Lecture Notes

Mathematical Foundations for Finance

© M. Schweizer and E. W. Farkas
Department of Mathematics
ETH Zürich

This version: December 15, 2020

Contents

1 Financial markets in finite discrete time

In this chapter, we introduce basic concepts in order to model trading in a frictionless financial market in finite discrete time. We recall the required notions from probability theory and stochastic processes and directly illustrate them by means of examples.

Standard concepts and results from (measure-theoretic) probability theory are assumed to be known; Chapter 8 contains a brief (and non-comprehensive) summary, and details can be found in Jacod/Protter [10] or Durrett [6].

1.1 Basic probabilistic concepts

Financial markets involve *uncertainty*, in particular about the future evolution of asset prices. We therefore start from a *probability space* $(\Omega, \mathcal{F}, P)$. Time evolves in discrete steps over a finite horizon; we label *trading dates* as $k = 0, 1, \ldots, T$ with $T \in \mathbb{N}$.

The flow of *information* over time is described by a *filtration* $\mathbb{F} = (\mathcal{F}_k)_{k=0,1,\ldots,T}$; this is a family of σ-fields $\mathcal{F}_k \subseteq \mathcal{F}$ which is *increasing* in the sense that $\mathcal{F}_k \subseteq \mathcal{F}_\ell$ for $k \leq \ell$. The interpretation is that $\mathcal{F}_k$ contains all events that are *observable* up to and including time k.

An ($\mathbb{R}^d$-valued) *stochastic process* in this discrete-time setting is simply a family $X = (X_k)_{k=0,1,\ldots,T}$ of ($\mathbb{R}^d$-valued) random variables which are all defined on the same probability space $(\Omega, \mathcal{F}, P)$. This can be used to describe the random evolution over time of d quantities, e.g. a bank account, asset prices, some liquidly traded options, or the holdings in a portfolio of assets. A stochastic process X is called *adapted* (to $\mathbb{F}$) if each X_k is $\mathcal{F}_k$-measurable, i.e. observable at time k; it is called *predictable* (with respect to $\mathbb{F}$) if each X_k is even $\mathcal{F}_{k-1}$-measurable, for $k = 1, \ldots, T$. (For the predictable processes X we use here, the value X_0 at time 0 is usually irrelevant.)

Example. If we think of a market where assets can be traded once each day (so that the time index k numbers days), then the price of a stock will usually be adapted because date k prices are known at date k. But if one wants to invest by selling or buying shares, one must make that decision before one knows where prices go in the next step; hence trading strategies must be predictable, unless one allows insiders or prophets. For a more

detailed discussion, see Section 1.2.

Example (multiplicative model). Suppose that we start with random variables $r_1, \dots, r_T$ and $Y_1, \dots, Y_T$. Take a constant $S_0^1 > 0$ and define

$$\widetilde{S}_k^0 := \prod_{j=1}^{k}(1 + r_j), \qquad \widetilde{S}_k^1 := S_0^1 \prod_{j=1}^{k} Y_j$$

for $k = 0, 1, \dots, T$. Note that we use here and throughout the *convention* that an empty product equals 1 and an empty sum equals 0. Suppose also that $r_k > -1$ and $Y_k > 0$ P-a.s. for $k = 1, \dots, T$. Then we have

$$\frac{\widetilde{S}_k^0}{\widetilde{S}_{k-1}^0} = 1 + r_k, \qquad \frac{\widetilde{S}_k^1}{\widetilde{S}_{k-1}^1} = Y_k,$$

or equivalently

$$\widetilde{S}_k^0 - \widetilde{S}_{k-1}^0 = \widetilde{S}_{k-1}^0 r_k, \qquad \widetilde{S}_k^1 - \widetilde{S}_{k-1}^1 = \widetilde{S}_{k-1}^1 (Y_k - 1),$$

with $\widetilde{S}_0^0 = 1$, $\widetilde{S}_0^1 = S_0^1$.

Interpretation. r_k describes the (simple) *interest rate* for the period $(k-1, k]$; so $\widetilde{S}^0$ models a *bank account* with that interest rate evolution, and $r_k > -1$ ensures that $\widetilde{S}^0 > 0$, in the sense that $\widetilde{S}_k^0 > 0$ P-a.s. for $k = 0, 1, \dots, T$. Similarly, $\widetilde{S}^1$ models a *stock*, say, and Y_k is the *growth factor* for the time period $(k-1, k]$. Of course, we could strengthen the analogy by writing $Y_k = 1 + R_k$; then $R_k > -1$ would describe the (simple) return on the stock for the period $(k-1, k]$.

How about the *filtration* in this example? For a general discussion, see Remark 1.1 below. The most usual choice for $\mathbb{F}$ is the filtration generated by Y, i.e.,

$$\mathcal{F}_k = \sigma(Y_1, \dots, Y_k) = \sigma(\widetilde{S}_0^1, \widetilde{S}_1^1, \dots, \widetilde{S}_k^1)$$

is the smallest σ-field that makes all stock prices up to time k observable. Then $\widetilde{S}^1$ is obviously adapted to $\mathbb{F}$. The bank account is naturally less risky than a stock, and in

particular the interest rate for the period $(k-1,k]$ is usually known at the beginning, i.e. at time $k-1$. So each r_k ought to be $\mathcal{F}_{k-1}$-measurable, i.e. the process $r = (r_k)_{k=1,\dots,T}$ should be predictable. Then $\widetilde{S}^0$ is also predictable (and vice versa). In particular, the interest rate r_k for the period $(k-1,k]$ then only depends on $Y_1,\dots,Y_{k-1}$ or equivalently on the stock prices $\widetilde{S}^1_0,\widetilde{S}^1_1,\dots,\widetilde{S}^1_{k-1}$, but not on other factors. This can be generalised.

Example (binomial model). Suppose all the r_k are constant with a value $r > -1$; this means that we have the same nonrandom interest rate over each period. Then the bank account evolves as $\widetilde{S}^0_k = (1+r)^k$ for $k = 0,1,\dots,T$.

Suppose also that $Y_1,\dots,Y_T$ are *independent* and only take two values, $1+u$ with probability p, and $1+d$ with probability $1-p$. In particular, this means that all the Y_k have the same distribution; they are *identically distributed* (with a particular two-point distribution). Usually, one also has $u > 0$ and $-1 < d < 0$ so that $1+u > 1$ and $0 < 1+d < 1$. Then the stock price at each step moves either up (by a factor $1+u$) or down (by a factor $1+d$), because

$$\frac{\widetilde{S}^1_k}{\widetilde{S}^1_{k-1}} = Y_k = \begin{cases} 1+u & \text{with probability } p \\ 1+d & \text{with probability } 1-p. \end{cases}$$

This is the so-called *Cox–Ross–Rubinstein (CRR) binomial model*.

Remark. If in the general multiplicative model, the r_k are all constant with the same value and $Y_1,\dots,Y_T$ are i.i.d., we have the *i.i.d. returns* model. If in addition the Y_k only take finitely many values (two or more), we get the *multinomial model*. ◇

Remark 1.1. (This remark is for mathematicians, but not only.) In the general multiplicative model, one could also start with the filtration

$$\mathcal{F}'_k := \sigma(Y_1,\dots,Y_k,r_1,\dots,r_k) = \sigma(\widetilde{S}^1_0,\widetilde{S}^1_1,\dots,\widetilde{S}^1_k,\widetilde{S}^0_0,\widetilde{S}^0_1,\dots,\widetilde{S}^0_k)$$

generated by both Y and r, or equivalently by both assets $\widetilde{S}^0$ and $\widetilde{S}^1$. In general, this filtration $\mathbb{F}'$ is bigger than $\mathbb{F}$, meaning that $\mathcal{F}'_k \supseteq \mathcal{F}_k$ for all k. But if one also assumes

that the process r (or, equivalently, the bank account $\widetilde{S}^0$) is predictable, one can show by induction that

$$\mathcal{F}'_k = \sigma(Y_1, \ldots, Y_k) = \mathcal{F}_k \qquad \text{for all } k.$$

This explains a posteriori why we have started above directly with $\mathbb{F}$ generated by Y. $\diamond$

1.2 Financial markets and trading

In this section, we present the basic model for a discrete-time financial market and explain how to describe dynamic trading in a mathematical way. This involves stochastic processes to describe asset prices and trading strategies, and gains or losses from trade are then naturally described by (discrete-time) stochastic integrals.

As Dieter Sondermann, the founder and first editor of the journal "Finance and Stochastics", once said: "The financial engineer always starts from a filtered probability space." In all the sequel in this chapter, we work on a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathbb{F} = (\mathcal{F}_k)_{k=0,1,\dots,T}$ for some $T \in \mathbb{N}$, without repeating this explicitly. We shall only be more specific when we want to exploit special properties of a particular model $(\Omega, \mathcal{F}, \mathbb{F}, P)$. We sometimes assume that $\mathcal{F}_0$ is *(P-)trivial*, i.e. $P[A] \in \{0, 1\}$ for all $A \in \mathcal{F}_0$; this equivalently means that any $\mathcal{F}_0$-measurable random variable is P-a.s. constant, and it represents a situation where we have no nontrivial information at time 0. For notational convenience, we sometimes also assume that $\mathcal{F} = \mathcal{F}_T$; this means that any event is observable by time T at the latest.

The basic *asset prices* in our financial market are specified by a strictly positive adapted process $\widetilde{S}^0 = (\widetilde{S}^0_k)_{k=0,1,\dots,T}$ and an $\mathbb{R}^d$-valued adapted process $\widetilde{S} = (\widetilde{S}_k)_{k=0,1,\dots,T}$. The interpretation is that $\widetilde{S}^0$ models a *reference asset* or *numeraire*; this explains why we assume that $\widetilde{S}^0_0 = 1$ and $\widetilde{S}^0$ is strictly positive, i.e. $\widetilde{S}^0_k > 0$ P-a.s. for all k. In many cases, we think of $\widetilde{S}^0$ as a *bank account* and then in addition also assume that $\widetilde{S}^0$ is predictable; see Section 1.1. In contrast, $\widetilde{S} = (\widetilde{S}^1, \dots, \widetilde{S}^d)$ describes the prices of d genuinely *risky assets* (often called *stocks*); so $\widetilde{S}^i_k$ is the price of asset i at time k, and because this becomes known at time k, but usually not earlier, each $\widetilde{S}^i$ and hence also the vector process $\widetilde{S}$ is adapted. For financial reasons, one might want $\widetilde{S}^i_k \geq 0$ P-a.s. for all i and k, but mathematically, this is not needed.

Prices (and values) are expressed in units of something, but it is economically not relevant what that is; all prices (and values) are *relative*. To simplify notations, we immediately switch to units of the reference asset $\widetilde{S}^0$; this is sometimes called "*discounting with* $\widetilde{S}^0$" or "*using* $\widetilde{S}^0$ *as numeraire*". Mathematically, it basically amounts to dividing at each time k every traded quantity by $\widetilde{S}^0_k$; so the discounted price of the reference asset is

simply $S^0_k := \widetilde{S}^0_k / \widetilde{S}^0_k = 1$ at all times, and the *discounted asset prices* $S = (S_k)_{k=0,1,\dots,T}$ are given by $S_k := \widetilde{S}_k / \widetilde{S}^0_k$. If $\widetilde{S}^0$ is viewed as a bank account, then in terms of interest rates, using discounted prices is equivalent to working with *zero interest*. We shall explain later how to re-incorporate interest rates; but our basic (discounted) model always has $S^0 \equiv 1$, and we usually call asset 0 the bank account.

Remark 2.1. It is important for this simplification by discounting that the reference asset 0 is also tradable. So while we have only d risky assets with discounted prices $S^1, \dots, S^d$, there are actually $d+1$ assets available for trading. This is almost always implicitly assumed in the literature, but not always stated explicitly.

