (a + \varepsilon)(x - y)^2 = -\varepsilon y^2 - (a + \varepsilon)x^2 + 2(a + \varepsilon)xy \leq 2(a + \varepsilon)\beta|y| - \varepsilon y^2$ . From this we obtain that  $((x - y)^2 + 1)\exp(ay^2)p(t, x - y)$  is bounded by a constant  $D$  say, on the indicated domain. Hence  $|f(y)q(t, x - y)| \leq CD\exp(-ay^2)|f(y)|$ , which is integrable in virtue of Condition 3.1. We then obtain that  $(t, x) \mapsto \int f(y)q(t, x - y)dy$  is continuous in  $(t, x)$  on  $E$  by the dominated convergence theorem, and then also on  $(0, 1/2a) \times \mathbb{R}$  by letting  $\beta \rightarrow \infty$  and  $\varepsilon, t_0 \rightarrow 0$ .

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

$$\frac{1}{h}(u(t, x + h) - u(t, x)) = \int_{-\infty}^{\infty} f(y)q(t, \theta_h(t, x) - y)dy,\tag{3.4}$$

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

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

which is what we had to show. The analogous results for the other derivatives can be derived similarly.

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  $f(x_0) = 0$ . Fix  $\varepsilon > 0$ , then there is  $\delta > 0$  such that  $|f(y)| \leq \varepsilon$  for  $|y - x_0| < \delta$ . We will show that  $|u(t, x)|$

is small for  $t < 1/2a$  and  $x \in (x_0 - \frac{1}{2}\delta, x_0 + \frac{1}{2}\delta)$ . Obviously we have

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

We treat the latter integral in detail, the other one can be dealt with similarly. So assume that  $t < 1/2a$ . Write it as  $\frac{1}{\sqrt{2\pi t}} \int_{x_0 + \delta}^{\infty} |f(y)| \exp(-ay^2) \times \exp(ay^2 - \frac{(x-y)^2}{2t}) dy$  and look at  $h(y) = ay^2 - \frac{(x-y)^2}{2t}$ . For  $t < 1/2a$  this terms is maximal at  $y_0 = \frac{x}{1-2at} \rightarrow x$ . For small enough  $t$  we have  $y_0 < x_0 + \frac{1}{2}\delta$ , which lies outside the integration interval. Hence it follows that for  $y \geq x_0 + \delta$  we have that  $h$  is maximal at  $x_0 + \delta$  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 the integral we are interested in can be bounded by  $\int |f(y)| \exp(-ay^2) dy$  times  $\exp(a(x_0 + \delta)^2)p(t, \frac{1}{2}\delta)$  which tends to 0. The other integral in (3.5) can be treated similarly.  $\square$

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 without proof a theorem that gives affirmative answers to these questions. See also Section A.7 of the Appendix.

**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  $\sup\{\exp(-ax^2)|u_i(t, x)| : (t, x) \in (0, T] \times \mathbb{R}\} < \infty$  for some  $a > 0$  ( $i = 1, 2$ ) and such that  $\lim_{t \downarrow 0, y \rightarrow x} u_1(t, y) = \lim_{t \downarrow 0, y \rightarrow x} u_2(t, y)$  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  $0 \leq T \leq \infty$  that solves the heat equation and assume that for all  $x \in \mathbb{R}$  the limit  $f(x) := \lim_{t \downarrow 0, y \rightarrow x} u(t, x)$  exists. Then  $u$  has for  $t > 0$  the integral representation  $u(t, x) = \int_{-\infty}^{\infty} f(y)p(t, x - y) dy$ .

The first assertion of Theorem 3.3 is proved in the Appendix, Section A.7. Also the second assertion of the theorem is 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 equations 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 we will also consider the *backward heat equation*

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

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  $\lim_{t \uparrow T, y \rightarrow x} u(t, y) =: f(x)$  exists. Then  $v$  has the integral representation  $v(t, x) = \int_{-\infty}^{\infty} f(x + y)p(T - t, y) dy$ .

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

The probabilistic interpretation of solutions  $v$  of the backward heat equation is  $v(t, x) = \mathbb{E} f(x + W(T - t))$ . But, 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 part (iv) of Proposition A.6, we can also write  $v(t, x) = \mathbb{E} [f(W_T) | W_t = x]$ . We will come back to this interpretation of  $v$  later on.

## 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  $u_t(t, x) = \int_{-\infty}^{\infty} f(y) p_t(t, x - y) dy$ .

**3.3** The polynomials in two variables  $H_n(t, x)$  ( $n \geq 0$ ) are defined by  $H_n(t, x) = \frac{\partial^n}{\partial \alpha^n} \exp(\alpha x - \frac{1}{2} \alpha^2 t) |_{\alpha=0}$ . Show that  $\frac{\partial}{\partial x} H_n(t, x) = n H_{n-1}(t, x)$  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. For which points  $(t, x)$  is  $u(t, x)$  well defined? Check that your answer is in agreement with Proposition 3.2. 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  $u(t, x) = v(t, e^{x - \frac{1}{2} \sigma^2 t})$ . Suppose that  $v$  is a solution of

$$v_t(t, x) + \frac{1}{2} \sigma^2 x^2 v_{xx}(t, x) = 0. \quad (3.7)$$

Show that  $u$  satisfies

$$u_t(t, x) + \frac{1}{2} \sigma^2 u_{xx}(t, x) = 0. \quad (3.8)$$

Conversely, let  $u$  be a solution of (3.8). For  $x > 0$  we put  $v(t, x) = u(t, \log x + \frac{1}{2} \sigma^2 t)$ . Show that  $v$  satisfies (3.7).

**3.6** Suppose that the function  $v$  satisfies

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

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

$$w_t(t, x) + \frac{1}{2} \sigma^2 x^2 w_{xx}(t, x) + r x w_x(t, x) - r w(t, x) = 0. \quad (3.10)$$

Conversely, if  $w$  is a solution of (3.10) and  $v(t, x) = w(t, x e^{rt}) e^{-rt}$ , then  $v$  is a solution to (3.9). *Remark:* Equation (3.10) is called the *Black-Scholes partial differential equation*, 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  $p = \frac{1}{\sqrt{2\pi t}} \exp(-x^2/2t)$ . Show that all partial derivatives of  $p$  satisfy the heat equation. Is this in agreement with Theorem 3.3?

**3.9** Let  $w(t, s)$  be given by (1.27). Show by direct computation that  $w$  is a solution of (3.10) (with  $x$  replaced by  $s$ ) and compute  $\lim_{t \rightarrow T} w(t, s)$ . *Hint:* Show and use  $\phi(d_2) = \frac{s}{K} \phi(d_1) \exp(r(T-t))$  and  $\frac{\partial d_2}{\partial t} = \frac{\partial d_1}{\partial t} + \frac{1}{2}\sigma(T-t)^{-1/2}$ .

## 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

$$X(t) = x_0 + at + \sigma W_t, \quad (4.1)$$

where  $x_0$  is a given real number (the initial value of the process  $X$ ) and  $a$  and  $\sigma$  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  $\sigma$ -algebra on this space, we take the Borel  $\sigma$ -algebra that is induced by the sup norm. Clearly, the law depends on  $a$  and  $\sigma^2$ . What can we say of the relation between these laws, if we let the parameters  $a$  and  $\sigma$  vary? We will see that for fixed  $\sigma^2$  the laws for varying  $a$  are *equivalent*, and for varying  $\sigma$  they are mutually *singular*. Let us define this terminology.

Let  $L_+^0$  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

$$\mathbb{E}_{\mathbb{Q}} X = \mathbb{E}_{\mathbb{P}} Z X, \quad \text{for all } X \in L_+^0. \quad (4.2)$$

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  $\frac{d\mathbb{Q}}{d\mathbb{P}}$ . Note that necessarily we have  $\mathbb{E}_{\mathbb{P}} Z = 1$  (take  $X = 1$ ). Every 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.

Let us also observe that  $\mathbb{Q}(Z > 0) = 1$ , because  $\mathbb{Q}(Z = 0) = \mathbb{E}_{\mathbb{P}} 1_{\{Z=0\}} Z = \mathbb{E}_{\mathbb{P}} 0 = 0$ . Is  $\mathbb{P}(Z > 0) = 1$ ? In general not, see Exercise 4.2. But if  $\mathbb{P}(Z = 0) = 0$ , then we also have  $\mathbb{P} \ll \mathbb{Q}$  and  $d\mathbb{P}/d\mathbb{Q} = 1/Z$ . 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  $\mathbb{Q} \ll \mathbb{P}$ .

Note that if  $\mathbb{Q} \ll \mathbb{P}$ , then  $\mathbb{Q}(F) = 0$  if  $\mathbb{P}(F) = 0$  (take  $X = 1_F$ ). 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 4.4. 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  $[0, T]$ .

We will use in the remainder of this section the filtration  $\mathbb{F} = \mathbb{F}^W$  and the  $\sigma$ -algebra  $\mathcal{F}_T$ . Define the random variable

$$Z(T) = \exp(-aW(T) - \frac{1}{2}a^2T). \quad (4.3)$$

Clearly we have  $Z(T) \geq 0$  and moreover  $\mathbb{E}_{\mathbb{P}}Z(T) = 1$  (see Exercise A.2). Hence we can use  $Z(T)$  to define a new probability measure  $\mathbb{Q}$  on  $\mathcal{F}_T$  by

$$\mathbb{Q}(F) = \mathbb{E}_{\mathbb{P}}1_F Z(T) \text{ for all } F \in \mathcal{F}_T. \quad (4.4)$$

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.2 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 of conditional expectation (Proposition A.6), we get

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[Z(T)|\mathcal{F}_t] &= Z(t) \mathbb{E}_{\mathbb{P}}[\exp(-a(W(T) - W(t)) - \frac{1}{2}a^2(T - t))|\mathcal{F}_t] \\ &= Z(t). \end{aligned} \quad (4.5)$$

Equation (4.5) tells us that the process  $\{Z(t) : t \in [0, T]\}$  is a martingale (w.r.t.  $\mathbb{F}$ ).