2) Economically, it should not matter whether one works in original or in discounted prices (except that one has of course different units and different numbers). Mathematically, however, things are more subtle. In finite discrete time, there is indeed an equivalence between undiscounted and discounted formulations, as discussed in Delbaen/ Schachermayer [4, Section 2.5]. But in models with infinitely many trading dates (whether in infinite discrete time or in continuous time), one must be more careful because there are pitfalls. ◇

We assume that we have a *frictionless financial market*, which includes quite a lot of assumptions. There are *no transaction costs* so that assets can be bought or sold at the same price (at any given time); money (in the bank account) can be borrowed or lent at the same (zero) interest rate; assets are available in arbitrarily small or large quantities; there are *no constraints* on the numbers of assets one holds, and in particular, one may decide to own a negative number of shares (so-called *short selling*); and *investors* are *small* so that their trading activities have no effect on asset prices (which means that S is an exogenously and a priori given and fixed stochastic process). All this is of course unrealistic; but for explaining and understanding basic concepts, one has to start with the simplest case, and a frictionless financial market is in many cases at least a reasonable first approximation.

Definition. A *trading strategy* is an $\mathbb{R}^{d+1}$-valued stochastic process $\varphi = (\varphi^0, \vartheta)$, where $\varphi^0 = (\varphi^0_k)_{k=0,1,\dots,T}$ is real-valued and adapted, and $\vartheta = (\vartheta_k)_{k=0,1,\dots,T}$ with $\vartheta_0 = 0$ is $\mathbb{R}^d$-valued and predictable. The *(discounted) value process* of a strategy φ is the real-valued adapted process $V(\varphi) = (V_k(\varphi))_{k=0,1,\dots,T}$ given by

$$V_k(\varphi) := \varphi^0_k S^0_k + \vartheta^{\mathrm{tr}}_k S_k = \varphi^0_k + \sum_{i=1}^{d} \vartheta^i_k S^i_k \qquad \text{for } k = 0, 1, \dots, T. \tag{2.1}$$

Interpretation. A trading strategy describes a *dynamically evolving portfolio* in the $d+1$ basic assets available for trade. At time k, we have φ^0_k units of the bank account and ϑ^i_k units (shares) of asset (stock) i, so that straightforward financial book-keeping gives (2.1) as the time k value, in units of the bank account, of the time k portfolio holdings.

A little bit more precisely, $\varphi_k = (\varphi^0_k, \vartheta_k)$ is the portfolio with which we arrive at time k. Because stock prices change at time k from S_{k-1} to S_k and we arrive with holdings ϑ_k, we could easily make profits if we could choose ϑ_k at time k. To avoid this and exclude insiders and prophets, ϑ_k must therefore already be determined and chosen at time $k-1$; so ϑ_k is $\mathcal{F}_{k-1}$-measurable, hence ϑ is predictable, and ϑ_k are actually the holdings in risky assets on $[k-1, k)$. In the same way, φ^0_k are the bank account holdings on $[k-1, k)$; but as the bank account is riskless (at least locally for each time step, by predictability), one can allow φ^0 to be adapted without giving investors any extra advantages. So φ^0_k can be $\mathcal{F}_k$-measurable, which means that φ^0 is adapted..

With the above interpretation, we arrive at time k with the portfolio $\varphi_k = (\varphi^0_k, \vartheta_k)$ and change this at time k to a new portfolio $\varphi_{k+1} = (\varphi^0_{k+1}, \vartheta_{k+1})$ with which we then leave for date $k+1$. Hence $V_k(\varphi)$ in (2.1) is more precisely the *pre-trade value* of the strategy φ at time k. Note that we have not (yet) said anything about how investors get the money to implement and update their chosen strategies.

Finally, as there are no activities before time 0, we demand via $\vartheta_0 = 0$ that investors start out without any shares. All they can do at time 0 is decide on their initial investment $V_0(\varphi) = \varphi^0_0$ into the reference asset or bank account.

Remark. If the numeraire $\widetilde{S}^0$ is just strictly positive and adapted, but not necessarily

predictable, then also φ^0 must be predictable. We shall see later in Proposition 2.3 that this is automatically satisfied if the strategy φ is self-financing. $\diamond$

Of course, investors must do book-keeping about their expenses (and income). To work out the *costs* associated to a trading strategy $\varphi = (\varphi^0, \vartheta)$, we first observe that apart from time 0, transactions only occur at the dates k when φ_k is changed to φ_{k+1}. So the *incremental cost* for φ over the time interval $(k, k+1]$ occurs at time k when we change from φ_k to φ_{k+1} at the time-k prices S_k, and it is given by

$$\begin{aligned} \Delta C_{k+1}(\varphi) &:= C_{k+1}(\varphi) - C_k(\varphi) \\ &= (\varphi^0_{k+1} - \varphi^0_k) S^0_k + (\vartheta_{k+1} - \vartheta_k)^{\mathrm{tr}} S_k \\ &= \varphi^0_{k+1} - \varphi^0_k + \sum_{i=1}^{d} (\vartheta^i_{k+1} - \vartheta^i_k) S^i_k. \end{aligned} \tag{2.2}$$

Note that this is again in units of the bank account, hence discounted; and note also that (2.2) is just a *book-keeping identity* with no room for alternative or artificial definitions. Finally, the *initial cost* for φ at time 0 comes from putting φ^0_0 into the bank account; so

$$C_0(\varphi) = \varphi^0_0 = V_0(\varphi). \tag{2.3}$$

We also point out that it is to some extent arbitrary whether we associate the above cost increment $\Delta C_{k+1}(\varphi)$ to the time interval $(k, k+1]$ or to $[k, k+1)$. The choice we have made simplifies notations, but is not financially compelling.

Remark. φ^0, ϑ and S are all stochastic processes, and so φ^0_{k+1}, φ^0_k, ϑ_{k+1}, ϑ_k and S_k are all random variables, i.e., functions on Ω (to $\mathbb{R}$ or $\mathbb{R}^d$). In consequence, the equality in (2.2) is really an equality between functions, and so (2.2) means that we have this equality whenever we plug in an argument, i.e. for all ω. In particular, what looks like one simple equation is in fact an entire system of equations.

Of course, this comment applies not only to (2.2), but to all equalities or inequalities between random variables. In addition, it is usually enough if the set of all ω for which the relevant equality or inequality holds has probability 1; so e.g. (2.2) only needs to

hold P-a.s., and a similar comment applies again in general. We often do not write P-a.s. explicitly unless this becomes important for some reason. ◇

Notation. For any stochastic process $X = (X_k)_{k=0,1,\dots,T}$, we denote the *increment* of X from $k-1$ to k by

$$\Delta X_k := X_k - X_{k-1}.$$

Elementary rewriting of (2.2) automatically brings up a new process as follows. By adding and subtracting $\vartheta_{k+1}^{\mathrm{tr}} S_{k+1}$, we write

$$\begin{aligned} \Delta C_{k+1}(\varphi) &= \varphi_{k+1}^0 - \varphi_k^0 + (\vartheta_{k+1} - \vartheta_k)^{\mathrm{tr}} S_k \\ &= \varphi_{k+1}^0 + \vartheta_{k+1}^{\mathrm{tr}} S_{k+1} - \varphi_k^0 - \vartheta_k^{\mathrm{tr}} S_k - \vartheta_{k+1}^{\mathrm{tr}} (S_{k+1} - S_k) \\ &= V_{k+1}(\varphi) - V_k(\varphi) - \vartheta_{k+1}^{\mathrm{tr}} \Delta S_{k+1} \\ &= \Delta V_{k+1}(\varphi) - \vartheta_{k+1}^{\mathrm{tr}} \Delta S_{k+1}. \end{aligned} \tag{2.4}$$

But now we note that ϑ_{k+1} is the share portfolio we have when arriving at time $k+1$, and ΔS_{k+1} is the asset price change at time $k+1$; hence $\vartheta_{k+1}^{\mathrm{tr}} \Delta S_{k+1}$ is the *(discounted) incremental gain or loss* arising over $(k, k+1]$ from our trading strategy due to the price fluctuations of S. (There is no such gain or loss from the bank account because its price $S^0 \equiv 1$ does not change over time.) This justifies the following

Definition. Let $\varphi = (\varphi^0, \vartheta)$ be a trading strategy. The *(discounted) gains process* associated to φ or to ϑ is the real-valued adapted process $G(\vartheta) = (G_k(\vartheta))_{k=0,1,\dots,T}$ with

$$G_k(\vartheta) := \sum_{j=1}^{k} \vartheta_j^{\mathrm{tr}} \Delta S_j \qquad \text{for } k = 0, 1, \dots, T \tag{2.5}$$

(where $G_0(\vartheta) = 0$ by the usual convention that a sum over an empty set is 0). The *(discounted) cost process* of φ is defined by

$$C_k(\varphi) := V_k(\varphi) - G_k(\varphi) \qquad \text{for } k = 0, 1, \dots, T, \tag{2.6}$$

as justified by (2.3) and (2.4).

Remark 2.2. If we think of a continuous-time model where successive trading dates are infinitely close together, then the increment ΔS in (2.5) becomes a differential $\mathrm{d}S$ and the sum becomes an integral. This explains why the *stochastic integral* $G(\vartheta) = \int \vartheta \,\mathrm{d}S$ provides the natural description of gains from trade in a continuous-time financial market model. As a mathematical aside, we also note that we should think of this stochastic integral as "$G(\vartheta) = \int \sum_{i=1}^{d} \vartheta^i \,\mathrm{d}S^i$", not as "$\sum_{i=1}^{d} \int \vartheta^i \,\mathrm{d}S^i$". It turns out in stochastic calculus that this does make a difference. ◇

By construction, $C_k(\varphi) = C_0(\varphi) + \sum_{j=1}^{k} \Delta C_j(\varphi)$ describes the *cumulative (total) costs* for the strategy φ on the time interval $[0, k]$. If we do not want to worry about how to pay these costs, we ideally try to make sure they never occur, by imposing this as a condition on φ. This motivates the next definition.

Definition. A trading strategy $\varphi = (\varphi^0, \vartheta)$ is called *self-financing* if its cost process $C(\varphi)$ is constant over time (and hence equal to $C_0(\varphi) = V_0(\varphi) = \varphi_0^0$).

Due to (2.2), a strategy is self-financing if and only if it satisfies for each k

$$\varphi_{k+1}^0 - \varphi_k^0 + (\vartheta_{k+1} - \vartheta_k)^{\mathrm{tr}} S_k = \Delta C_{k+1}(\varphi) = 0 \qquad P\text{-a.s.} \tag{2.7}$$

As it should, from economic intuition, this means that changing the portfolio from φ_k to φ_{k+1} at time k can be done cost-neutrally, i.e. with zero gains or losses at that time. In particular, all losses from the portfolio due to stock price changes must be fully compensated by gains from the bank account holdings and vice versa, without infusing or draining extra funds. Due to (2.6), another *equivalent description* of a self-financing strategy $\varphi = (\varphi^0, \vartheta)$ is that it satisfies $C(\varphi) = C_0(\varphi)$ or

$$V(\varphi) = V_0(\varphi) + G(\vartheta) = \varphi_0^0 + G(\vartheta) \tag{2.8}$$

(in the sense that $V_k(\varphi) = V_0(\varphi) + G_k(\vartheta)$ P-a.s. for each k). This gives the following very useful result.

Proposition 2.3. *Any self-financing trading strategy $\varphi = (\varphi^0, \vartheta)$ is uniquely determined by its initial wealth V_0 and its "risky asset component" ϑ. In particular, any pair (V_0, ϑ), where V_0 is an $\mathcal{F}_0$-measurable random variable and ϑ is an $\mathbb{R}^d$-valued predictable process with $\vartheta_0 = 0$, specifies in a unique way a self-financing strategy. We sometimes write $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ for the resulting strategy φ.*

Moreover, if $\varphi = (\varphi^0, \vartheta)$ is self-financing, then $(\varphi^0_k)_{k=1,\dots,T}$ is automatically predictable.