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  $\sigma$ -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 4.7. □

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

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

because

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

which follows from Exercise 4.5. Using the same arguments that led us to the result of (4.5), we can also compute  $\mathbb{E}_{\mathbb{Q}}W(t) = -at$  (Exercise 4.6). 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

$$W^{\mathbb{Q}}(t) = W(t) + at, \tag{4.6}$$

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 (4.6) is Brownian motion on the time domain  $[0, T]$  under the probability  $\mathbb{Q}$  defined by (4.4).*

**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  $W^{\mathbb{Q}}(t_k) - W^{\mathbb{Q}}(t_{k-1})$ ,  $k = 1, \dots, n$ , where  $t_0 = 0 \leq t_1, \dots, \leq t_n$ . We compute

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

by Proposition 4.1. 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

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

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

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

We finally conclude that

$$\mathbb{E}_{\mathbb{Q}} e^{i \sum \lambda_k (W^{\mathbb{Q}}(t_k) - W^{\mathbb{Q}}(t_{k-1}))} = e^{-\frac{1}{2} \sum_k \lambda_k^2 (t_k - t_{k-1})}. \quad (4.7)$$

From Equation (4.7) we deduce the independence under  $\mathbb{Q}$  of the increments of  $W^{\mathbb{Q}}$  and normality with the right parameters (see also Section A.2).  $\square$

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.  $\mathbb{P}$ ). Define the probability  $\mathbb{Q}$  on  $(\Omega, \mathcal{F}_T)$  by  $\frac{d\mathbb{Q}}{d\mathbb{P}} = Z(T) = \exp(-\gamma W_T - \frac{1}{2} \gamma^2 T)$ , 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 4.8.

**Remark 4.4** It is possible to show that the probability measure  $\mathbb{Q}$  of Corollary 4.3 is the unique probability measure such that  $X$  can be written as  $X(t) = bt + \sigma W(t)$ , where  $W$  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  $\sigma$ . Can we also make a measure transformation, such that  $\sigma$  changes into another parameter,  $\tau$  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, 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.

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?

**Proposition 4.5** *Let  $\mathbb{Q}$  be defined by (4.4) 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*

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t] = \frac{\mathbb{E}_{\mathbb{P}}[XZ(T)|\mathcal{F}_t]}{Z(t)}. \quad (4.8)$$

**Proof** We will repeatedly use properties of the conditional expectation, see Section (A.4 of the Appendix. Let  $F \in \mathcal{F}_t$ . We compute

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

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  $\mathbb{E}_{\mathbb{P}}[XZ(T)|\mathcal{F}_t] = Z(t) \mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$ .  $\square$

Note that we can write Equation (4.8), with  $Z(T)$  as in (4.3), in the equivalent form

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t] = \mathbb{E}_{\mathbb{P}}[X e^{-a(W(T)-W(t))} | \mathcal{F}_t] e^{-\frac{1}{2}a^2(T-t)}. \quad (4.9)$$

## 4.2 Exercises

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

**4.2** 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  $\mathbb{P}(Z > 0) < 1$ .

**4.3** 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)).

**4.4** Let  $\Omega = \{\omega_1, \omega_2, \dots\}$ ,  $\mathcal{F}$  the power set of  $\Omega$  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  $X$ .

**4.5** 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.

**4.6** 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  $\sigma$ -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  $d\mathbb{Q}/d\mathbb{P} = Z$ . 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.

**4.7** 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  $W$ .

**4.8** Prove Corollary 4.3.

**4.9** Let  $W$  be a Brownian motion under  $\mathbb{P}$  and define  $\mathbb{Q}$  by (4.4). Show that (4.9) follows from (4.8).

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

**4.11** Let  $0 = t_0 < t_1 < \dots < t_n = T$  and  $a(t) = \sum_{i=1}^n a_i 1_{(t_{i-1}, t_i]}(t)$ . 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  $X(t) = \int_0^t a(s) ds + \sigma W(t)$ . Let  $\Delta W_i = W(t_i) - W(t_{i-1})$  and  $Z = \exp\left(-\sum_{i=1}^n \frac{a_i}{\sigma} \Delta W_i - \frac{1}{2} \int_0^T \frac{a^2(s)}{\sigma^2} ds\right)$ . Define  $\mathbb{Q}$  by  $\frac{d\mathbb{Q}}{d\mathbb{P}} = Z$ . Argue that under  $\mathbb{Q}$  the process  $X$  becomes a Brownian motion on  $[0, T]$  with variance parameter  $\sigma^2$ .

## 5 Black Scholes market

We have established in Proposition 1.4 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 Brownian motion with drift  $X$  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*  $\mathbb{Q}$ .

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

where  $W^{\mathbb{Q}}$  is a Brownian motion under the probability measure  $\mathbb{Q}$ .

### 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

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

We have seen in the discussion of Girsanov's theorem, 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 the results of that section 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  $\sigma$ -algebra that makes the  $W(s)$  ( $s \leq T$ ) measurable functions on  $\Omega$ . For  $t \leq T$  we have a similar definition of  $\mathcal{F}_t$ . Note that  $\mathcal{F}_T$  is also the smallest  $\sigma$ -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 4.11 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  $L^2[0, T]$  and this enables one to prove that any probability measure  $\mathbb{P}$  that is such that

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

with  $a \in L^2[0, T]$  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  $d\mathbb{P}/d\mathbb{Q}$  and  $d\mathbb{Q}/d\mathbb{P}$  would look like. We come back to this point later. 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

$$V(t) = x_t S(t) + y_t B(t). \quad (5.4)$$

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

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

Let us for a while agree on the following definition (in the spirit of Exercise 1.6) 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 1.6). This is no longer true in the continuous time case (see Exercise 5.2). But still we are able to give an expression for the value process  $V$  of this 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 (1.8) and (1.9). Whereas in Section (1 we emphasized the value process, here we turn our attention to the structure of the self-financing portfolio process.

We set  $r = 0$  in the motivation below of the definition of a self-financing portfolio. 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

$$\begin{aligned} V_n(u) &= x_n S_n(u) + y_n \\ V_n(d) &= x_n S_n(d) + y_n. \end{aligned}$$

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

$$x_n = \frac{DV_n}{DS_n}. \quad (5.6)$$

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 1.6). 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

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

We consider what happens if  $u \downarrow 1$ ,  $d \uparrow 1$  so that  $u - d \rightarrow 0$ . 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  $\{0, \dots, N\}$ , 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  $n = \lfloor \frac{t}{T}N \rfloor$  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 index  $N$ . Let us now focus on the values  $V_n^N = v_n^N(S_n^N)$ . Recall 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 one 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 have 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.

**Definition 5.1** A martingale portfolio in the Black-Scholes market that is such that the value process  $V$  can be written as  $V(t) = v(t, S(t))$ , 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  $\bar{V}(t) = \bar{v}(t, \bar{S}(t))$  and 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’. One might wonder why the function  $v$  in Definition 5.1 has a partial derivative w.r.t. the second variable. 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.

**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.  $\square$

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 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 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 this claim 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, 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 definition of a self-financing 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  $e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$ .

This formula 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 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 we need 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. 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

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

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 (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 conclude this paragraph with the statement that any simple claim  $F(S(T))$  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

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

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 5.3** *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), which is beyond our purposes.

**Proof of Theorem 5.3** 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 Equation (5.9) we know that  $v(t, x) := e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[F(S(T))|S(t) = x]$  is the solution of that equation. In particular,  $v_x$  exists and we can define  $x_t = v_x(t, S(t))$ . With  $y_t = e^{-rt}(v(t, S(t)) - x_t S(t))$  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 hedging portfolio of our claim.  $\square$

## 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** Suppose that a portfolio is such that the value process is not constant and of the form  $v(t, S(t))$ . Take as an alternative portfolio the one in which  $x_t \equiv 0$  and  $y_t \equiv v(t, S(t))$ . Show that this portfolio is not self-financing.

**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.2.

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

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

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

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

where  $\Phi$  is the distribution function of the standard normal distribution. Use this result to obtain the expression for  $C_t$  in the general case.

## 6 Elementary Itô calculus

The modern theory of option pricing uses the technology and calculus of the *Stochastic (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’,  $d\bar{V}(t) = x_t d\bar{S}(t)$ . Now we are immediately faced with the question how to interpret this equation, and what is meant by a ‘differential’ like  $d\bar{V}(t)$ ? If  $\bar{V}$  would be a differentiable function, we could rely on ordinary differential calculus to give a meaning to this equation. In the present context it turns out to be the case that we don’t deal with differentiable functions  $\bar{V}$  and  $\bar{S}$ . 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. To get some feeling for this, 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  $\mathbb{P}(|(W(t+h) - W(t))/h| \geq N) = 2(1 - \Phi(N\sqrt{h})) \rightarrow 1$  as  $h \rightarrow 0$ . We cannot expect the difference quotient to have a limit in any reasonable sense.

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  $\sigma$ -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

$$\int_0^T a(s) dW(s) \tag{6.1}$$

and will be called the *stochastic 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 can be written as

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

where  $0 = t_0 < t_1 < \dots < t_n = T$  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  $\|a\| = (\mathbb{E} \int_0^T a(s)^2 ds)^{1/2}$ . Strictly speaking, this is 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  $\int_0^T a(s) dW(s)$  should be. We define it as the sum

$$\int_0^T a(s) dW(s) = \sum_{j=1}^n a_j (W(t_j) - W(t_{j-1})). \quad (6.3)$$

It is not difficult to show that this integral is independent of the chosen representation of  $a$ . For  $t < T$  we define  $\int_0^t a(s) dW(s)$  as  $\int_0^T 1_{(0,t]}(s) a(s) dW(s)$ . Note that also  $a 1_{(0,t]}$  is a simple process, if  $a$  is one. Let us write  $I(a)$  for the process defined on  $[0, T]$  by  $I_t(a) = \int_0^t a(s) dW(s)$ . 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)  $\mathbb{E} (I_t(a) - I_s(a))^2 = \mathbb{E} \int_s^t a(u)^2 du$  for  $t > s$ .
- (iii)  $\mathbb{E} [(I_t(a) - I_s(a))^2 | \mathcal{F}_s] = \mathbb{E} [\int_s^t a(s)^2 ds | \mathcal{F}_s]$ .
- (iv)  $I_T$  considered as an operator on the space of simple processes is linear and an isomorphism from  $\mathcal{S}$  with  $\|\cdot\|$  into  $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ .

**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.13. Continuity is obvious.

(ii) follows from (iii), which *you* do as an exercise 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  $\mathbb{E} I_T(a)^2 = \mathbb{E} \int_0^T a(s)^2 ds$ .  $\square$

**Example 6.2** Let the times  $t_i$  be fixed and consider the process  $W^n(t) = \sum_{i=1}^n W(t_{i-1}) 1_{(t_{i-1}, t_i]}(t)$ . Application of the definition of the integral results after some manipulation in

$$\int_0^T W^n(s) dW(s) = \frac{1}{2} W(T)^2 - \frac{1}{2} \sum_{i=1}^n (W(t_i) - W(t_{i-1}))^2. \quad (6.4)$$

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.

Here we pause for a while in our treatment to outline why we cannot use Stieltjes or Lebesgue integration theory. A main result for Stieltjes integrals is that

the integral  $\int_0^T a(s) dW(s)$  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. This trick is 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  $\int_0^T a(s) dW(s)$  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. The following proposition shows that we can approximate piecewise continuous processes by simple processes.

**Proposition 6.3** *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 and bounded. Let  $0 = t_0^n < \dots < t_n^n = T$  be a subdivision of  $[0, T]$ . Consider the approximating (adapted) simple processes  $a^n$  defined by

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

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  $\delta$  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  $a^n(t, \omega) \rightarrow a(t, \omega)$ . Boundedness allows us to invoke the dominated convergence theorem to establish  $\|a^n - a\| \rightarrow 0$ , which shows the assertion of the proposition for bounded continuous  $a$ . If  $a$  is bounded 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. Then we can finish the proof as before.

If  $a$  is not bounded we approximate  $a$  first with the bounded piecewise continuous function  $a^N = a 1_{|a| \leq N}$ . Indeed we have  $\|a - a^N\|^2 = \mathbb{E} \int_0^T |a^N(t) - a(t)|^2 dt = \mathbb{E} \int_0^T a^2(t) 1_{|a(t)| > N} dt$ , 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.  $\square$

We are now ready to define the stochastic integral  $\int_0^T a(t) dW(t)$  for processes in  $\mathcal{P}$ . Let  $a$  be such a process, and let  $a^n$  be a sequence of simple processes such that  $\|a^n - a\| \rightarrow 0$ . 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. Hence the sequence has a limit in this space. We notice 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 6.4** *For every  $a \in \mathcal{P}$  there exists a 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  $I_t(a) = I(a1_{(0,t]})$ , 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** In all what follows the  $a^n$  are processes in  $\mathcal{S}$  such that  $\|a^n - a\| \rightarrow 0$ . We will need the properties that also  $\mathbb{E}(I_t(a^n) - I_t(a))^2 \rightarrow 0$  and that  $\mathbb{E} I_t(a^n)^2 \rightarrow \mathbb{E} I_t(a)^2$  for  $t \leq T$ . 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  $\mathbb{E}(I_t(a) - I_s(a))^2 = \mathbb{E} \int_s^t a(u)^2 du$ , 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 that  $\mathbb{E} 1_A I_t(a) = \mathbb{E} 1_A I_s(a)$  for all  $A \in \mathcal{F}_s$ . We have

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

Application of the Cauchy-Schwartz inequality gives  $(\mathbb{E} 1_A (I_t(a) - I_t(a^n)))^2 \leq \mathbb{E} (I_t(a) - I_t(a^n))^2$ , which tends to zero. Similarly,  $\mathbb{E} 1_A (I_s(a^n) - I_s(a))$  has limit zero.

In much the same way we have for any  $A \in \mathcal{F}_s$

$$\begin{aligned} \mathbb{E} 1_A (I_t(a) - I_s(a))^2 &= \lim \mathbb{E} 1_A (I_t(a^n) - I_s(a^n))^2 \\ &= \lim \mathbb{E} 1_A \int_s^t a^n(u)^2 du \\ &= \mathbb{E} 1_A \int_s^t a(u)^2 du. \end{aligned}$$

That  $\mathbb{E} I_T(a)^2 = \mathbb{E} \int_0^T a(u)^2 du$  now immediately follows. Linearity of the operator  $I_T$  on  $\mathcal{P}$  follows similarly. The latter implies the last assertion of the theorem.  $\square$

Henceforth we denote the stochastic integral  $I_T(a)$  by  $\int_0^T a(s) dW(s)$  and  $I_t(a)$  by  $\int_0^t a(s) dW(s)$ . Let us consider as an example the integral  $\int_0^T W(s) dW(s)$ . Take as approximations the functions  $W^n$  of Example 6.2. Note that we have  $\|W - W^n\|^2 = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \mathbb{E} (W(t) - W(t_{i-1}))^2 dt = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (t - t_{i-1}) dt = \frac{1}{2} \sum_{i=1}^n (t_i - t_{i-1})^2 \rightarrow 0$ , if the mesh of the subdivision tends to zero. Hence to get  $\int_0^T W(s) dW(s)$  we have to compute the limit of the sum in Equation (6.4). But this we have done in Section 2.3, when we discussed Brownian motion. With this in mind we conclude

$$\int_0^T W(s) dW(s) = \frac{1}{2} W(T)^2 - \frac{1}{2} T. \quad (6.5)$$

We immediately see, that  $\mathbb{E} \int_0^T W(s) dW(s) = 0$ . Other properties of the stochastic integral, those of Theorem 6.4, can be directly verified (this is Exercise 6.2).

We also need the quadratic variation of the process  $I.(a)$ .

**Proposition 6.5** *Let  $a \in \mathcal{P}$ . The quadratic variation of  $I.(a)$  over the interval  $[0, t]$  is given by  $\int_0^t a(u)^2 du$ .*

**Proof** Consider first the case where  $a$  is piecewise constant. The result then immediately follows from Proposition 2.6. Let now  $a \in \mathcal{P}$  be arbitrary but bounded (we omit the proof of the general case). Consider partitions  $0 = t_0^n < \dots < t_n^n = t$  of  $[0, t]$  with mesh tending to zero and  $a^n$  the processes that are on each of the intervals  $(t_{k-1}^n, t_k^n]$  and equal to  $a(t_{k-1}^n)$ . Since  $\int_0^t (a^n(u))^2 du \rightarrow \int_0^t a(u)^2 du$  a.s. and in the  $L^2$ -sense ( $a$  is bounded), we show that the difference of  $Q_n = \sum_k (\int_{t_{k-1}^n}^{t_k^n} a dW)^2$  and  $\int_0^t a^n(u)^2 du$  tends to zero in the  $L^2$ -sense. Write

$$Q_n = \sum_k \left( \int_{t_{k-1}^n}^{t_k^n} (a - a^n) dW + \sum_k \int_{t_{k-1}^n}^{t_k^n} a^n dW \right)^2.$$

We write  $Q_n$  as the sum of the terms  $\sum_k (\int_{t_{k-1}^n}^{t_k^n} (a - a^n) dW)^2$ ,  $\sum_k (\int_{t_{k-1}^n}^{t_k^n} a^n dW)^2$  and the cross term  $2 \sum_k (\int_{t_{k-1}^n}^{t_k^n} (a - a^n) dW \int_{t_{k-1}^n}^{t_k^n} a^n dW)$ . The expectation of the first term tends to zero by properties of the stochastic integral. Below we will prove

$$\mathbb{E} \left( \sum_k \left( \int_{t_{k-1}^n}^{t_k^n} a^n dW \right)^2 - \int_0^t a^n(u)^2 du \right)^2 \rightarrow 0. \quad (6.6)$$

Under the assumption that this result holds, the expectation of the cross terms tends to zero by the Cauchy-Schwartz inequality. Let us therefore now focus

on (6.6). The stochastic integral and the ordinary integral in this expression are actually sums and their difference 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 (6.6) is bounded by  $C^4 \sum_k \mathbb{E}((\Delta W_k^n)^2 - \Delta t_k^n)^2$ , and we can finish as we did in the proof of Proposition 2.6.  $\square$

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

$$X = X(0) + \int_0^\cdot b(s) ds + \int_0^\cdot a(s) dW(s). \quad (6.7)$$

We need the following lemma:

**Lemma 6.6** *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  $g$ .*

**Proof** Exercise 6.3  $\square$

The consequence of this lemma and Proposition 6.5 is that the decomposition of a continuous semimartingale is unique (Exercise 6.4). Hence we can unambiguously define the stochastic integral of a process  $h \in \mathcal{P}$  w.r.t. the semimartingale  $X$  as

$$\int_0^T h(t) dX(t) = \int_0^T h(t)b(t) dt + \int_0^T h(t)a(t) dW(t), \quad (6.8)$$

where the latter is the stochastic integral, and it is well-defined under the condition  $\mathbb{E} \int_0^T h(t)^2 a(t)^2 dt < \infty$ . For Equation (6.8) we also use the shorthand notation in stochastic differentials

$$h(t)dX(t) = h(t)b(t) dt + h(t)a(t) dW(t). \quad (6.9)$$

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 other possibilities. One of them leads to the Stratonovich integral. Contrary to the Itô integral, the Stratonovich integrals w.r.t. Brownian motion are in general not martingales, see Exercise 6.8.

## 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 6.7** *Let  $f$  be a twice continuously differentiable function, such that  $\mathbb{E} \int_0^T f'(W(t))^2 dt < \infty$ . Then  $f(W)$  is a semimartingale and one has a.s.*

$$f(W(T)) = f(W(0)) + \int_0^T f'(W(t)) dW(t) + \frac{1}{2} \int_0^T f''(W(t)) dt. \quad (6.10)$$

**Proof (sketch).** Let  $0 = t_0 < \dots < t_n = T$  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  $W(t_{i-1})$

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

By properties of the stochastic integral the first term converges in  $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  to  $\int_0^T f'(W(t)) dW(t)$  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  $\int_0^T f''(W(t)) dt$ . To treat the second one, we state that we can replace  $f''(W_i^*)$  with  $f_i = f''(W(t_{i-1}))$  (without making this precise) and we use an argument like in the proof of the fact that Brownian motion has quadratic variation  $\langle W \rangle(t) = t$ . If we also assume that  $\mathbb{E} f_i^2 < \infty$ , we have  $\mathbb{E} (\sum_{i=0}^{n-1} f_i (\Delta W_i^2 - \Delta_i))^2 = 2 \sum_{i=1}^n \mathbb{E} f_i^2 \Delta_i^2$ , which tends to zero, for partitions whose mesh tend to zero. This finishes the sketch proof.  $\square$

Lemma 6.7 can be extended to functions  $f$  of two variables.