The important feature of Proposition 2.3 is that it allows us to describe self-financing strategies in a very simple way. We just have to specify the initial wealth V_0 and the strategy ϑ we use for the risky assets; then the self-financing condition automatically tells us how the bank account component φ^0 must evolve. The proof simply makes that intuition precise, and so we give the short argument to get some practice.

Proof of Proposition 2.3. By (2.8) (or directly from the definitions of self-financing and of $C(\varphi)$ in (2.6), a strategy φ is self-financing if and only if for each k,

$$V_k(\varphi) = V_0(\varphi) + G_k(\vartheta) \qquad P\text{-a.s.}$$

Because $V_k(\varphi) = \varphi^0_k + \vartheta^{\mathrm{tr}}_k S_k$ by definition, we can rewrite the above equation for φ^0_k to get

$$\varphi^0_k = V_0(\varphi) + G_k(\vartheta) - \vartheta^{\mathrm{tr}}_k S_k,$$

which already shows that φ^0 is determined from V_0 and ϑ by the self-financing condition. To see that φ^0 is predictable, we note that

$$G_k(\vartheta) - G_{k-1}(\vartheta) = \Delta G_k(\vartheta) = \vartheta^{\mathrm{tr}}_k \Delta S_k = \vartheta^{\mathrm{tr}}_k (S_k - S_{k-1}).$$

Therefore

$$\begin{aligned} \varphi^0_k &= V_0(\varphi) + G_{k-1}(\vartheta) + \Delta G_k(\vartheta) - \vartheta^{\mathrm{tr}}_k S_k \\ &= V_0(\varphi) + G_{k-1}(\vartheta) - \vartheta^{\mathrm{tr}}_k S_{k-1} \end{aligned}$$

is directly seen to be $\mathcal{F}_{k-1}$-measurable, because $G(\vartheta)$ and S are adapted and ϑ is predictable. **q.e.d.**

Remarks. 1) The notion of a strategy being self-financing is a kind of *economic budget constraint*. Exactly like the cost process, this is formulated via basic *financial bookkeeping* requirements, and hence there cannot be any alternative (different) definitions that make sense financially. This is a clear example where basic modelling sense must override mathematical convenience. (In fact, there have been some attempts in continuous time to use a different concept of stochastic integral, the so-called Wick integral, to define the notion of a self-financing strategy. This has led to mathematical results which were easier to derive; but the approach has subsequently been demonstrated to be economically meaningless.)

2) We have expressed all prices and values in units of the bank account. However, as basic intuition suggests, this has no effect on whether or not a strategy is self-financing; indeed, because $\widetilde{S}^0_k > 0$, (2.7) is equivalent to

$$(\varphi^0_{k+1} - \varphi^0_k)\widetilde{S}^0_k + (\vartheta_{k+1} - \vartheta_k)^{\mathrm{tr}}\widetilde{S}_k = 0 \tag{2.9}$$

if we recall that $S = \widetilde{S}/\widetilde{S}^0$. But (2.9) is clearly the self-financing condition expressed in terms of the original units. The same argument shows that the notion of self-financing is *numeraire-invariant* in the sense that it does not depend on the units in which we do calculations. [→ *Exercise*] Note that it also does not matter here whether $\widetilde{S}^0$ is predictable or only adapted. ◇

Example (Stopping a process at a random time). Let $\tau : \Omega \to \{0, 1, \ldots, T\}$ be some mapping to be thought of as some *random time*; one specific example might be the first time that stock i's price exceeds that of stock j. We should like to use the "strategy" to "buy and then hold until time τ", because we believe for some reason that this might be a good idea. For ease of notation, we take $d = 1$ so that there is just one risky asset.

Formally, let us take $V_0 := S_0$ and

$$\vartheta_k(\omega) := I_{\{k \le \tau(\omega)\}} = \begin{cases} 1 & \text{for } k = 1, \ldots, \tau(\omega) \\ 0 & \text{for } k = \tau(\omega) + 1, \ldots, T, \end{cases}$$

which means exactly that we hold one unit of S up to and including time $\tau(\omega)$, but no further. The value process of the corresponding self-financing "strategy" $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ is

then by (2.8) and (2.5) given by

$$\begin{aligned} V_k(\varphi) &= V_0 + G_k(\vartheta) \\ &= S_0 + \sum_{j=1}^{k} \vartheta_j \Delta S_j \\ &= S_0 + \sum_{j=1}^{k} I_{\{j \le \tau\}} (S_j - S_{j-1}) \\ &= S_0 + \begin{cases} S_k - S_0 & \text{if } \tau > k \\ S_\tau - S_0 & \text{if } \tau \le k \end{cases} \\ &= S_{k \wedge \tau} = \begin{cases} S_k & \text{if } k < \tau \\ S_\tau & \text{if } k \ge \tau, \end{cases} \end{aligned}$$

where we use the standard notation $a \wedge b := \min(a, b)$.

The "stochastic process" $S^\tau = (S^\tau_k)_{k=0,1,\dots,T}$ defined by

$$S^\tau_k(\omega) := S_{k \wedge \tau}(\omega) := S_{k \wedge \tau(\omega)}(\omega)$$

is called the *process S stopped at τ*, because it clearly behaves like S up to time τ and remains constant after time τ. Of course, for every $\omega \in \Omega$, this operation and notation per se make sense for any stochastic process and any "random time" τ as above.

However, a closer look shows that one must be a little more careful. For one thing, S^τ could fail to be a stochastic process because $S^\tau_k = S_{k \wedge \tau}$ could fail to be a random variable, i.e. could fail to be measurable. But (in discrete time like here) this is not a problem if we assume that τ is *measurable*, which is mild and reasonable enough.

While the measurability question is mainly technical, there is a second and financially much more relevant issue. For φ to be a strategy, we need ϑ to be predictable, and this translates into the equivalent requirement that τ should be a so-called *stopping time*, meaning that $\tau : \Omega \to \{0, 1, \dots, T\}$ satisfies

$$\{\tau \le j\} \in \mathcal{F}_j \qquad \text{for all } j. \tag{2.10}$$

To see this, note that $\vartheta_k = I_{\{k \le \tau\}}$ is $\mathcal{F}_{k-1}$-measurable if and only if $\{\tau \ge k\} \in \mathcal{F}_{k-1}$, and to have this for all k is equivalent to (2.10) by passing to complements. By definition,

(2.10) means that τ is a stopping time (with respect to $\mathbb{F}$, to be accurate). Intuitively, (2.10) says that at each time j, we can observe from the then available information $\mathcal{F}_j$ whether or not τ is already past, i.e., whether the event corresponding to τ has already occurred. Typical *examples* are the first (or, by induction, n-th) time that an adapted process does something that only involves looking at the past, e.g.

$$\tau(\omega) := \inf\{k : S^i_k(\omega) > S^j_k(\omega)\} \wedge T$$

(the first time that stock i's price exceeds that of stock j) or

$$\tau'(\omega) := \inf\left\{k : S^1_k(\omega) \geq 10 \max_{j=0,1,\ldots,k-1} S^1_j(\omega)\right\} \wedge T$$

(the first time that stock 1's price goes above ten times its past maximum value). On the other hand, times looking at the future like

$$\tau''(\omega) := \sup\{k : S^\ell_k(\omega) > 5\} \vee 0$$

(the *last* time that stock ℓ's price exceeds 5) are typically *not* stopping times; so they cannot be used for constructing such buy-and-hold strategies. This makes intuitive sense.

Example (A doubling strategy). Suppose we have a model where the stock price can in each step only go up or down. A well-known idea for a strategy to force winnings is then to bet on a rise and keep on betting, doubling the stakes at each date, until the rise occurs.

More formally, consider the *binomial model* with parameters $u > 0$, $-1 < d < 0$ and $r = 0$; so the stock price S_k is either $(1+u)S_{k-1}$ or $(1+d)S_{k-1}$. To simplify computations, suppose $u = -d$ so that the growth factors $Y_k = S_k/S_{k-1}$ are symmetric around 1. Note that as seen earlier,

$$\Delta S_k = S_k - S_{k-1} = S_{k-1}(Y_k - 1). \tag{2.11}$$

Now denote by

$$\tau := \inf\{k : Y_k = 1 + u\} \wedge T \tag{2.12}$$

the (random) time of the first stock price rise and define

$$\vartheta_k := \frac{1}{S_{k-1}} 2^{k-1} I_{\{k \le \tau\}}. \tag{2.13}$$

Then τ is a stopping time, because

$$\{\tau \le j\} = \{\max(Y_1, \dots, Y_j) \ge 1 + u\} \in \mathcal{F}_j$$

for each j, and so ϑ is predictable because each ϑ_k is $\mathcal{F}_{k-1}$-measurable. Note that this uses $\{k \le \tau\} = \{\tau < k\}^c = \{\tau \le k-1\}^c$. Moreover,

$$\vartheta_{k+1} S_k = 2^k I_{\{\tau \ge k+1\}} = 2 \times 2^{k-1} (I_{\{\tau \ge k\}} - I_{\{\tau = k\}}) = 2\vartheta_k S_{k-1} - 2^k I_{\{\tau = k\}}$$

shows that while we are not successful, the value of our stock holdings (not the amount of shares of the strategy itself) indeed doubles from one step to the next.

For $V_0 := 0$, we now take the self-financing strategy φ corresponding to (V_0, ϑ). Its value process is by (2.8) and (2.5) given by

$$V_k(\varphi) = G_k(\vartheta) = \sum_{j=1}^{k} \vartheta_j \Delta S_j = \sum_{j=1}^{k} 2^{j-1} I_{\{j \le \tau\}} (Y_j - 1),$$

using (2.11) and (2.13). By the definition (2.12) of τ, we have $Y_j = 1 + d$ for $j < \tau$ and $Y_j = 1 + u$ for $j = \tau$; so

$$\begin{aligned} V_k(\varphi) &= I_{\{\tau > k\}} \sum_{j=1}^{k} 2^{j-1} d + I_{\{\tau \le k\}} \Big(\sum_{j=1}^{\tau-1} 2^{j-1} d + 2^{\tau-1} u \Big) \\ &= (2^k - 1) d \, I_{\{\tau > k\}} + \big((2^{\tau-1} - 1) d + 2^{\tau-1} u \big) I_{\{\tau \le k\}}. \end{aligned}$$

Because $u = -d$ and $d < 0$, we can write this as

$$V_k(\varphi) = |d| I_{\{\tau \le k\}} - |d| (2^k - 1) I_{\{\tau > k\}},$$

which says that we obtain a value, and hence net gain, of $|d|$ in all the (usually many) cases that S goes up at least once up to time k, and make a (big) loss of $|d|(2^k - 1)$ in the (hopefully unlikely) event that S always goes down up to time k.

One problem with the doubling strategy in the above example is that while it does produce a gain in many cases, its value process goes very far below 0 in those cases where "things go badly". In continuous time or over an infinite time horizon, one obtains quite pathological effects if one does not forbid such strategies in some way. The next definition aims at that.

Definition. For $a \geq 0$, a trading strategy φ is called *a-admissible* if its value process $V(\varphi)$ is uniformly bounded from below by $-a$, i.e. $V(\varphi) \geq -a$ in the sense that $V_k(\varphi) \geq -a$ P-a.s. for all k. A trading strategy is *admissible* if it is a-admissible for some $a \geq 0$.

Interpretation. An admissible strategy has some credit line which imposes a lower bound on the associated value process; so one may make debts, but only within clearly defined limits. Note that while every admissible strategy has some credit line, the level of that can be different for different strategies.

Remarks. 1) If Ω (or more generally $\mathcal{F}$) is finite, any random variable can only take finitely many values; for any model with finite discrete time, every trading strategy is then admissible. But if $\mathcal{F}$ (or the time horizon) is infinite or time is continuous, imposing admissibility is usually a genuine and important restriction. We return to this point later.

2) Note that all our prices and values are discounted and hence expressed in units of the reference asset 0. Imposing a constant lower bound on a value process like admissibility does is therefore obviously not invariant if we change to a different reference asset for discounting. This is the root of the pitfalls mentioned earlier in Remark 2.1. $\diamond$

1.3 Some important martingale results

Martingales are ubiquitous in mathematical finance, as we shall see very soon. This section collects a number of important facts and results we shall use later on.

Let $(\Omega, \mathcal{F}, Q)$ be a probability space with a filtration $\mathbb{F} = (\mathcal{F}_k)_{k=0,1,\dots,T}$. A (real-valued) stochastic process $X = (X_k)_{k=0,1,\dots,T}$ is called a *martingale* (with respect to Q and $\mathbb{F}$) if it is adapted to $\mathbb{F}$, is Q-integrable in the sense that $X_k \in \mathcal{L}^1(Q)$ for each k, and satisfies the *martingale property*

$$E_Q[X_\ell \,|\, \mathcal{F}_k] = X_k \qquad Q\text{-a.s. for } k \leq \ell. \tag{3.1}$$

Intuitively, this means that the best prediction for the later value X_ℓ given the earlier information $\mathcal{F}_k$ is just the current value X_k; so the changes in a martingale cannot be predicted. If we have "$\leq$" in (3.1) (a tendency to go down), X is called a *supermartingale*; if we have "$\geq$", then X is a *submartingale*. An $\mathbb{R}^d$-valued process X is a martingale if each coordinate X^i is a martingale.

It is important to note that the property of being a martingale depends on the probability we use to look at a process. The same process can very well be a martingale under some Q, but not a martingale under another Q' or P.

Example. In the binomial model on $(\Omega, \mathcal{F}, \mathbb{F}, P)$ with parameters r, u, d, the discounted stock price $\widetilde{S}^1/\widetilde{S}^0$ is a P-martingale if and only if $r = pu + (1-p)d$.

Indeed, $\widetilde{S}^1/\widetilde{S}^0$ is obviously adapted and takes only finitely many values; so it is bounded and hence integrable. Moreover, by induction, one easily sees that it is enough to check (the one-step martingale property) that

$$E_P\left[\frac{\widetilde{S}^1_{k+1}}{\widetilde{S}^0_{k+1}} \,\middle|\, \mathcal{F}_k\right] = \frac{\widetilde{S}^1_k}{\widetilde{S}^0_k} \qquad \text{for each } k$$

or equivalently that

$$1 = E_P\left[\frac{\widetilde{S}^1_{k+1}}{\widetilde{S}^0_{k+1}} \Big/ \frac{\widetilde{S}^1_k}{\widetilde{S}^0_k} \,\middle|\, \mathcal{F}_k\right] = E_P\left[\frac{Y_{k+1}}{1+r} \,\middle|\, \mathcal{F}_k\right].$$

But Y_{k+1} is independent of $\mathcal{F}_k$ and takes the values $1+u, 1+d$ with probabilities $p, 1-p$. Therefore

$$\begin{aligned} E_P\left[\frac{Y_{k+1}}{1+r}\,\middle|\,\mathcal{F}_k\right] &= \frac{1}{1+r}E_P[Y_{k+1}] \\ &= \frac{1}{1+r}\big(p(1+u)+(1-p)(1+d)\big) \\ &= \frac{1+pu+(1-p)d}{1+r}. \end{aligned}$$

This equals 1 if and only if $r = pu + (1-p)d$, which proves the assertion.

For mathematical reasons and arguments, the following generalisation of martingales is extremely useful.

Definition. An adapted process $X = (X_k)_{k=0,1,\ldots,T}$ null at 0 (i.e. with $X_0 = 0$) is called a *local martingale* (with respect to Q and $\mathbb{F}$) if there exists a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ increasing to T such that for each $n \in \mathbb{N}$, the stopped process $X^{\tau_n} = (X_{k\wedge\tau_n})_{k=0,1,\ldots,T}$ is a $(Q,\mathbb{F})$-martingale. We then call $(\tau_n)_{n\in\mathbb{N}}$ a *localising sequence*.

Remarks. **1)** Especially in continuous time, local martingales can be substantially different from (true) martingales; the concept is rather subtle.

2) In parts of the recent finance literature, local martingales have come up in studies of price bubbles. ◇

The next result gives a whole class of examples of local martingales.

Theorem 3.1. *Suppose $X = (X_k)_{k=0,1,\ldots,T}$ is an $\mathbb{R}^d$-valued martingale or local martingale null at 0. For any $\mathbb{R}^d$-valued predictable process ϑ, the* stochastic integral process *$\vartheta \bullet X$ defined by*

$$\vartheta \bullet X_k := \sum_{j=1}^{k} \vartheta_j^{\mathrm{tr}} \Delta X_j \qquad \text{for } k = 0, 1, \ldots, T$$

is then a (real-valued) local martingale null at 0. *If* X *is a martingale and* ϑ *is bounded, then* $\vartheta \bullet X$ *is even a martingale.*

Note that if we think of $X = S$ as discounted asset prices, then $\vartheta \bullet S = G(\vartheta)$ is the discounted gains process of the self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$.

Proof of Theorem 3.1. This result is important enough to deserve at least a partial proof. So suppose X is a Q-martingale and ϑ is bounded. Then $\vartheta \bullet X$ is also Q-integrable, it is always adapted, and

$$\begin{aligned} E_Q[\vartheta \bullet X_{k+1} - \vartheta \bullet X_k \,|\, \mathcal{F}_k] &= E_Q[\vartheta^{\mathrm{tr}}_{k+1} \Delta X_{k+1} \,|\, \mathcal{F}_k] \\ &= \sum_{i=1}^{d} E_Q[\vartheta^i_{k+1} \Delta X^i_{k+1} \,|\, \mathcal{F}_k]. \end{aligned}$$

But ϑ^i_{k+1} is bounded and $\mathcal{F}_k$-measurable because ϑ is predictable, and ΔX^i_{k+1} is Q-integrable because X is a Q-martingale; so

$$E_Q[\vartheta^i_{k+1} \Delta X^i_{k+1} \,|\, \mathcal{F}_k] = \vartheta^i_{k+1} E_Q[\Delta X^i_{k+1} \,|\, \mathcal{F}_k] = 0$$

again because X^i is a Q-martingale. So $\vartheta \bullet X$ also has the martingale property.

For the mathematicians: Because ϑ is predictable,

$$\sigma_n := \inf\{k : |\vartheta_{k+1}| > n\} \wedge T$$

is a stopping time, and $|\vartheta_k| \le n$ for $k \le \sigma_n$ by definition. So if $(\tau_n)_{n \in \mathbb{N}}$ is a localising sequence for X, one can easily check with the above argument that $\tau'_n := \tau_n \wedge \sigma_n$ yields a localising sequence for $\vartheta \bullet X$. This gives the general result. **q.e.d.**

We have seen earlier that if τ is any stopping time, then $\vartheta_k := I_{\{k \le \tau\}}$ is predictable, and of course bounded. So if we note that $\vartheta \bullet X = X^\tau - X_0$, an immediate consequence of Theorem 3.1 is

Corollary 3.2. *For any martingale* X *and any stopping time* τ, *the stopped process* X^τ *is again a martingale. In particular,* $E_Q[X_{k \wedge \tau}] = E_Q[X_0]$ *for all* k.

Interpretation. A martingale describes a *fair game* in the sense that one cannot predict where it goes next. Corollary 3.2 says that one cannot change this fundamental character by cleverly stopping the game — and Theorem 3.1 says that as long as one can only use information from the past, not even complicated clever betting (in the form of trading strategies) will help.

Remark. Corollary 3.2 still holds if we replace "martingale" by either "supermartingale" or "submartingale". However, such a generalisation is not true in general for Theorem 3.1. [→ *Exercise*] ◇

In general, the stochastic integral with respect to a local martingale is only a local martingale — and in continuous time, it may fail to be even that in the most general case. But there is one situation where things are very nice in discrete time, and this is tailor-made for applications in mathematical finance, as one can see by looking at the definition of self-financing and admissible strategies.

Theorem 3.3. *Suppose that X is an $\mathbb{R}^d$-valued local Q-martingale null at 0 and ϑ is an $\mathbb{R}^d$-valued predictable process. If the stochastic integral process $\vartheta \bullet X$ is uniformly bounded below (i.e. $\vartheta \bullet X_k \geq -b$ Q-a.s. for all k, with a constant $b \geq 0$), then $\vartheta \bullet X$ is a Q-martingale.*

Proof. See Föllmer/Schied [9, Theorem 5.15]. A bit more generally, this relies on the result that in discrete (possibly infinite) time, a local martingale that is uniformly bounded below is a true martingale. More precisely: If $L = (L_k)_{k \in \mathbb{N}_0}$ is a local Q-martingale with $E_Q[|L_0|] < \infty$ and $T \in \mathbb{N}$ is such that $E_Q[L_T^-] < \infty$, then the stopped process $L^T = (L_k)_{k=0,1,\ldots,T}$ is a Q-martingale. **q.e.d.**

We shall see later that Theorem 3.3 is extremely useful.

Remark. We have formulated everything here for the setting $k = 0, 1, \ldots, T$ of finite

discrete time. The same definitions and results also apply for the setting $k \in \mathbb{N}_0$ of infinite discrete time; the only required change is that one must replace T by ∞ in an appropriate manner. $\diamond$

1.4 An example: The multinomial model

In this section, we take a closer look at the multinomial model already introduced briefly in Section 1.1. Recall that this is the multiplicative model with i.i.d. returns given by

$$\frac{\widetilde{S}^0_k}{\widetilde{S}^0_{k-1}} = 1 + r > 0 \qquad \text{for all } k,$$

$$\frac{\widetilde{S}^1_k}{\widetilde{S}^1_{k-1}} = Y_k \qquad \text{for all } k,$$

where $\widetilde{S}^0_0 = 1$, $\widetilde{S}^1_0 = S^1_0 > 0$ is a constant, and $Y_1, \ldots, Y_T$ are i.i.d. and take the finitely many values $1 + y_1, \ldots, 1 + y_m$ with respective probabilities $p_1, \ldots, p_m$. To avoid degeneracies and fix the notation, we assume that all the probabilities p_j are > 0 and that $y_m > y_{m-1} > \cdots > y_1 > -1$. This also ensures that $\widetilde{S}^1$ remains strictly positive.

The *interpretation* for this model is very simple. At each step, the bank account changes by a factor of $1+r$, while the stock changes by a random factor that can only take the m different values $1+y_j$, $j = 1, \ldots, m$. The choice of these factors happens randomly, with the same mechanism (identically distributed) at each date, and independently across dates. Intuition suggests that for a reasonable model, the sure factor $1 + r$ should lie between the minimal and maximal values $1 + y_1$ and $1 + y_m$ of the (uncertain) random factor; we come back to this issue in the next chapter when we discuss absence of arbitrage.

The simplest and in fact *canonical model* for this setup is a *path space*. Let

$$\begin{aligned}\Omega &= \{1, \ldots, m\}^T \\ &= \big\{\omega = (x_1, \ldots, x_T) : x_k \in \{1, \ldots, m\} \text{ for } k = 1, \ldots, T\big\}\end{aligned}$$

be the set of all strings of length T formed by elements of $\{1, \ldots, m\}$. Take $\mathcal{F} = 2^\Omega$, the family of all subsets of Ω, and define P by setting

$$P[\{\omega\}] = p_{x_1} p_{x_2} \cdots p_{x_T} = \prod_{k=1}^{T} p_{x_k}. \tag{4.1}$$

Finally, define $Y_1, \ldots, Y_T$ by

$$Y_k(\omega) := 1 + y_{x_k} \tag{4.2}$$

so that $Y_k(\omega) = 1 + y_j$ if and only if $x_k = j$. This mathematically formalises the idea that at each step k, we choose the value $1 + y_j$ for Y_k with probability p_j, and we do this independently over k because P is obtained by multiplication. A nice way to graphically illustrate the construction of this canonical model $(\Omega, \mathcal{F}, P)$ is to draw a (non-recombining) *tree* of length T with m *branches* going out from each *node*. We then place the p_j as *one-step transition probabilities* into each branching, and the probability of each single trajectory ω is obtained by multiplying the one-step transition probabilities along the way. *[A figure to illustrate this is very helpful.]*

As usual, we take as *filtration* the one generated by $\widetilde{S}^1$ (or, equivalently, by Y) so that

$$\mathcal{F}_k = \sigma(Y_1, \ldots, Y_k) \qquad \text{for } k = 0, 1, \ldots, T.$$

Intuitively, this means that up to time k, we can observe the values of $Y_1, \ldots, Y_k$ and hence the first k "bits" of the trajectory or string ω. *Formally*, this translates as follows.