**Lemma 6.8** *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  $\mathbb{E} \int_0^T f_x(t, W(t))^2 dt < \infty$ . Then one has a.s.*

$$\begin{aligned} f(T, W(T)) &= f(0, W(0)) + \int_0^T f_t(t, W(t)) dt \\ &\quad + \int_0^T f_x(t, W(t)) dW(t) + \frac{1}{2} \int_0^T f_{xx}(t, W(t)) dt. \end{aligned} \quad (6.11)$$

**Proof** Use the same technique as in the proof of Lemma 6.7. We omit the details.  $\square$

For Equation (6.11) we use, similar to (6.9) the abbreviation

$$df(t, W(t)) = f_t(t, W(t)) dt + f_x(t, W(t)) dW(t) + \frac{1}{2} f_{xx}(t, W(t)) dt. \quad (6.12)$$

**Remark 6.9** The conditions involving finite expectations in lemmas 6.7 and 6.8 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 6.10** *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*

$$dY(t) = g_t(t, X(t)) dt + g_x(t, X(t)) dX(t) + \frac{1}{2} g_{xx}(t, X(t)) f_x(t, W(t))^2 dt,$$

where the term with  $dX(t)$  has to be understood as a shorthand notation for an expression like (6.9).

**Proof** Let  $h$  be the decomposition  $g \circ f$ . Compute the partial derivatives of  $h$ , express them in those of  $f$  and  $g$  and use Equation (6.9).  $\square$

Now we have seen in Theorem 6.10 that functions and compositions of functions of time and Brownian motion are again semimartingales, we can use Lemma 6.6 to compute the quadratic variation of  $X$  and  $Y$  as they appear in this theorem.

**Proposition 6.11** *Let  $X$  and  $Y$  be as in Theorem 6.10. Then  $X$  has quadratic variation process given by  $\langle X \rangle_t = \int_0^t f_x(s, W(s))^2 ds$  and  $Y$  has quadratic variation given by  $\langle Y \rangle_t = \int_0^t h_x(s, W(s))^2 ds$ , where  $h = g \circ f$ . Moreover these quadratic variations are related by  $\langle Y \rangle_t = \int_0^t g_x(s, X(s))^2 d\langle X \rangle_s$ .*

**Proof** The proof of the expression for  $\langle X \rangle_t$  goes along the same lines as the proof of the convergence of the quadratic variation term in Lemma 6.7, as well as the proof of the expression for  $\langle Y \rangle_t$ . The relation between the two quadratic variation processes then follows.  $\square$

**Corollary 6.12** *In the situation of Theorem 6.10 we have*

$$dY(t) = g_t(t, X(t)) dt + g_x(t, X(t)) dX(t) + \frac{1}{2} g_{xx}(t, X(t)) d\langle X \rangle_t. \quad (6.13)$$

**Proof** Just use the expression for  $\langle X \rangle_t$  of Proposition 6.11.  $\square$

Corollary 6.12 can be shown to be generalized to the situation, where  $Y(t) = g(A(t), X(t))$  with  $A$  a process given by  $A(t) = \int_0^t a(s) ds$ , where  $a$  is a continuous process. We just state, rather informally, the result.

**Proposition 6.13** *Let  $X$  be a semimartingale,  $A$  a process of the form  $A(t) = \int_0^t a(s) ds$  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*

$$\begin{aligned} dY(t) &= g_t(A(t), X(t))a(t) dt \\ &+ g_x(A(t), X(t)) dX(t) + \frac{1}{2} g_{xx}(A(t), X(t)) d\langle X \rangle_t. \end{aligned} \quad (6.14)$$

**Example 6.14** We can apply Proposition 6.13 to the following interesting case. Let  $X(t)$  be of the form  $X(t) = f(t, W(t))$ ,  $A(t) = \int_0^t f_x(s, W(s))^2 ds$ ,  $g(t, x) = \exp(x - \frac{1}{2}t)$  and  $Y(t) = g(A(t), X(t))$ . Then we have (check it yourself!)

$$Y(t) = 1 + \int_0^t Y(s) dX(s). \quad (6.15)$$

For processes  $X$  and  $Y$  that are related by Equation (6.15) we use the notation  $Y = \mathcal{E}(X)$ .  $Y$  is said to be the Doléans exponential of  $X$ .

The Itô rules that we have given above, in particular (6.14) 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. Then  $Y$  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

$$\begin{aligned} dY(t) &= f_t dt + f_{x_1} dX_1(t) + f_{x_2} dX_2(t) \\ &+ \frac{1}{2} (f_{x_1 x_1} d\langle X_1 \rangle + 2f_{x_1 x_2} d\langle X_1, X_2 \rangle + f_{x_2 x_2} d\langle X_2 \rangle). \end{aligned} \quad (6.16)$$

We notice that we have seen already an instance of this formula, namely (6.14). You check why! One special case deserves our attention.

Consider  $f(t, x_1, x_2) = x_1 x_2$ . Then Equation (6.16) becomes the product formula for semimartingales. Omitting the time index, we get

$$d(X_1 X_2) = X_1 dX_2 + X_2 dX_1 + d\langle X_1, X_2 \rangle. \quad (6.17)$$

This equation can be specialized further. Let  $X_1 = X$  be an arbitrary semimartingale and  $X_2 = A$  a semimartingale of the form  $A = \int_0^\cdot a(t) dt$ . then we have

$$d(X(t)A(t)) = X(t)a(t) dt + A(t) dX(t). \quad (6.18)$$

Check yourself that the result (6.15) also follows from (6.18).

### 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 4.11 an example was given of a change measure that transformed Brownian motion in a Brownian motion with piecewise linear drift (and vice versa). In this exercise we looked at the exponential of  $-\sum \frac{a_i}{\sigma} \Delta W_i - \frac{1}{2} \int_0^T a(s)^2 ds$ . For  $\sigma = -1$  the first sum we now understand as  $\int_0^T a(s) dW(s)$ . Having a closer look at this exercise, we suspect that any random variable  $Z = \exp(\int_0^T a(s) dW(s) - \frac{1}{2} \int_0^T a(s)^2 ds)$  with  $a \in \mathcal{P}$  can be used to define a new probability measure  $\mathbb{Q}$  with  $\frac{d\mathbb{Q}}{d\mathbb{P}} = Z$ . 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  $d\mathbb{Q} = Z(T)d\mathbb{P}$ . But to have that  $\mathbb{Q}$  is a probability measure, we need  $\mathbb{E} Z(T) = 1$ . If  $a$  is a non-random function in  $L^2[0, T]$ , this follows from Exercise 6.6. 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  $a$ , so  $a \in \mathcal{S}$  and let us look at  $Z(t) = \exp(\int_0^t a(s) dW(s) - \frac{1}{2} \int_0^t a(s)^2 ds)$ . Application of the Itô rule gives

$$Z(t) = 1 + \int_0^t Z(s)a(s) dW(s). \quad (6.19)$$

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  $\mathbb{E} Z(T)^2 < \infty$ . Observe by the way, that we can use the Doléans exponential to write  $Z = \mathcal{E}(Y)$ , with  $Y = \int_0^\cdot a(s) dW(s)$ .

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

$$Z(t) = \exp\left(\int_0^t a(s) dW(s) - \frac{1}{2} \int_0^t a(s)^2 ds\right).$$

Without checking technical conditions, we simply apply the Itô rule again, to get the same representation (6.19). It is in this case not clear whether  $Z$  is a martingale. For instance, if we have functions  $a^n$  that converge to  $a$  w.r.t. the norm of Section 6.1 we have  $L^2$ -convergence of the resulting stochastic integrals  $I_T(a^n)$  to  $I_T(a)$ . With  $Z^n = \mathcal{E}(a^n)$ , 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 (6.19). To have that  $Z$  is a martingale, it is sufficient to impose that  $aZ$  belongs to  $\mathcal{P}$  (we then also have  $\mathbb{E}_{\mathbb{P}} \int_0^T Z(s)^2 a(s)^2 ds < \infty$ ). 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  $L^2[0, T]$ , or if  $a$  is simple.

**Theorem 6.15** *Let  $a \in \mathcal{P}$ , in particular  $\mathbb{E}_{\mathbb{P}} \int_0^T a(t)^2 dt < \infty$  and let  $Z(t) = \exp\left(\int_0^t a(s) dW(s) - \frac{1}{2} \int_0^t a(s)^2 ds\right)$ . Assume that  $\mathbb{E}_{\mathbb{P}} Z(T)^2 < \infty$ . Define the measure  $\mathbb{Q}$  on  $\mathcal{F}_T$  by  $d\mathbb{Q} = Z(T) d\mathbb{P}$ . Then  $\mathbb{Q}$  is a probability measure and the process  $W^{\mathbb{Q}}$  defined by  $W^{\mathbb{Q}}(t) = W(t) - \int_0^t a(s) ds$  is a Brownian motion under  $\mathbb{Q}$ .*

**Proof** We give the proof only under the restricting condition that  $a$  is bounded process,  $a \leq K$ , for some  $K > 0$ . First we show that  $\mathbb{Q}$  is a probability measure, which follows as soon as we know that  $\mathbb{E}_{\mathbb{P}} Z(T) = 1$ . Consider the sequence of  $a^n \in \mathcal{S}$  that converges to  $a$ , which we took in the proof of Proposition 6.3, then  $a^n < K$  for all  $n$ . Moreover  $\int_0^T |a^n(s) - a(s)| ds$  converges to zero  $\mathbb{P}$ -a.s. Put  $Z^n(t) = \exp\left(\int_0^t a^n(s) dW(s) - \frac{1}{2} \int_0^t a^n(s)^2 ds\right)$ . 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  $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ . It then follows that  $\mathbb{E}_{\mathbb{P}} Z(T) = 1$ .