Recall that for a general probability space $(\Omega, \mathcal{F}, P)$, a set B is an atom of a σ-field $\mathcal{G} \subseteq \mathcal{F}$ if $B \in \mathcal{G}$, $P[B] > 0$ and any $C \in \mathcal{G}$ with $C \subseteq B$ has either $P[C] = 0$ or $P[C] = P[B]$. In that sense, atoms of a σ-field $\mathcal{G}$ are minimal elements of $\mathcal{G}$, where minimal is measured with the help of P.

In the above path-space setting, the only set of probability zero is the empty set, and so $P[C] = 0$ and $P[C] = P[B|$ translate into $C = \emptyset$ and $C = B$, respectively. A set $A \subseteq \Omega$ is therefore an *atom of* $\mathcal{F}_k$ if and only if there exists a string $(\bar{x}_1, \ldots, \bar{x}_k)$ of length k with elements $\bar{x}_i \in \{1, \ldots, m\}$ such that A consists of all those $\omega \in \Omega$ that start with the substring $(\bar{x}_1, \ldots, \bar{x}_k)$, i.e.

$$A = A_{\bar{x}_1,\ldots,\bar{x}_k} := \big\{\omega = (x_1, \ldots, x_T) \in \{1, \ldots, m\}^T : x_i = \bar{x}_i \text{ for } i = 1, \ldots, k\big\}.$$

This has the following consequences for our path-space model:

- Each $\mathcal{F}_k$ is parametrised by substrings of length k and therefore contains precisely m^k atoms.

- When going from time k to time $k+1$, each atom $A = A_{\bar{x}_1,\ldots,\bar{x}_k}$ from $\mathcal{F}_k$ splits into precisely m subsets $A_1 = A_{\bar{x}_1,\ldots,\bar{x}_k,1}, \ldots, A_m = A_{\bar{x}_1,\ldots,\bar{x}_k,m}$ that are atoms of $\mathcal{F}_{k+1}$. So

we can see very precisely and graphically how information about the past, i.e. the initial part of trajectories ω, is growing and refining over time.

It is clear from the above description that for any k, the atoms of $\mathcal{F}_k$ are pairwise disjoint and their union is Ω; in other words, the atoms of $\mathcal{F}_k$ form a partition of Ω so that we can write

$$\Omega = \bigcup_{(\bar{x}_1,\ldots,\bar{x}_k)\in\{1,\ldots,m\}^k} A_{\bar{x}_1,\ldots,\bar{x}_k} \qquad \text{with the } A_{\bar{x}_1,\ldots,\bar{x}_k} \text{ pairwise disjoint.}$$

Finally, each set B in $\mathcal{F}_k$ is a union of atoms of $\mathcal{F}_k$; so the family $\mathcal{F}_k$ of events observable up to time k consists of 2^{m^k} sets (because for each of the m^k atoms, we can either include it or not when forming B).

Remark. For many (but not all) purposes in the multinomial model, it is enough if one looks at time k only at the current value $\widetilde{S}^1_k$ of the stock. In graphical terms, this means that one makes the underlying tree *recombining* by collapsing at each time k into one (big) node all those nodes where $\widetilde{S}^1_k$ has the same value. In terms of σ-fields, this amounts to looking at time k only at $\mathcal{G}_k = \sigma(\widetilde{S}^1_k)$. It is clear that $\mathcal{G}_k$ (as a collection of subsets of Ω, i.e. $\mathcal{G}_k \subseteq 2^\Omega$) is substantially smaller than $\mathcal{F}_k$ and also that the recombining tree is much less complicated. However, note that the family $(\mathcal{G}_k)_{k=0,1,\ldots,T}$ is in general not a filtration; we do not have $\mathcal{G}_k \subseteq \mathcal{G}_\ell$ for $k \le \ell$. ◇

With the help of the atoms introduced above, we can also give a very precise and intuitive description of all *probability measures* Q on $\mathcal{F}_T$. First of all, we identify each atom in $\mathcal{F}_k$ with a node at time k of the non-recombining tree, namely that node which is reached via the substring $(\bar{x}_1,\ldots,\bar{x}_k)$ that parametrises the atom. For any atom $A = A_{\bar{x}_1,\ldots,\bar{x}_k}$ of $\mathcal{F}_k$, we then look at its m successor atoms $A_1 = A_{\bar{x}_1,\ldots,\bar{x}_k,1}, \ldots, A_m = A_{\bar{x}_1,\ldots,\bar{x}_k,m}$ of $\mathcal{F}_{k+1}$, and we *define* the *one-step transition probabilities* for Q at the node corresponding to A by the conditional probabilities (note that $A_j \cap A = A_j$ as $A_j \subseteq A$)

$$Q[A_j \mid A] = \frac{Q[A_j]}{Q[A]} \qquad \text{for } j = 1,\ldots,m. \tag{4.3}$$

Because A is the disjoint union of $A_1, \ldots, A_m$, we have $0 \leq Q[A_j \,|\, A] \leq 1$ for $j = 1, \ldots, m$ and $\sum_{j=1}^{m} Q[A_j \,|\, A] = 1$. (If $Q[A]$ is zero, then so are all the $Q[A_j]$ because $A_j \subseteq A$, and we can for instance define the ratios to be $\frac{1}{m}$, to make sure they are ≥ 0 and sum to 1.) By attaching all these one-step transition probabilities to each branch from each node, we then have by construction a decomposition or factorisation of Q in such a way that for every trajectory $\omega \in \Omega$, its probability $Q[\{\omega\}]$ is the product of the successive one-step transition probabilities along ω. This follows in an elementary way from the definition of conditional probabilities, $Q[C \cap D] = Q[C]\, Q[D \,|\, C]$, and by iteration. In more detail, we can write, for $\bar{\omega} = (\bar{x}_1, \ldots, \bar{x}_T)$,

$$\begin{aligned} Q[\{\bar{\omega}\}] &= Q[A_{\bar{x}_1,\ldots,\bar{x}_T}] \\ &= Q[A_{\bar{x}_1,\ldots,\bar{x}_T} \,|\, A_{\bar{x}_1,\ldots,\bar{x}_{T-1}}]\, Q[A_{\bar{x}_1,\ldots,\bar{x}_{T-1}}] \\ &= q_{\bar{x}_T}(\bar{x}_1, \ldots, \bar{x}_{T-1}) Q[A_{\bar{x}_1,\ldots,\bar{x}_{T-1}}] \end{aligned}$$

and iterate from here to obtain

$$Q[\{\bar{\omega}\}] = q_{\bar{x}_1} \prod_{j=1}^{T-1} q_{\bar{x}_{j+1}}(\bar{x}_1, \ldots, \bar{x}_j).$$

In the above procedure, we have factorised a given probability measure Q on $(\Omega, \mathcal{F})$ into its one-step transition probabilities. However, this idea also works the other way round. If we take for each node m numbers in $[0, 1]$ that sum to 1 and attach them to the branches from that node as "one-step transition probabilities", then defining $Q[\{\omega\}]$ for each $\omega \in \Omega$ to be as in (4.1) the product of the numbers along ω defines a probability measure Q on $\mathcal{F}_T$ whose one-step transition probabilities, defined as above in (4.3) via atoms, coincide with the a priori chosen numbers at each node. Indeed, just using (4.1) gives in (4.3) that $Q[A_j \,|\, A] = Q[A_{\bar{x}_1,\ldots,\bar{x}_k,j} \,|\, A_{\bar{x}_1,\ldots,\bar{x}_k}] = q_j(\bar{x}_1, \ldots, \bar{x}_k)$. Hence we can describe Q equivalently either via its global weights $Q[\{\omega\}]$ or via its local transition behaviour. The latter description is particularly useful when computing conditional expectations under Q, as we shall see later in Sections 2.1, 2.3 or 3.3.

For a general Q, one can have different one-step transition probabilities at every node in the tree. The (coordinate) variables $Y_1, \ldots, Y_T$ from (4.2) are *independent* under Q if and only if for each k, the one-step transition probabilities are the same for each node at

time k (but they can still differ across dates k). Finally, $Y_1, \ldots, Y_T$ are *i.i.d.* under Q if and only if at each node throughout the tree, the one-step transition probabilities are the same. Probability measures with this particular structure can therefore be described by $m - 1$ parameters; recall that the m one-step transition probabilities at any given node must sum to 1, which eliminates one degree of freedom.

Remark. We have discussed the path space formulation for the multinomial model where each node in the tree has the same number of successor nodes and in that sense is homogeneous in time. But of course, the same considerations can be done for any model where the final σ-algebra $\mathcal{F}_T$ is finite. The only difference is that the corresponding event tree is no longer nicely symmetric and homogeneous, which makes the notation (but not the basic considerations) more complicated. $\diamond$

2 Arbitrage and martingale measures

Our goal in this chapter is to formalise the idea that a reasonable financial market model should not allow the construction of riskless yet profitable investment strategies, and to characterise this by an equivalent mathematical property. **Throughout the chapter**, we consider a discounted financial market in finite discrete time on $(\Omega, \mathcal{F}, I\!F, P)$ with $I\!F = (\mathcal{F}_k)_{k=0,1,\dots,T}$, where discounted asset prices are given by the processes $S^0 \equiv 1$ and $S = (S_k)_{k=0,1,\dots,T}$, the latter taking values in $I\!R^d$.

2.1 Arbitrage

Recall from Proposition 1.2.3 that any pair (V_0, ϑ) consisting of $V_0 \in L^0(\mathcal{F}_0)$ and an $I\!R^d$-valued $I\!F$-predictable process ϑ can be identified with a self-financing strategy φ, whose value process is then given by $V(\varphi) = V_0 + G(\vartheta) = V_0 + \int \vartheta \, \mathrm{d}S = V(V_0, \vartheta)$. We shortly write $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$. (Of course, we work throughout in units of asset 0.) Hence $G(\vartheta) = V(0, \vartheta)$ describes the cumulative gains or losses one can generate from initial capital 0 through self-financing trading via $\varphi \mathrel{\widehat{=}} (0, \vartheta)$. We also recall that a strategy φ is a-admissible if $V(\varphi) \geq -a$, and admissible if it is a-admissible for some $a \geq 0$. Note that these notions depend on the chosen accounting unit or numeraire (here S^0), except for 0-admissibility.

Definition. An *arbitrage opportunity* is an admissible self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with zero initial wealth, with $V_T(\varphi) \geq 0$ P-a.s. and with $P[V_T(\varphi) > 0] > 0$. The financial market $(\Omega, \mathcal{F}, I\!F, P, S^0 \equiv 1, S)$ or shortly S is called *arbitrage-free* if there exist no arbitrage opportunities. Sometimes one also says that S *satisfies (NA)*.

Interpretation. An arbitrage opportunity produces *something* (nonnegative final wealth $V_T(\varphi) \geq 0$, with a genuine chance of having strictly positive final wealth) *out of nothing* (zero initial capital) *without any risk* (because the strategy is self-financing). In a well-functioning market, such "money pumps" cannot exist (for long) because they would quickly be exploited and hence would vanish. So absence of arbitrage is a *natural eco-*

nomic/financial requirement for a reasonable model of a financial market.