From Equation (6.19) and the analogous expression for  $Z^n$  we obtain

$$\begin{aligned} Z^n(T) - Z(T) &= \int_0^T (a^n(t)Z^n(t) - a(t)Z(t)) dW(t) \\ &= \int_0^T (a^n(t) - a(t))Z(t) dW(t) \\ &\quad + \int_0^T (Z^n(t) - Z(t))a^n(t) dW(t). \end{aligned}$$

Hence we get

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} |Z^n(T) - Z(T)|^2 &\leq 2\mathbb{E}_{\mathbb{P}} \int_0^T |a^n(t) - a(t)|^2 Z(t)^2 dt \\ &\quad + 2\mathbb{E}_{\mathbb{P}} \int_0^T |Z^n(t) - Z(t)|^2 a^n(t)^2 dt \\ &\leq 2\mathbb{E}_{\mathbb{P}} \int_0^T |a^n(t) - a(t)|^2 Z(t)^2 dt \\ &\quad + 2K^2 \mathbb{E}_{\mathbb{P}} \int_0^T |Z^n(t) - Z(t)|^2 dt. \end{aligned}$$

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

$$F^n(T) \leq A^n(T) + B \int_0^T F^n(t) dt. \quad (6.20)$$

It follows from Gronwall's inequality (Exercise 6.10) that we can deduce that  $F^n(T) \leq A^n(T) \exp(BT)$ . So to achieve our aim,  $\mathbb{E}_{\mathbb{P}} |Z^n(T) - Z(T)|^2 \rightarrow 0$ , it is sufficient to show that  $A^n$  converges to zero, and this follows from the dominated convergence theorem.

Recall from Exercise 4.11 that under  $\mathbb{Q}^n$  we have that  $W^n$  defined by  $W^n(t) = W(t) + \int_0^t a^n(s) ds$  is a Brownian motion. This means that any event of the type  $E^n = \{W^n(t_1) \leq x_1, \dots, W^n(t_m) \leq x_m\}$  has a stationary  $\mathbb{Q}^n$ -probability, i.e.  $\mathbb{Q}^n(E^n)$  is the same for all  $n$ . We will show that  $\mathbb{Q}(E)$  is equal to  $\mathbb{Q}^n(E^n)$ .

Consider

$$\begin{aligned} |\mathbb{Q}(E) - \mathbb{Q}^n(E^n)| &= \mathbb{E}_{\mathbb{P}} |Z 1_E - Z^n 1_{E^n}| \\ &\leq \mathbb{E}_{\mathbb{P}} |Z(1_E - 1_{E^n})| + \mathbb{E}_{\mathbb{P}} |Z^n - Z|. \end{aligned} \quad (6.21)$$

The second term in (6.21) tends to zero, because of the just shown convergence of  $Z^n(T)$  to  $Z(T)$ . Since we the  $1_{E^n}$  converge to  $1_E$   $\mathbb{P}$ -a.s., we use the dominated convergence theorem to conclude that also the first term in (6.21) eventually vanishes. This finishes the proof of the theorem, the conclusion is that  $W^{\mathbb{Q}}$  given by  $W^{\mathbb{Q}}(t) = W(t) + \int_0^t a(s) ds$  is a Brownian motion under  $\mathbb{Q}$ .  $\square$

## 6.4 Exercises

**6.1** Prove the third assertion of Proposition 6.1

**6.2** Show by direct computation that the process  $I(t) = \int_0^t W(s) dW(s)$  satisfies all properties of Theorem 6.4.

**6.3** Prove Lemma 6.6.

**6.4** Use Lemma 6.6 to show that the semimartingale decomposition (6.7) is unique.

**6.5** Compute from one of the Itô rules  $dZ_t$  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  $X_t = W_t^2 - t$ .

**6.6** Let  $\sigma$  be a piecewise continuous function on any interval  $[0, t]$ . Let  $W$  be Brownian motion and put  $X(t) = \int_0^t \sigma(s) dW(s)$ . Let  $Z(t) = \exp(iuX(t))$ , for fixed  $u \in \mathbb{R}$ . Use the Itô rule to find

$$Z(t) = 1 + \int_0^t iuZ(s) dW(s) - \frac{1}{2}u^2 \int_0^t \sigma(s)^2 Z(s) ds.$$

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

**6.7** Let  $X$  be given by  $X(t) = \exp(\gamma t + \sigma W(t))$ , with  $\gamma$  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  $\sigma^2 \int_0^t X(u)^2 du$ .

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

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

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

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

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

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  $J(\alpha)$ .

**6.9** 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 and define for every  $\lambda \in \mathbb{R}$  and  $s \leq t$  the random variables  $\phi(s, t) = \mathbb{E}[e^{i\lambda(W(t) - W(s))} | \mathcal{F}_s]$ . We have reached our goal as soon as we can prove  $\phi(s, t) = e^{-\frac{1}{2}\lambda^2(t-s)}$  (why?). Let  $Y(t) = e^{i\lambda W(t)}$ . Use the Itô rule to get

$$Y(t) - Y(s) = i\lambda \int_s^t e^{i\lambda W(u)} dW(u) - \frac{1}{2}\lambda^2 \int_s^t e^{i\lambda W(u)} ds.$$

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

**6.10** Let  $f : [0, T] \rightarrow [0, \infty)$  satisfy the inequality  $f(t) \leq a(t) + \beta \int_0^t f(s) ds$ , with  $a$  an integrable function and  $\beta \geq 0$ . Then  $f(t) \leq a(t) + \beta \int_0^t a(s)e^{\beta(t-s)} ds$  for all  $t \leq T$ . If the function  $a$  is constant, say  $\alpha$ , then  $f(t) \leq \alpha e^{\beta t}$ .

**6.11** 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,  $X$  has bounded quadratic variation and  $Y$  has bounded variation over  $[0, T]$ .

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

## 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 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 6.8, we obtain the following stochastic differential representation of  $S$

$$dS(t) = rS(t) dt + \sigma S(t) dW^{\mathbb{Q}}(t). \quad (7.1)$$

We could do the same for the representation of  $S$  under the probability measure  $\mathbb{P}$  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

$$dS(t) = a(t)S(t) dt + \sigma S(t) dW^{\mathbb{P}}(t). \quad (7.2)$$

We use Girsanov's theorem 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  $d\mathbb{P}/d\mathbb{Q}$  and  $d\mathbb{Q}/d\mathbb{P}$  (as always,  $\mathbb{Q}$  and  $\mathbb{P}$  restricted to  $\mathcal{F}_T$ ). Indeed, with

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\int_0^T \frac{r - a(t)}{\sigma} dW^{\mathbb{P}}(t) - \frac{1}{2} \int_0^T \left(\frac{r - a(t)}{\sigma}\right)^2 dt\right), \quad (7.3)$$

and

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \exp\left(\int_0^T \frac{a(t) - r}{\sigma} dW^{\mathbb{Q}}(t) - \frac{1}{2} \int_0^T \left(\frac{a(t) - r}{\sigma}\right)^2 dt\right), \quad (7.4)$$

we have the relation  $W^{\mathbb{P}}(t) = W^{\mathbb{Q}}(t) - \int_0^t \frac{a(s) - r}{\sigma} ds$ . Of course we take  $a$  such that the stochastic integrals in (7.3) and in (7.4) are well defined. 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  $dt$  differ.

In Section 1.2 we also considered the cumulative return process. We found in Proposition 1.5 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 suggests 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

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

Note that the Radon-Nikodym derivatives in (7.3) and in (7.4) transform the two representations for  $R$  into 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)

$$dS(t) = S(t) dR(t). \quad (7.5)$$

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 (1.6).

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

$$dB(t) = rB(t) dt. \quad (7.6)$$

We observe the similarity of the  $dt$  terms in equations (7.1) and (7.6). 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. Therefore  $\mathbb{Q}$  is also called the risk neutral probability measure.

## 7.2 Self-financing portfolios and hedging

Consider a portfolio process  $(x_t, y_t)$  with value process  $V$ , so  $V(t) = x_t S(t) + y_t B(t)$ . 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  $V$ .

**Definition 7.1** A portfolio process is called self-financing if it satisfies the integrability condition  $\mathbb{E}_{\mathbb{Q}} \int_0^T x(u)^2 S(u)^2 du < \infty$  and if we have for all  $t \in [0, T]$   $\mathbb{Q}$ -almost surely  $V(t) = V(0) + \int_0^t x_u dS(u) + \int_0^t y_s dB(s)$ .

Like in discrete time, it is possible to show that this definition can equivalently be stated in terms of discounted processes. We have

**Proposition 7.2** A portfolio process is self-financing iff we have for all  $t \in [0, T]$  almost surely  $\bar{V}(t) = V(0) + \int_0^t x_u d\bar{S}(u)$ . A self-financing portfolio process is therefore such that the process  $\bar{V}$  is a martingale under  $\mathbb{Q}$ .

This proposition tells us that the discounted value process of a self-financing portfolio is a martingale and  $\mathbb{E}_{\mathbb{Q}}\bar{V}(T)^2 < \infty$ . 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 7.1, if we restrict the class of martingales under consideration.

**Proposition 7.3** *Let a process  $\bar{V}$  be a martingale under  $\mathbb{Q}$  and suppose that there exists a function  $v$  such that  $\bar{V}(t) = v(t, \bar{S}(t))$ . Then the portfolio defined by  $x_t = v_x(t, \bar{S}(t))$  and  $y_t = \bar{V}(t) - x_t \bar{S}(t)$  is self-financing and its discounted value process is just  $\bar{V}$ .*

**Proof** From Section 5 we know that the function  $v$  is a solution to the partial differential equation  $v_t(t, x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t, x) = 0$ . Now we apply Theorem 6.10 to  $\bar{V}$ . We obtain

$$d\bar{V}(t) = v_t(t, \bar{S}(t)) dt + v_x(t, \bar{S}(t)) d\bar{S}(t) + \frac{1}{2}\sigma^2 \bar{S}(t)^2 v_{xx}(t, \bar{S}(t)) dt.$$

In this equation all the terms with  $dt$  vanish. Hence we are left with  $d\bar{V}(t) = v_x(t, \bar{S}(t)) d\bar{S}(t)$ . So, choosing  $x_t$  as  $v_x(t, \bar{S}(t))$  and  $y_t$  as  $y_t = \bar{V}(t) - x_t \bar{S}(t)$ , we get the desired result.  $\square$

Proposition 7.3 has as an immediate consequence that every simple claim can be hedged.

**Corollary 7.4** *Every simple claim  $X = F(S(T))$  with  $\mathbb{E}_{\mathbb{Q}}|X| < \infty$  can be hedged.*

**Proof** We have seen before that there exists a function  $v$  such that the martingale  $\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$  can be written as  $v(t, S(t))$ . Since  $v(T, S(T)) = F(S(T))$  the proof is complete.  $\square$

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  $(\frac{1}{T} \int_0^T S(t) dt - K)^+$ .

Let  $Z(t) = \int_0^t g(u, S(u)) du$  with some given function  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  and assume that  $Z(t)$  is well defined for all  $t$  and consider the claim  $F(S(T), Z(T))$ . 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. We thus have  $\bar{V}(t) = \mathbb{E}_{\mathbb{Q}}[\bar{F}(S(T), Z(T))|\mathcal{F}_t]$ . It would be nice if we could write  $\bar{V}(t) = v(t, S(t), Z(t))$  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  $F(S(T), \int_0^T g(u, S(u)) du)$ . We decompose as follows.  $S(T) = \frac{S(T)}{S(t)} S(t)$  and  $\int_0^T g(u, S(u)) du = \int_0^t g(u, S(u)) du + \int_t^T g(u, S(u)) du$ .

Given  $\mathcal{F}_t$  we completely know in the latter expression the first summand, which is just  $Z(t)$ . The second summand we write as  $\int_t^T g(u, \frac{S(u)}{S(t)} S(t)) du$ . Now we exploit the independence of  $\frac{S(u)}{S(t)}$  and the whole past of  $S$  up to  $t$ . Carrying out the conditional expectation with values  $S(t) = s$  and  $Z(t) = z$  then leaves us with the unconditional expectation

$$\mathbb{E}_{\mathbb{Q}} F\left(\frac{S(T)}{S(t)} s, z + \int_t^T g(u, \frac{S(u)}{S(t)} s) du\right),$$

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

**Theorem 7.5** *Let  $v$ , a sufficiently smooth function of  $t, x, z$  be the solution of the partial differential equation*

$$\begin{aligned} v_t(t, x, z) + xr v_x(t, x, z) \\ + \frac{1}{2} \sigma^2 x^2 v_{xx}(t, x, z) + g(t, x) v_z(t, x, z) - rv(t, x, z) = 0, \end{aligned} \quad (7.7)$$

with boundary condition  $v(T, x, z) = F(x, z)$ . The portfolio consisting of  $x_t = v_x(t, S(t), Z(t))$  and  $y_t = (v(t, S(t), Z(t)) - x_t S(t)) e^{-rt}$  is self-financing and hedges the claim  $F(S(T), Z(T))$ .

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

$$\begin{aligned} dV(t) &= v_t dt + v_x dS(t) + v_z dZ(t) + \frac{1}{2} v_{xx} d\langle S \rangle_t \\ &= v_t dt + v_x r S(t) dt + v_x \sigma S(t) dW(t) + v_z g(t, S(t)) dt \\ &\quad + \frac{1}{2} v_{xx} \sigma^2 S(t)^2 dt \\ &= rV(t) dt + v_x \sigma S(t) dW(t), \end{aligned}$$

where we used in the last equality that  $v$  was a solution of (7.7). Hence the discounted value process  $\bar{V}$ , recall that  $\bar{V}(t) = e^{-rt} V(t)$ , satisfies

$$d\bar{V}(t) = -r\bar{V}(t) dt + e^{-rt} dV(t) \quad (7.8)$$

$$= e^{-rt} v_x \sigma S(t) dW(t) \quad (7.9)$$

$$= v_x d\bar{S}(t). \quad (7.10)$$

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  $d\bar{V}(t) = x_t d\bar{S}(t)$ , so that the portfolio is self-financing. From the boundary condition on  $v$  we obtain  $v(T, S(T), Z(T)) = F(S(T), Z(T))$ , this portfolio is a hedging portfolio for the claim  $F(S(T), Z(T))$ .  $\square$

As a corollary to Theorem 7.5 we get

**Corollary 7.6** *Every claim  $F(S(T), Z(T))$  that satisfies an appropriate integrability condition can be hedged. The hedging portfolio is given in Theorem 7.5.*

We have seen above 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. A most effective one is to demand that a claim  $X$  has finite second moment under  $\mathbb{Q}$ . In that case we can apply a non-trivial result in probability, called the martingale representation theorem. It states that it is possible to represent the claim as  $X = \int_0^T \phi(s) dW^{\mathbb{Q}}(s)$ , for a process  $\phi$  satisfying  $\mathbb{E}_{\mathbb{Q}} \int_0^T \phi(s)^2 ds < \infty$ . The hedging strategy in this case is given by  $x_t = \frac{\phi(t)}{\sigma S(t)}$  and  $y_t = \int_0^t \phi(s) dW^{\mathbb{Q}}(s) - \frac{\phi(t)}{\sigma} \bar{S}(t)$ . In general, the process  $\phi$  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  $x_t$  as 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  $\frac{DV_n}{DS_n}$ , 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  $x_t = \frac{DV}{DS}(t)$ .

### 7.3 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  $\tau$  may assume any value between 0 and  $T$  and this moment may depend on the random evolution of the stock price,  $\tau$  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  $t \leq T$  an event  $\{\tau \leq t\}$  depends on the values of  $S$  at times  $s \leq t$  only. In technical terms, the event  $\{\tau \leq t\}$  is  $\mathcal{F}_t$ -measurable, and when this is true for all  $t \leq T$ , we call  $\tau$  a stopping time. 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, that seems profitable. The problem here becomes to find an optimal strategy that tells an investor when to exercise the claim. Problems of this type—we could also look at other claims than the American call—are hard to solve. But in this special case 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 this American claim is thus the same as that of the European call option with same maturity time and same strike price, although the owner of such a 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  $C(t)$ . Clearly we should have  $C_A(t) \geq C(t)$  for all  $t$ . Consider a market with interest rate  $r > 0$ . We compare two financial products. 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  $C(t)$ . We compare the values of the two products at time  $T$ . They are  $S(T) - K$  and  $C(T) = (S(T) - K)^+$ , respectively and we have the obvious inequality  $S(T) - K \leq C(T)$ . But then this inequality between the two values is preserved at any time  $t < T$  and we get  $S(t) - Ke^{-r(T-t)} \leq C(t)$ . Hence we also have the strict inequality  $S(t) - K < C(t) \leq C_A(t)$ . Now suppose the owner of an American call wants to exercise his option at the chosen time  $\tau = t$ , 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. 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  $C(t)$  of the European claim,  $C_A(t) = C(t)$ .

This surprising result has no counterpart for the American put-option. It is possible to show that the value of an American put is 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 7.8).

## 7.4 The Greeks

Starting point of this section is the Black-Scholes formula (1.27) 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  $C(t, s, r, \sigma)$  for the expression (1.27) and introduce the following.

- (a)  $\Delta = \frac{\partial C}{\partial s}$  (*delta*)
- (b)  $\Gamma = \frac{\partial^2 C}{\partial s^2}$  (*gamma*)
- (c)  $\rho = \frac{\partial C}{\partial r}$  (*rho*)
- (d)  $\Theta = \frac{\partial C}{\partial t}$  (*theta*)
- (e)  $\mathcal{V} = \frac{\partial C}{\partial \sigma}$  (*vega*)

Of course the ‘Greeks’ in turn depend on all parameters as well, see Proposition 7.7, 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 7.7** *Let  $\phi$  be the density of the standard normal distribution. The following hold.*

$$\begin{aligned}\Delta &= \Phi(d_1(t)) \\ \Gamma &= \frac{\phi(d_1(t))}{s\sigma\sqrt{T-t}} \\ \rho &= K(T-t)e^{-r(T-t)}\Phi(d_2(t)) \\ \Theta &= -\frac{s\phi(d_1(t))\sigma}{2\sqrt{T-t}} - rKe^{-r(T-t)}\Phi(d_2(t)) \\ \mathcal{V} &= s\phi(d_1(t))\sqrt{T-t}.\end{aligned}$$

**Proof** Exercise 7.9. □

For a number of reasons it may be attractive to manage portfolios that are not sensitive for, also called *neutral* to one or more parameters. 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 parameters) and we wish to make it  $\Delta$ -neutral by addition 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 a basis for a trading strategy, is not attractive, since one has to rebalance the portfolio at every time to keep it  $\Delta$ -neutral. Moreover, if the delta of the portfolio itself is subject to big changes, a big value for  $|\Gamma|$ , the result at any time will be major adjustments of the portfolio ( $z$  depends on  $t$ ). In such a situation, one may extend the portfolio with yet another product in order to accomplish  $\Gamma$ -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)

$$z = -\frac{\Gamma_V}{\Gamma_D}$$

$$w = \frac{\Gamma_V}{\Gamma_D} \Delta_D - \Delta_V.$$

And the story doesn't end here ...

## 7.5 Exercises

**7.1** Consider the Black-Scholes market with claim  $\int_0^T S(u) du$ . Show that the fair price of this claim at time  $t$  is given by  $V(t)$  with  $V(t) = e^{r(t-T)} Z(t) + S(t) \frac{1-e^{r(t-T)}}{r}$ , where  $Z(t) = \int_0^t S(u) du$  and  $r \neq 0$  the interest rate. What is  $V(t)$  if  $r = 0$ ?

Write  $V(t)$  as  $v(t, S_t, Z_t)$  and determine a partial differential equation for  $v$ . Find the hedge strategy for this claim.

**7.2** 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  $C_t - P_t = S_t - e^{r(t-T)} K$ .

**7.3** Consider in the Black-Scholes market a self-financing portfolio  $(x, y)$  with value process  $V$ , so

$$dV(t) = x_t dB(t) + y_t dS(t).$$

Show by using the Itô formula that

$$d\bar{V}(t) = y_t d\bar{S}(t),$$

with  $\bar{S}$  the discounted stock price and  $\bar{V}$  the discounted value process.

**7.4** 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

$$dV(t) = rV(t)dt + g(t) dW^{\mathbb{Q}}(t).$$

**7.5** Consider in a Black-Scholes market the claim  $\Phi(S_T) = K1_{(a,b)}(S_T)$  (this is called a binary spread). Determine for each time  $t \leq T$  the fair price of this claim.