Remarks. 1) An arbitrage opportunity in the sense of the above definition is actually a specific form of an arbitrage opportunity of the first kind. More generally, one can look at self-financing strategies $\varphi \widehat{=} (V_0, \vartheta)$ with $V_T(\varphi) = V_0 + G_T(\vartheta) \geq 0$ P-a.s. and $V_0(\varphi) \leq 0$ P-a.s. An *arbitrage opportunity of the first kind* then has in addition $P[V_T(\varphi) > 0] > 0$, while an *arbitrage opportunity of the second kind* has in addition $P[V_0(\varphi) < 0] > 0$.

2) One can also introduce the condition (NA_+) which says that it is impossible to produce something out of nothing with *0-admissible* self-financing strategies, or (NA') which does the same for *all* (not necessarily admissible) self-financing strategies. Then we clearly have the implications $(\mathrm{NA}') \Longrightarrow (\mathrm{NA}) \Longrightarrow (\mathrm{NA}_+)$, and the distinction is important in continuous time or with an infinite time horizon. But for finite discrete time, the three concepts are all equivalent; see Proposition 1.1 below. $\diamond$

Example. If there exist an asset i_0 and a date k_0 such that $S^{i_0}_{k_0+1} \leq S^{i_0}_{k_0}$ P-a.s. and $P[S^{i_0}_{k_0+1} < S^{i_0}_{k_0}] > 0$, then S admits arbitrage.

Indeed, the price process S^{i_0} can only go down from time k_0 to $k_0 + 1$ and does so in some cases (i.e., with positive probability); so if we sell short that asset at time k_0, we run no risk and have the chance of a genuine profit. Formally, the strategy $\varphi \widehat{=} (0, \vartheta)$ with

$$\vartheta^i_{k+1} := -I_{\{i=i_0\}} I_{\{k+1=k_0\}} \qquad \text{for } k = 0, 1, \ldots, T-1$$

gives an arbitrage opportunity, as one easily checks. [$\rightarrow$ *Exercise*] This also illustrates the well-known wisdom that "bad news is better than no news" .

Let us introduce a useful notation. For any σ-field $\mathcal{G} \subseteq \mathcal{F}$, we denote by $L^0_{(+)}(\mathcal{G})$ the space of all (equivalence classes, for the relation of equality P-a.s., of) (nonnegative) $\mathcal{G}$-measurable random variables. Then for example, we can write $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$ more compactly as $V_T(\varphi) \in L^0_+(\mathcal{F}_T) \setminus \{0\}$.

Proposition 1.1. *For a discounted financial market in finite discrete time, the following are equivalent:*

1) *S is arbitrage-free.*

2) *There exists no self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with zero initial wealth and satisfying $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$; in other words, S satisfies (NA′).*

3) *For every (not necessarily admissible) self-financing strategy φ with $V_0(\varphi) = 0$ P-a.s. and $V_T(\varphi) \geq 0$ P-a.s., we have $V_T(\varphi) = 0$ P-a.s.*

4) *For the space*

$$\mathcal{G}' := \{G_T(\vartheta) : \vartheta \text{ is } \mathbb{R}^d\text{-valued and predictable}\}$$

of all final wealths that one can generate from zero initial wealth through some self-financing trading $\varphi \mathrel{\widehat{=}} (0, \vartheta)$, we have

$$\mathcal{G}' \cap L^0_+(\mathcal{F}_T) = \{0\}.$$

Remarks. **1)** Proposition 1.1 and its proof substantiate the above comment that all three above formulations for absence of arbitrage are equivalent in finite discrete time.

2) The mathematical relevance of Proposition 1.1 is that it translates the no-arbitrage condition (NA) into the formulation in 4) which has a very useful *geometric interpretation*. We shall exploit this in the next section. ◇

Proof of Proposition 1.1. "2) ⇔ 3)" is obvious, and "2) ⇔ 4)" is a direct consequence of the parametrisation of self-financing strategies in Proposition 1.2.3. It is also clear that (NA′) as in 2) implies (NA) as in 1). Finally, the argument for "1) ⇒ 2)" is indirect and even shows a bit more: We claim that if one has a self-financing strategy φ which produces something out of nothing, one can construct from φ a *0-admissible* self-financing strategy $\tilde{\varphi}$ which also produces something out of nothing. Indeed, if φ is not already 0-admissible itself, then the set $A_k := \{V_k(\varphi) < 0\}$ has $P[A_k] > 0$ for some k. We take as k_0 the largest of these k and then define $\tilde{\varphi}$ simply as the strategy φ on A_{k_0} after time k_0. In words, we wait until we can start on some set with a negative initial capital and transform that via φ into something nonnegative. As this turns something nonpositive

into something nonnegative and keeps wealth nonnegative by construction, it produces the desired arbitrage opportunity.

(Writing out the above verbal argument in formal terms and checking all the details is an excellent [→ *exercise*] necessarily increase the financial understanding.) **q.e.d.**

Our next intermediate goal is to give a simple probabilistic condition that excludes arbitrage opportunities. Recall that two probability measures Q and P on $\mathcal{F}$ are *equivalent* (on $\mathcal{F}$), written as $Q \approx P$ (on $\mathcal{F}$), if they have the same nullsets (in $\mathcal{F}$), i.e. if for each set A (in $\mathcal{F}$), we have $P[A] = 0$ if and only if $Q[A] = 0$. Intuitively, this means that while P and Q may differ in their quantitative assessments, they qualitatively agree on what is "possible or impossible".

Example. If we construct the *multinomial model* as in Section 1.4 as an event tree on the canonical path space $\Omega = \{1, \dots, m\}^T$ with $\mathcal{F} = 2^\Omega$, then we know that any probability measure on $(\Omega, \mathcal{F})$ can be described by its collection of one-step transition probabilities, which all lie between 0 and 1, i.e. in $[0, 1]$.

Now consider two probability measures P and Q on $(\Omega, \mathcal{F})$. If some of the transition probabilities p_{ij} of P are 0 (or 1), a characterisation of Q being equivalent to P is a bit involved, and so we assume (as for example in the multinomial model) that $P[\{\omega\}] > 0$ for all $\omega \in \Omega$. This means that all one-step transition probabilities p_{ij} for P lie in the open interval $(0, 1)$, and then we have $Q \approx P$ if and only if all one-step transition probabilities q_{ij} for Q lie in $(0, 1)$ as well.

Now we go back to the general case.

Lemma 1.2. *If there exists a probability measure $Q \approx P$ on $\mathcal{F}_T$ such that S is a Q-martingale, then S is arbitrage-free.*

Proof. If S is a Q-martingale and $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ is an admissible self-financing strategy, then $V(\varphi) = G(\vartheta) = \vartheta \bullet S$ is a stochastic integral of S and uniformly bounded below (by

some $-a$ with $a \geq 0$). By Theorem 1.3.3, $V(\varphi)$ is thus also a Q-martingale and so

$$E_Q[V_T(\varphi)] = E_Q[V_0(\varphi)] = 0.$$

Now suppose in addition that $Q \approx P$ on $\mathcal{F}_T$, so that Q-a.s. and P-a.s. are the same thing for all events in $\mathcal{F}_T$. If $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ is an admissible self-financing strategy with $V_T(\varphi) \geq 0$ P-a.s., then also $V_T(\varphi) \geq 0$ Q-a.s. But $E_Q[V_T(\varphi)] = 0$ by the above argument, and so we must have $V_T(\varphi) = 0$ Q-a.s., hence also $V_T(\varphi) = 0$ P-a.s. By Proposition 1.1, S is therefore arbitrage-free. **q.e.d.**

Remark 1.3. 1) It would be enough if S is only a *local* Q-martingale, because we could still use Theorem 1.3.3.

2) An alternative proof of Lemma 1.2 goes as follows. This is attractive because it proves a more general result, and the proof still works (with one reference changed) in continuous or infinite discrete time. Suppose that $Q \approx P$ on $\mathcal{F}_T$ is such that S is a local Q-martingale and take an admissible self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$. Then $V(\varphi) = G(\vartheta) = \vartheta \bullet S$ is a local Q-martingale by Theorem 1.3.1, with $V_0(\varphi) = 0$, and $V(\varphi)$ is bounded below because φ is admissible. (In continuous time, the argument and reference here are bit different.) But then $V(\varphi)$ is a Q-supermartingale (this is easily argued via localising and passing to the limit with the help of Fatou's lemma [$\rightarrow$ *exercise*]), and so we get $E_Q[V_T(\varphi)] \leq E_Q[V_0(\varphi)] = 0$. If in addition $V_T(\varphi) \geq 0$ P-a.s., we also get $V_T(\varphi) \geq 0$ Q-a.s., hence $V_T(\varphi) = 0$ Q-a.s., and then also $V_T(\varphi) = 0$ P-a.s. This allows us to conclude as before.

3) We can also give a complete proof of Lemma 1.2 which relies only on proved results. We still use that with $\varphi \mathrel{\widehat{=}} (0, \vartheta)$, we have $V(\varphi) = G(\vartheta) = \vartheta \bullet S$. Now because ϑ is predictable, the process $\vartheta^{(n)}$ defined by $\vartheta_k^{(n)} := \vartheta_k I_{\{|\vartheta_k| \leq n\}}$ is again predictable and bounded. So if S is a martingale under Q, then $\vartheta^{(n)} \bullet S$ is again a Q-martingale by (the simple and proved part of) Theorem 1.3.1. Moreover, the definition of $\vartheta^{(n)}$ yields

$$-(\vartheta_k^{(n)})^{\mathrm{tr}} \Delta S_k = -\vartheta_k^{\mathrm{tr}} \Delta S_k I_{\{|\vartheta_k| \leq n\}} \leq -\vartheta_k^{\mathrm{tr}} \Delta S_k I_{\{\vartheta_k^{\mathrm{tr}} \Delta S_k \leq 0\}} I_{\{|\vartheta_k| \leq n\}} \leq -\vartheta_k^{\mathrm{tr}} \Delta S_k I_{\{\vartheta_k^{\mathrm{tr}} \Delta S_k \leq 0\}}$$

so that $((\vartheta_k^{(n)})^{\mathrm{tr}} \Delta S_k)^- \leq (\vartheta_k^{\mathrm{tr}} \Delta S_k)^-$ for all k and hence $(\vartheta^{(n)} \bullet S)^- \leq (\vartheta \bullet S)^-$. But $V(\varphi)$ is bounded below by $-a$ because φ is admissible, and therefore the entire sequence

$(G(\vartheta^{(n)}))_{n\in\mathbb{N}} = (\vartheta^{(n)} \bullet S)_{n\in\mathbb{N}}$ is also bounded below by $-a$. This allows us to use Fatou's lemma and conclude from the martingale property of each $G(\vartheta^{(n)})$ that $V(\varphi) = \vartheta \bullet S$ is a Q-supermartingale; indeed,

$$E_Q[G_k(\vartheta) \,|\, \mathcal{F}_{k-1}] = E_Q\Big[\lim_{n\to\infty} G_k(\vartheta^{(n)}) \,\Big|\, \mathcal{F}_{k-1}\Big] \le \liminf_{n\to\infty} E_Q[G_k(\vartheta^{(n)}) \,|\, \mathcal{F}_{k-1}]$$
$$= \liminf_{n\to\infty} G_{k-1}(\vartheta^{(n)}) = G_{k-1}(\vartheta).$$

Then we can finish the proof as before in 2).

4) In continuous time, Theorem 1.3.3 no longer holds; then it is useful and important to have for proofs the alternative route via 2). Also for discrete but infinite time, one must be careful about the behaviour at ∞. $\diamond$

Example. Consider the *multinomial model* on the canonical path space $\Omega = \{1, \dots, m\}^T$ and suppose as usual that $P[\{\omega\}] > 0$ for all $\omega \in \Omega$. (We can also assume that the returns $Y_1, \dots, Y_T$ are i.i.d. under P, but this is actually not needed for the subsequent reasoning.) To find $Q \approx P$ such that $S^1 = \widetilde{S}^1/\widetilde{S}^0$ is a Q-martingale (recall that we always work in units of asset 0), we need to find one-step transition probabilities in the open interval $(0, 1)$ such that

$$E_Q[\widetilde{S}^1_k/\widetilde{S}^0_k \,|\, \mathcal{F}_{k-1}] = \widetilde{S}^1_{k-1}/\widetilde{S}^0_{k-1} \qquad \text{for all } k.$$

Because

$$\frac{\widetilde{S}^1_k/\widetilde{S}^0_k}{\widetilde{S}^1_{k-1}/\widetilde{S}^0_{k-1}} = \frac{\widetilde{S}^1_k/\widetilde{S}^1_{k-1}}{\widetilde{S}^0_k/\widetilde{S}^0_{k-1}} = \frac{Y_k}{1+r},$$

we equivalently need $E_Q[Y_k/(1+r) \,|\, \mathcal{F}_{k-1}] = 1$ for all k.

Now fix k and look at a node corresponding to an atom $A^{(k-1)} = A_{\bar{x}_1,\dots,\bar{x}_{k-1}}$ of $\mathcal{F}_{k-1}$ at time $k-1$ with corresponding one-step transition probabilities $q_1, \dots, q_m$. (We sometimes omit to write the indices for $q_j = q_j(A^{(k-1)}) = q_j(\bar{x}_1, \dots, \bar{x}_{k-1})$, but of course the one-step transition probabilities can depend on the atom $A^{(k-1)}$ and hence on the time k.) For the associated probability measure Q, the quantities $q_j(A^{(k-1)}) = Q[Y_k = 1 + y_j \,|\, A^{(k-1)}]$ for branch $j = 1, \dots, m$ then describe the (one-step) conditional distribution of Y_k given

$\mathcal{F}_{k-1}$ at that node, and so

$$\text{on the atom } A^{(k-1)}, \qquad E_Q[Y_k \,|\, \mathcal{F}_{k-1}] = E_Q[Y_k \,|\, A^{(k-1)}]$$
$$= \sum_{j=1}^{m} q_j(A^{(k-1)})(1+y_j)$$
$$= 1 + \sum_{j=1}^{m} q_j(A^{(k-1)})y_j$$

which implies that

$$E_Q[Y_k \,|\, \mathcal{F}_{k-1}] = \sum_{\text{atoms } A^{(k-1)} \in \mathcal{F}_{k-1}} I_{A^{(k-1)}} E_Q[Y_k \,|\, A^{(k-1)}]$$
$$= 1 + \sum_{\text{atoms } A^{(k-1)} \in \mathcal{F}_{k-1}} I_{A^{(k-1)}} q_j(A^{(k-1)})y_j,$$

and we want this to equal $1+r$. Note that although we have started with a particular time k and atom $A^{(k-1)}$, the resulting condition always looks the same; this is due to the homogeneity in the structure of the multinomial model. The above conditional expectation equals $1+r$ if and only if the equation

$$\sum_{j=1}^{m} q_j(A^{(k-1)})y_j = r$$

has a solution $q_1(A^{(k-1)}), \ldots, q_m(A^{(k-1)})$. Because we want all the $q_j(A^{(k-1)})$ to lie in $(0,1)$ and because we have $y_m > y_{m-1} > \cdots > y_1 > -1$ by the assumed labelling, this can clearly be achieved if and only if $y_m > r > y_1$, i.e. if and only if the riskless interest rate r for the bank account lies strictly between the smallest and largest return values, y_1 and y_m, for the stock. Moreover, we can then choose the $q_j(A^{(k-1)})$ independently of k and $A^{(k-1)}$, and if we do that, the corresponding probability measure Q has the property that the returns $Y_1, \ldots, Y_T$ are i.i.d. under Q. But we also see that there are clearly many $Q' \approx P$ on $\mathcal{F}_T$ such that $\widetilde{S}^1/\widetilde{S}^0$ is a Q'-martingale, but $Y_1, \ldots, Y_T$ are not i.i.d. under Q'.

In summary, we obtain the following result.

Corollary 1.4. *In the multinomial model with parameters $y_1 < \cdots < y_m$ and r, there exists a probability measure $Q \approx P$ such that $\widetilde{S}^1/\widetilde{S}^0$ is a Q-martingale if and only if $y_1 < r < y_m$.*

The *interpretation* of the condition $y_1 < r < y_m$ is very intuitive. It says that in comparison to the riskless bank account $\widetilde{S}^0$, the stock $\widetilde{S}^1$ has the potential for both higher and lower growth than $\widetilde{S}^0$. Hence $\widetilde{S}^1$ is genuinely more risky than $\widetilde{S}^0$. One has the feeling that this should not only be sufficient to exclude arbitrage opportunities, but necessary as well. That feeling is correct, as we shall see in the next section; alternatively, one can also prove this directly. [→ *Exercise*]

For the special case of the binomial model, we can even say a bit more.

Corollary 1.5. *In the binomial model with parameters $u > d$ and r, there exists a probability measure $Q \approx P$ such that $\widetilde{S}^1/\widetilde{S}^0$ is a Q-martingale if and only if $u > r > d$. In that case, Q is unique (on $\mathcal{F}_T$) and characterised by the property that $Y_1, \ldots, Y_T$ are i.i.d. under Q with parameter*

$$Q[Y_k = 1 + u] = q^* = \frac{r - d}{u - d} = 1 - Q[Y_k = 1 + d].$$

Proof. The martingale condition $\sum_{j=1}^m q_j(A^{(k-1)})y_j = r$ reduces, with $m = 2$, $y_1 = d$, $y_2 = u$ and $q := q_2(A^{(k-1)})$, to the equation $(1 - q)d + qu = r$, which has the unique solution q^*. Because the one-step transition probabilities for Q are thus the same in each node throughout the tree, the i.i.d. description under Q follows as in Section 1.4 and in the preceding discussion. **q.e.d.**

2.2 The fundamental theorem of asset pricing

We have already seen in Lemma 1.2 a sufficient condition for S to be arbitrage-free. Moreover, the multinomial model has led us to suspect that this condition might be necessary as well. In this section, we shall prove that this is indeed so, for every financial market model in finite discrete time. To give the result a crisp formulation, we first introduce a new and very important concept.

Definition. An *equivalent (local) martingale measure (E(L)MM)* for S is a probability measure Q equivalent to P on $\mathcal{F}_T$ such that S is a (local) Q-martingale. We denote by $I\!P_{\mathrm{e}}(S)$ or simply $I\!P_{\mathrm{e}}$ the set of all EMMs for S and by $I\!P_{\mathrm{e,loc}}$ the set of all ELMMs for S. Clearly, $I\!P_{\mathrm{e}} \subseteq I\!P_{\mathrm{e,loc}}$.

Saying that $I\!P_{\mathrm{e(,loc)}}(S)$ is non-empty is the same as saying that there exists an equivalent (local) martingale measure Q for S. By Lemma 1.2 and the discussion around it, both these properties imply that S is arbitrage-free or, equivalently, that S satisfies (NA). It is very remarkable and important that the converse implication holds as well.

Theorem 2.1 (**Dalang/Morton/Willinger**)**.** *Consider a (discounted) financial market model in finite discrete time. Then S is arbitrage-free if and only if there exists an equivalent martingale measure for S. In brief:*

$$\textit{(NA)} \quad \Longleftrightarrow \quad I\!P_{\mathrm{e}}(S) \neq \emptyset \quad \Longleftrightarrow \quad I\!P_{\mathrm{e,loc}}(S) \neq \emptyset.$$

This result deserves a number of *comments*:

1) The crucial *significance* of Theorem 2.1 is that it translates the economic/financial condition of absence of arbitrage into an equivalent, purely mathematical/probabilistic condition. This opens the door for the use of martingale theory, with its many tools and results, for the study of financial market models.

2) The classical theorems in martingale theory on gambling say that one cannot win in a systematic way if one bets on a martingale (see the stopping theorem or Doob's systems

theorem). Theorem 2.1 can be viewed as a *converse*; it says that if one cannot win by betting on a given process, then that process must be a martingale — at least after an equivalent change of probability measure.

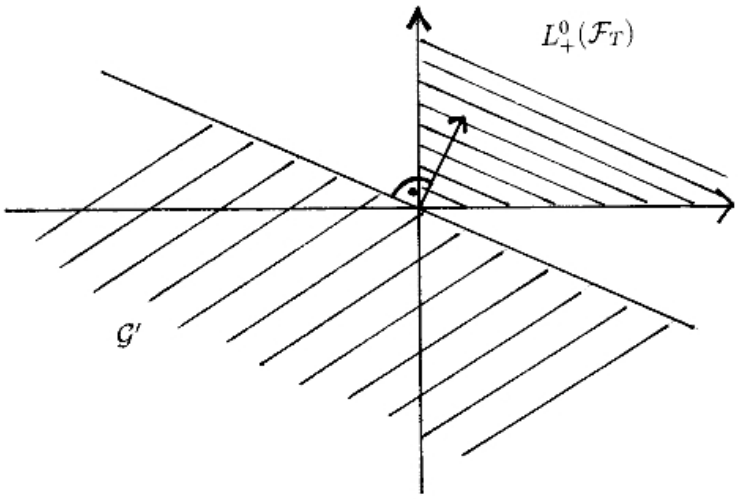
3) Note that we make no integrability assumptions about S (under P); so it is also noteworthy that S, being a Q-martingale, is automatically integrable under (some) Q. (To put this into perspective, one should add that it is a minor point; one can always easily construct [$\rightarrow$ *exercise*] a probability measure R equivalent to P such that S becomes under R as nicely integrable as one wants. But of course such an R will in general not be a martingale measure for S.)

Proving Theorem 2.1 is not elementary if one wants to allow models where the underlying probability space $(\Omega, \mathcal{F}, P)$ is infinite, or more precisely if one of the σ-fields $\mathcal{F}_k$, $k \leq T$, is infinite. This level of generality is needed very quickly, for instance as soon as we want to work with returns which take more than only a finite number of values; the simplest example would be to have the Y_k lognormal, and other typical examples come up when one wants to study GARCH-type models. In that sense, the result in Theorem 2.1 is really needed in full generality. However, we content ourselves here with an explanation of the *key geometric idea* behind the proof, and with the exact argument for the case where Ω (or rather $\mathcal{F}_T$) is finite (like for instance in the canonical setting for the multinomial model).

Due to Lemma 1.2 (plus Remark 1.3) and $\mathbb{P}_{\mathrm{e}} \subseteq \mathbb{P}_{\mathrm{e,loc}}$, we only need to prove that absence of arbitrage implies the existence of an equivalent martingale measure for S. By Proposition 1.1, (NA) is equivalent to $\mathcal{G}' \cap L^0_+(\mathcal{F}_T) = \{0\}$, where

$$\mathcal{G}' = \{G_T(\vartheta) : \vartheta \text{ is } \mathbb{R}^d\text{-valued and predictable}\}$$

is the space of all final positions one can generate from initial wealth 0 by self-financing (but not necessarily admissible) trading. In geometric terms, this means that the upper-right quadrant of nonnegative random variables, $L^0_+(\mathcal{F}_T)$, intersects the linear subspace $\mathcal{G}'$ only in the point 0.



Graphical illustration of the condition $\mathcal{G}' \cap L^0_+(\mathcal{F}_T) = \{0\}$

As a consequence, the two sets $L^0_+(\mathcal{F}_T)$ and $\mathcal{G}'$ can be *separated by* a *hyperplane*, and the normal vector defining that hyperplane then yields (after suitable normalisation) the (density of the) desired EMM.

As one can see from the above scheme of proof, the existence of an EMM follows from the existence of a separating hyperplane between two sets. In that sense, the proof is (not surprisingly) not constructive, and it is also clear that we cannot expect uniqueness of an EMM in general. The latter fact can also easily be seen directly: Because the set $\mathbb{P}_{\mathrm{e}}(S)$ is obviously convex [$\rightarrow$ *exercise*], it is either empty, or contains exactly one element, or contains infinitely (uncountably) many elements.