**7.6** 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. Find this portfolio.

**7.7** Consider in the Black-Scholes market a *bull-spread*. This is a claim with payoff  $\min\{\max\{S(T), A\}, B\}$ , where  $B > A > 0$ . Like a straddle (Exercise 7.6), 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.

**7.8** 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.

**7.9** First show

$$s\phi(d_1(t)) - Ke^{-r(T-t)}\phi(d_2(t)) = 0,$$

and use this to derive the expressions for the Greeks in Proposition 7.7.

## A Some results in Probability and in Analysis

### A.1 Normal random variables

If  $X$  is a normally distributed real random variable with mean  $\mu$  and variance  $\sigma^2 > 0$ , then by definition it has density  $p$  given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

An extension of this formula holds for the case where  $X = (X_1, \dots, X_n)^\top$  is a random vector in  $\mathbb{R}^n$ . If  $\mathbb{E} X = \mu \in \mathbb{R}^n$  and  $X$  has covariance matrix  $\text{Cov}(X) = \Sigma$ , which is assumed to be strictly positive definite, then by definition  $X$  has a density on  $\mathbb{R}^n$  given by

$$p(x) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right). \quad (\text{A.1})$$

A random vector  $X$  with such a density is said to have a nondegenerate multivariate normal distribution. If the matrix  $\Sigma$  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 rank, then  $Y = AX + b$  is again multivariate normal with  $\mathbb{E} Y = A\mu + b$  and  $\text{Cov}(Y) = A\Sigma A^\top$ . Note that  $\text{Cov}(Y)$  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  $\text{Cov}(X^1, X^2) = 0$ , then  $X^1$  and  $X^2$  are independent random vectors.

Here is an important consequence of this last property. Let  $X^1$  be a subvector of  $X$  with covariance matrix  $\Sigma_1$ . Then we have  $Z := X^2 - \text{Cov}(X^2, X^1)\Sigma_1^{-1}X^1$  is *independent* of  $X^1$ . Hence the conditional expectation (see Section A.4 for a definition and some properties)  $\mathbb{E}[X^2|X^1]$  is equal to  $\text{Cov}(X^2, X^1)\Sigma_1^{-1}X^1 + \mathbb{E} Z = \mathbb{E} X^2 + \text{Cov}(X^2, X^1)\Sigma_1^{-1}(X^1 - \mathbb{E} X^1)$ .

Assume that  $\mathbb{E} X^1$  and  $\mathbb{E} X^2$  are zero. Then  $\mathbb{E}[X^2|X^1] = \text{Cov}(X^2, X^1)\Sigma_1^{-1}X^1$  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.2 Characteristic functions

Let  $X$  be a random vector in  $\mathbb{R}^n$  ( $n \geq 1$ ). The characteristic function of  $X$  is  $\lambda \mapsto \mathbb{E} \exp(i\lambda^\top X)$ . 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.3. 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  $\phi$  be its characteristic function. Then

$$\phi(\lambda) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(i\lambda x - \frac{1}{2}x^2) dx.$$

With methods from complex integration it is not difficult to show that  $\phi(\lambda) = \exp(-\frac{1}{2}\lambda^2)$ . Knowing this, it is easy to show that the characteristic function of  $N(\mu, \sigma^2)$  distributed random variable is given by  $\phi(\lambda) = \exp(i\mu\lambda - \frac{1}{2}\sigma^2\lambda^2)$ . But then it is immediate that for this case

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-i\lambda x) \phi(\lambda) d\lambda. \tag{A.2}$$

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  $L^1(\mathbb{R})$ .

Other properties that we use are the following.  $X$  and  $Y$  are independent random vectors iff  $\mathbb{E} \exp(i\lambda^\top X + i\mu^\top Y) = \mathbb{E} \exp(i\lambda^\top X) \mathbb{E} \exp(i\mu^\top Y)$ . 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  $\mathbb{E} \exp(i\lambda X + i\mu Y) = \exp(-\frac{1}{2}(\sigma^2\lambda^2 + \tau^2\mu^2))$ . Another property concerns derivatives of characteristic functions. Let  $X$  be a real random variable and let  $\phi$  be its characteristic function. If  $\mathbb{E}|X|^k < \infty$ , then  $\phi$  is  $k$  times differentiable and  $\phi^{(k)}(0) = i^k \mathbb{E} X^k$ .

### A.3 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. Consider a sequence  $Z_1, Z_2, \dots$  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  $\mathbb{E} f(Z_n) \rightarrow \mathbb{E} f(Z)$ . 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.1** *A sequence  $Z_1, Z_2, \dots$  of random variables converges in distribution to a random variable  $Z$  iff for all bounded continuous functions  $h$  the convergence  $\mathbb{E} h(Z_n) \rightarrow \mathbb{E} h(Z)$  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  $\mathbb{P}(Z_n \leq x) = \mathbb{E} 1_{(-\infty, x]}(Z_n)$ . Let  $\varepsilon > 0$  be given and let  $h^\varepsilon$  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  $h_\varepsilon \leq 1_{(-\infty, x]} \leq h^\varepsilon$  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  $\varepsilon \rightarrow 0$  we obtain  $\limsup_{n \rightarrow \infty} \mathbb{P}(Z_n \leq x) \leq \mathbb{P}(Z \leq x)$  and  $\liminf_{n \rightarrow \infty} \mathbb{P}(Z_n \leq x) \geq \mathbb{P}(Z < x)$ . 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  $\mathbb{E} h(Z_n) = \sum_k \mathbb{E} h(Z_n) 1_{I_k^N}(Z_n) \leq \sum_k \bar{h}_k^N \mathbb{P}(L_k^N < Z_n \leq R_k^N) \rightarrow \sum_k \bar{h}_k^N \mathbb{P}(L_k^N < Z \leq R_k^N)$  as  $n \rightarrow \infty$ . 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  $\varepsilon \rightarrow 0$  to finish the proof for continuous functions with compact support. Use this result to prove the assertion for arbitrary bounded continuous  $h$  (Exercise A.8).  $\square$

The most important example of weak convergence is the content of the Central limit theorem:

**Theorem A.2** *Let  $X_1, X_2, \dots$  be a sequence of independent and identically distributed random variables with mean  $\mu$  and variance  $\sigma^2 < \infty$ . Then for all  $x \in \mathbb{R}$  we have*

$$\mathbb{P}\left(\frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} \leq x\right) \rightarrow \Phi(x),$$

as  $n$  tends to  $\infty$ , where  $\Phi$  is the distribution function of the standard normal distribution.

To connect the statement of Theorem A.2 with the introduction of this section it is sufficient to take  $Z_n = \sum_{i=1}^n (X_i - \mu)/\sigma\sqrt{n}$ , since  $\Phi$  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.2.

**Proof of Theorem A.2.** Without loss of generality we assume that  $\mathbb{E} X_k = 0$ . We start off with some facts from analysis.

Recall that for complex  $z_n \rightarrow z$  we have the convergence  $(1 + \frac{z_n}{n})^n \rightarrow e^z$ . Let  $R_n(x) = e^{ix} - \sum_{k=0}^n (ix)^k/k!$  for  $x \in \mathbb{R}$ . Then (this is part of Exercise A.6)

$$|R_n(x)| \leq \min\left\{\frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!}\right\}. \quad (\text{A.3})$$

Let  $\phi$  be the characteristic function of the  $X_k$ . It then follows (this is another part of Exercise A.6) that

$$|\phi(\lambda) - 1 + \frac{1}{2}\sigma^2\lambda^2| \leq \mathbb{E} |R_2(\lambda X)| \leq \lambda^2 \mathbb{E} \left( \min\{|X|^2, \frac{1}{6}|\lambda||X|^3\} \right). \quad (\text{A.4})$$

Letting  $\lambda \rightarrow 0$ , 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  $\lambda \rightarrow 0$ .

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  $\phi$ , 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.  $\square$

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  $k = 1, \dots, k_n$ . The family  $\{X_{n,k} : n \in \mathbb{N}, k = 1, \dots, k_n\}$  is called a triangular array, a reasonable name if  $k_n$  is increasing in  $n$ .

**Theorem A.3** Consider a triangular array of random variables  $\{X_{n,k} : n \in \mathbb{N}, k = 1, \dots, k_n\}$  with  $\lim_{n \rightarrow \infty} k_n = \infty$ . Assume that for each  $n$  the random variables  $X_{n,k}$  for  $k = 1, \dots, k_n$  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  $\lim_{n \rightarrow \infty} \mathbb{E} S_n^2 = \sigma^2$  with  $\sigma^2 < \infty$ . Then we have weak convergence of the distributions of the  $S_n$  to the  $N(0, \sigma^2)$  distribution.

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.4 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  $\hat{f}(Y)$ .

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.4. Replacing  $X$  above by  $(X, Y)$  we can also define conditional expectations like  $\mathbb{E}[f(X, Y)|Y]$ . We have the following

**Proposition A.4** (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  $\mathbb{E}[f(X, Y)|Y] = \sum_{x \in \mathcal{X}} f(x, Y)\mathbb{P}(X = x)$ .*

The random variable  $Y$  induces a sub- $\sigma$ -algebra of  $\mathcal{F}$  on  $\Omega$ , call it  $\mathcal{G}$ . It is generated by the sets  $\{Y = y\}$ , that constitute a partition of  $\Omega$ . Consider again the function  $\hat{f}$  above. It defines another function on  $\Omega$ ,  $F$  say, according to  $F(\omega) = \hat{f}(y)$  on the set  $\{\omega : Y(\omega) = y\}$ . 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 that  $\mathbb{E}(F1_{Y=y}) = \mathbb{E}(f(X)1_{Y=y})$ . 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

$$\mathbb{E}(F1_G) = \mathbb{E}(f(X)1_G) \text{ for every set } G \in \mathcal{G}. \quad (\text{A.5})$$

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.5).

**Definition A.5** If  $X$  is a random variable with  $\mathbb{E}|X| < \infty$  and  $\mathcal{G}$  a sub- $\sigma$ -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  $\mathbb{E}X1_G = \mathbb{E}\hat{X}1_G$  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. 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  $\sigma$ -algebra that  $Y$  induces on  $\Omega$ . It is a theorem in probability theory that any version of the conditional expectation given  $Y$  can be represented by a (measurable) function of  $Y$ . The most important properties of conditional expectation that we use in this course are collected in the following proposition.

**Proposition A.6** *Let  $X$  be a random variable with  $\mathbb{E}|X| < \infty$ .*

(i) *If  $X$  is independent of  $\mathcal{G}$ , then  $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$ .*

(ii) *If  $\mathcal{H}$  is a sub- $\sigma$ -algebra of  $\mathcal{G}$ , then  $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ . (Repeated conditioning)*

(iii) *If  $Y$  is  $\mathcal{G}$ -measurable and  $\mathbb{E}|XY| < \infty$ , then  $\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y$ .*

(iv) *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  $X$ .*

(v) *The conditional expectation is a linear operator on the space of random variables with finite expectation.*

## A.5 Filtrations and Martingales

In probabilistic terms the loose term ‘information’ can be expressed by means of  $\sigma$ -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- $\sigma$ -algebras of  $\mathcal{F}$ . So we have a family  $\mathbb{F} = (\mathcal{F}_n)_{n \geq 0}$  where the  $\mathcal{F}_n$  are sub- $\sigma$ -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  $X = (X_1, X_2, \dots)$ , defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ . For each  $n$  we then put  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ , the smallest  $\sigma$ -algebra that makes  $X_1, \dots, X_n$  measurable functions on  $\Omega$ . In this case one often speaks of the filtration generated by the sequence  $X_1, X_2, \dots$  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  $S_1, \dots, S_n$  iff one knows all  $X_1, \dots, X_n$ .

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.7** 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  $n \geq 0$ .

The standard and easiest example of a martingale is the sum of independent random variables with zero expectation (this is Exercise A.12). 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  $H = (H_0, H_1, H_2, \dots)$  is said to be predictable w.r.t. a filtration  $\mathbb{F}$  if for all  $n$  the random variable  $H_n$  is  $\mathcal{F}_{n-1}$ -measurable. Here we take  $\mathcal{F}_{-1}$  the trivial  $\sigma$ -algebra.

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

$$V_n = \sum_{k=0}^n H_k \Delta M_k = \sum_{k=0}^n H_k (M_k - M_{k-1}), \quad (\text{A.6})$$

(with  $M_{-1} = 0$ ). Then also  $V$  is a martingale w.r.t.  $\mathbb{F}$  (this is Exercise A.13).

The relevant notions in continuous time are similarly defined. A filtration in continuous time is an increasing collection of sub- $\sigma$ -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- $\sigma$ -algebras of  $\mathcal{F}$  that satisfy  $\mathcal{F}_s \subset \mathcal{F}_t$  for all  $t \geq s \geq 0$ .

A martingale in continuous time is a process that satisfies the conditions of A.7 ‘with  $n$  replaced by  $t$ ’. If a 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  $\sigma$ -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.6 Modes of convergence

In this section we briefly treat various convergence concepts for random variables. Let us start by defining them.

**Definition A.8** 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  $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ .
- (ii)  $X_n$  converges to  $X$  almost surely (a.s.), if  $\mathbb{P}(\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1$ .
- (iii)  $X_n$  converges to  $X$  in  $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$  if  $\lim_{n \rightarrow \infty} \mathbb{E}|X_n - X| = 0$ .
- (iv)  $X_n$  converges to  $X$  in  $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$  if  $\lim_{n \rightarrow \infty} \mathbb{E}|X_n - X|^2 = 0$ .

Here are some relations between the different modes of convergence.

**Proposition A.9** 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  $\mathbb{E} X_n \rightarrow \mathbb{E} X$ .
- (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 subsequence  $(X_{n_k})$  such that  $X_{n_k}$  converges to  $X$  almost surely.

**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  $\mathbb{P}(\liminf_{k \rightarrow \infty} E_k^c) = 1$ . But for  $\omega \in \liminf_{k \rightarrow \infty} E_k^c$  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  $X(\omega)$ .  $\square$

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.10** Let  $X_n \rightarrow X$  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) (Dominated convergence) 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  $L^1(\Omega, \mathcal{F}, \mathbb{P})$ .

(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.

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.7 Uniqueness and the heat equation

In the main text we announced that under regularity conditions the heat equation has a unique solution. 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. Let for given  $T > 0$ , and real constants  $A$  and  $B$  let  $D \subset \mathbb{R}^2$  be the 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  $\partial D$  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.11** If the function  $u$  above satisfies the heat inequality

$$u_t(t, x) \leq \frac{1}{2}u_{xx}(t, x), (t, x) \in D, \quad (\text{A.7})$$

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.7). So let us assume that for the time being. Let  $\varepsilon \in (0, T)$  and consider  $D_\varepsilon = \{(t, x) \in$

$D : t < T - \varepsilon\}$ . Suppose that  $u$  has a maximum on  $\bar{D}_\varepsilon$  that is attained at an interior point of  $D_\varepsilon$ . 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 inequality (A.7) would be violated. We conclude that the maximum of  $u$  on  $\bar{D}_\varepsilon$  is attained on the boundary  $\partial D_\varepsilon$ . Suppose now that  $u$  is maximal at  $(T - \varepsilon, x_0) \in \partial D_\varepsilon \setminus \partial_0 D_\varepsilon$ . 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.7). Hence we conclude

$$\max_{(t,x) \in \bar{D}_\varepsilon} u(t, x) = \max_{(t,x) \in \partial_0 D_\varepsilon} u(t, x). \quad (\text{A.8})$$

Now we let  $\varepsilon \downarrow 0$  and use uniform continuity of  $u$  on  $\bar{D}$  to conclude that the left hand side of (A.8) 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.7). 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  $\max_{(t,x) \in \partial_0 D_\varepsilon} u'(t, x) \leq \max_{(t,x) \in \partial_0 D_\varepsilon} u(t, x)$ . The result then follows.  $\square$

The uniqueness theorem for solutions to the heat equation is a consequence of the following

**Theorem A.12** *Let  $u \in C^{1,2}((0, \infty) \times \mathbb{R})$  and continuous on  $[0, \infty) \times \mathbb{R}$ . Suppose that there are (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 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 notice 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  $D = D_h = \{(t, x) : t \in (0, h), |x| < h\}$ . We apply the parabolic maximum principle to  $v$  with respect to the domain  $D$  and we estimate the function  $v$  on  $\partial_0 D$ . If  $|x| = h$ , then

$$\begin{aligned} v(t, x) &\leq A \exp(Bh^2) - \delta w(t, h) \\ &\leq A \exp(Bh^2) - \delta (2T)^{-1/2} \exp(h^2/4T), \end{aligned} \quad (\text{A.9})$$

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 ( $h^2 > \log \frac{\delta}{A\sqrt{2T}} / (B - t/4)$ ), we can make (A.9) 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  $\bar{D}$ . Note that it now follows that  $v(t, x) \leq 0$  on  $[0, T] \times \mathbb{R}$  by letting  $h \rightarrow \infty$ . 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  $u'(t, x) = u(t + T, x)$ . 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.  $\square$

**Corollary A.13** *Under the conditions of Theorem A.12 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.12 yields the result.  $\square$

## A.8 Exercises

**A.1** Show by computation that the density of a non-degenerate multivariate normal random vector has integral equal to 1.

**A.2** 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  $\mu$  and  $\sigma^2$  if  $\log Z$  has a  $N(\mu, \sigma^2)$  distribution. What is the expectation of a log-normally distributed random variable with parameters  $\mu$  and  $\sigma^2$ ?

**A.3** Show by computation of the integral that Equation (A.2) holds for standard normal random variables.

**A.4** 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.5** Let for each  $n \in \mathbb{N}$  the random variable  $X_n$  be degenerate in  $1/n$ . Show that  $\mathbb{E} f(X_n) \rightarrow f(0)$  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.6** Define for  $x \in \mathbb{R}$  the functions  $R_n$  by  $R_n(x) = e^{ix} - \sum_{k=0}^n (ix)^k / k!$ . Show that  $|R_0(x)| \leq \min\{2, |x|\}$ ,  $R_n(x) = i \int_0^x R_{n-1}(t) dt$  and finally that the estimates (A.3) and (A.4) are valid.

**A.7** Suppose that  $X_1, X_2, \dots$  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  $\lim_{n \rightarrow \infty} \mathbb{P}(a_n < X_n \leq b_n) = \Phi(b) - \Phi(a)$ .

**A.8** Finish the proof of Theorem A.1.

**A.9** Prove Theorem A.3.

**A.10** Prove Proposition A.4.

**A.11** Let  $X$  and  $Y$  be finite valued random variables and  $f$  a function defined on the range of  $X$ . Let  $\hat{f}$  be a function defined on the range of  $Y$  with the property that  $\mathbb{E}(\hat{f}(y)1_{\{Y=y\}}) = \mathbb{E}(f(X)1_{\{Y=y\}})$  for all  $y$ . Show that  $\hat{f}(y) = \mathbb{E}[f(X)|Y = y]$  for all  $y$  with  $\mathbb{P}(Y = y) > 0$  and that  $\hat{Y} = \mathbb{E}[f(X)|Y]$ .

**A.12** Let  $X_1, X_2, \dots$  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.13** Show that the process  $V$  defined by Equation (A.6) is a martingale.

**A.14** Let  $X_n$  be multivariate normal random vectors, with expectations  $\mu_n$  and covariance matrices  $\Sigma_n$ . Suppose that the  $X_n$  converge in  $L^2$ -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.15** Let  $u : \mathbb{R}^2 \rightarrow \mathbb{R}$  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  $\alpha$ ) the points of  $\bar{D}$  where  $u$  attains a maximum. Make a sketch of the level sets of  $u$ .

**A.16** Here is an example of nonuniqueness of solutions to the heat equation. Let the initial condition be  $u(0, \cdot) = 0$ . One solution is  $u(t, x) \equiv 0$ . Let  $\phi$  be an infinitely many times differentiable function. Show that  $u(t, x) = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \frac{d^n}{dt^n} \phi(t)$  is a solution to the heat equation. Next we make the special choice  $\phi(t) = \exp(-1/t^2)1_{t>0}$ . 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.12?

**A.17** 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  $[0, T] \times \mathbb{R}$ .