Proof of Theorem 2.1 for Ω (or $\mathcal{F}_T$) finite. If Ω (or $\mathcal{F}_T$) is finite, then every random variable on $(\Omega, \mathcal{F}_T)$ can take only a finite number (n, say) of values, and so we can identify $L^0(\mathcal{F}_T)$ with the finite-dimensional space $\mathbb{R}^n$ and $L^0_+(\mathcal{F}_T)$ with $\mathbb{R}^n_+$. (More precisely, as pointed out below, we must take n as the number of atoms of $\mathcal{F}_T$.) The set $\mathcal{G}' \subseteq L^0(\mathcal{F}_T)$, which is obviously linear, can then be identified with a linear subspace $\mathcal{H}$ of $\mathbb{R}^n$, and so (NA) translates into $\mathcal{H} \cap \mathbb{R}^n_+ = \{0\}$ due to Proposition 1.1.

Recall that a set $A \in \mathcal{F}_T$ is an atom in $\mathcal{F}_T$ if $P[A] > 0$ and if any $B \in \mathcal{F}_T$ with $B \subseteq A$ has either $P[B] = 0$ or $P[B] = P[A]$. Then any $\mathcal{F}_T$-measurable random variable Z has

the form $Z = \sum_{A \text{ atom in } \mathcal{F}_T} Z I_A = \sum_{A \text{ atom in } \mathcal{F}_T} z_A I_A$ with $z_A \in \mathbb{R}$. We consider the set of all $\mathcal{F}_T$-measurable $Z \geq 0$ with $\sum_{A \text{ atom in } \mathcal{F}_T} z_A = 1$ and identify this with the subset

$$\mathcal{K} = \left\{ z \in \mathbb{R}^n_+ : \sum_{i=1}^n z_i = 1 \right\}$$

of $\mathbb{R}^n_+$, where n denotes the (finite, by assumption) number of atoms in $\mathcal{F}_T$. Then $\mathcal{K} \subseteq \mathbb{R}^n_+$ and $\mathcal{K}$ does not contain the vector 0, so that we must have $\mathcal{H} \cap \mathcal{K} = \emptyset$. Moreover, $\mathcal{K}$ is convex and compact, and so a classical separation theorem for sets in $\mathbb{R}^n$ (see e.g. Lamberton/Lapeyre [12, Theorem A.3.2] implies that there exists a vector $\lambda \in \mathbb{R}^n$ with $\lambda \neq 0$ such that

$$\lambda^{\text{tr}} h = 0 \qquad \text{for all } h \in \mathcal{H} \tag{2.1}$$

(which says that λ is a normal vector to the hyperplane separating $\mathcal{H}$ and $\mathcal{K}$) and

$$\lambda^{\text{tr}} z > 0 \qquad \text{for all } z \in \mathcal{K} \tag{2.2}$$

(which says that the hyperplane strictly separates $\mathcal{H}$ and $\mathcal{K}$).

Now we normalise λ. By the definition of $\mathcal{K}$, choosing as z in turn all the unit coordinate vectors in $\mathbb{R}^n$, the property (2.2) implies that all coordinates of λ must be strictly positive, and so the numbers

$$\rho_i := \frac{\lambda_i}{\sum_{i=1}^n \lambda_i}$$

lie in $(0, 1)$ and sum to 1 so that they define a probability measure Q on $\mathcal{F}_T$ via

$$Q[A_i] := \rho_i \qquad \text{for all atoms } A_i \text{ of } \mathcal{F}_T;$$

recall that $\mathcal{F}_T$ by assumption has only n atoms because it is finite, and any set in $\mathcal{F}_T$ is a union of atoms in $\mathcal{F}_T$. Because $P[A] > 0$ for all n atoms $A \in \mathcal{F}_T$, it is clear that Q is equivalent to P on $\mathcal{F}_T$; and the property (2.1) that $\lambda^{\text{tr}} h = 0$ for all $h \in \mathcal{H}$ translates via the identification of $\mathcal{H}$ and $\mathcal{G}'$ and the definition of $\mathcal{G}'$ into

$$E_Q[G_T(\vartheta)] = 0 \qquad \text{for all } \mathbb{R}^d\text{-valued predictable } \vartheta.$$

Choosing $\vartheta := I_{\{\text{time}\,=\,k\}} I_{\{\text{asset number}\,=\,i\}} I_A$ with $A \in \mathcal{F}_{k-1}$ gives $G_T(\vartheta) = I_A(S^i_k - S^i_{k-1})$. But the fact that this has Q-expectation 0 for arbitrary $A \in \mathcal{F}_{k-1}$ simply means that $E_Q[S^i_k - S^i_{k-1} \,|\, \mathcal{F}_{k-1}] = 0$ for all k, and so S is clearly a Q-martingale. Note that integrability is not an issue here because Ω (or $\mathcal{F}_T$) is finite. **q.e.d.**

In continuous time or with an infinite time horizon, existence of an EMM still implies (NA), but the converse is not true. One needs a sort of topological strengthening which excludes not only arbitrage from each single strategy, but also the possibility of creating "arbitrage in the limit by using a sequence of strategies". The resulting condition is called *(NFLVR)* for *"no free lunch with vanishing risk"*, and the corresponding equivalence theorem, due to Freddy Delbaen and Walter Schachermayer in its most general form, is called the *fundamental theorem of asset pricing (FTAP)*. (To be accurate, we should mention that also the concept of EMM must be generalised a little to obtain that theorem.) The basic idea for proving the FTAP is still the same as in our above proof, but the techniques and arguments are much more advanced. One reason is that for infinite $\mathcal{F}_k$, $k \leq T$, already the proof of Theorem 2.1 needs separation arguments for *infinite-dimensional spaces*. The second, more important reason is that the continuous-time formulation also needs the full arsenal and machinery of *general stochastic calculus* for semimartingales. This is rather difficult. For a detailed treatment, we refer to Delbaen/Schachermayer [4, Chapters 8, 9, 14]

Remark. While Theorem 2.1 is a very nice result, one should also be aware of its *assumptions* and in consequence its limitations. The most important of these assumptions are frictionless markets and small investors — and if one tries to relax these to have more realism, the theory even in finite discrete time becomes considerably more complicated and partly does not even exist yet. The same of course applies to continuous-time models and theorems. ◇

In some specific models, we have already studied when there exists a probability measure $Q \approx P$ such that $\widetilde{S}^1/\widetilde{S}^0$ is a Q-martingale; see Corollaries 1.4 and 1.5. Combining

this with Theorem 2.1 now immediately gives the following results.

Corollary 2.2. *The multinomial model with parameters $y_1 < \cdots < y_m$ and r is arbitrage-free if and only if $y_1 < r < y_m$.*

Note that this confirms the intuition stated after Corollary 1.4.

Corollary 2.3. *The binomial model with parameters $u > d$ and r is arbitrage-free if and only if $u > r > d$. In that case, the EMM Q^* for $\widetilde{S}^1/\widetilde{S}^0$ is unique (on $\mathcal{F}_T$) and is given as in Corollary 1.5.*

2.3 Equivalent (martingale) measures

We can already see from the FTAP in its simplest form in Theorem 2.1 that EMMs play an important role in mathematical finance. This becomes even more pronounced when we turn to questions of option pricing or hedging, as we shall see in later chapters. In this section, we therefore start to study how one can relate computations and probabilistic properties under Q and under P to each other if $Q \approx P$, and we also have a look at how one might actually construct an EMM for a given process S in certain situations.

We begin with $(\Omega, \mathcal{F})$ and a filtration $\mathbb{F} = (\mathcal{F}_k)_{k=0,1,\dots,T}$ in finite discrete time. On $\mathcal{F}$, we have two probability measures Q and P, and we assume that $Q \approx P$. Then the *Radon–Nikodým* theorem tells us that there exists a *density* $\frac{\mathrm{d}Q}{\mathrm{d}P} =: \mathcal{D}$; this is a random variable $\mathcal{D} > 0$ P-a.s. (because $Q \approx P$) such that $Q[A] = E_P[\mathcal{D} I_A]$ for all $A \in \mathcal{F}$, or more generally

$$E_Q[Y] = E_P[Y\mathcal{D}] \qquad \text{for all random variables } Y \geq 0. \tag{3.1}$$

In particular, $E_P[\mathcal{D}] = E_Q[1] = 1$. One sometimes writes (3.1) in integral form as

$$\int_\Omega Y \,\mathrm{d}Q = \int_\Omega Y\mathcal{D} \,\mathrm{d}P,$$

which explains the notation to some extent. The point of these formulae is that they tell us how to compute Q-expectations in terms of P-expectations and vice versa. Sometimes one also writes $\mathcal{D} = \frac{\mathrm{d}Q}{\mathrm{d}P}|_{\mathcal{F}}$ to emphasise that we have $Q[A] = E_P[\mathcal{D} I_A]$ for all $A \in \mathcal{F}$, and one sometimes explicitly calls $\mathcal{D}$ the density of Q with respect to P on $\mathcal{F}$.

To get similar transformation rules for conditional expectations, we introduce the P-martingale Z (sometimes denoted more explicitly by Z^Q or $Z^{Q;P}$) by

$$Z_k := E_P[\mathcal{D} \,|\, \mathcal{F}_k] = E_P\left[\frac{\mathrm{d}Q}{\mathrm{d}P} \,\middle|\, \mathcal{F}_k\right] \qquad \text{for } k = 0, 1, \dots, T.$$

Because $\mathcal{D} > 0$ P-a.s., the process $Z = (Z_k)_{k=0,1,\dots,T}$ is strictly positive in the sense that $Z_k > 0$ P-a.s. for each k, or also $P[Z_k > 0 \text{ for all } k] = 1$. Z is called the *density process (of Q, with respect to P)*; the next result makes it clear why.

Lemma 3.1. 1) *For every $k \in \{0, 1, \dots, T\}$ and any $A \in \mathcal{F}_k$ or any $\mathcal{F}_k$-measurable random variable $Y \geq 0$ or $Y \in L^1(Q)$, we have*

$$Q[A] = E_P[Z_k I_A] \quad \text{and} \quad E_Q[Y] = E_P[Z_k Y],$$

respectively. This means that Z_k is the density of Q with respect to P on $\mathcal{F}_k$, and we also write sometimes $Z_k = \frac{\mathrm{d}Q}{\mathrm{d}P}|_{\mathcal{F}_k}$.

2) *If $j \leq k$ and U_k is $\mathcal{F}_k$-measurable and either ≥ 0 or in $L^1(Q)$, then we have the* Bayes formula

$$E_Q[U_k \,|\, \mathcal{F}_j] = \frac{1}{Z_j} E_P[Z_k U_k \,|\, \mathcal{F}_j] \qquad Q\text{-a.s.} \tag{3.2}$$

This tells us how conditional expectations under Q and P are related to each other.

3) *A process $N = (N_k)_{k=0,1,\dots,T}$ which is adapted to $\mathbb{F}$ is a Q-martingale if and only if the product ZN is a P-martingale. This tells us how martingale properties under P and Q are related to each other.*

The proof of Lemma 3.1 is a standard exercise from probability theory in the use of conditional expectations. We do not give it here, but strongly recommend to do this as an [→ *exercise*]. Note that if $\mathcal{F}_T$ is smaller than $\mathcal{F}$, we have $Z_T \neq \mathcal{D}$ in general.

Because Z is strictly positive, we can define

$$D_k := \frac{Z_k}{Z_{k-1}} \qquad \text{for } k = 1, \dots, T.$$

The process $D = (D_k)_{k=1,\dots,T}$ is adapted, strictly positive and satisfies by its definition

$$E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1,$$

because Z is a P-martingale. Again because Z is a martingale and by Lemma 3.1,

$$E_P[Z_0] = E_P[Z_T] = E_P[Z_T I_\Omega] = Q[\Omega] = 1,$$

and we can of course recover Z from Z_0 and D via

$$Z_k = Z_0 \prod_{j=1}^{k} D_j \qquad \text{for } k = 0, 1, \ldots, T.$$

So every $Q \approx P$ induces via Z a pair (Z_0, D). If we conversely start with a pair (Z_0, D) with the above properties (i.e. Z_0 is $\mathcal{F}_0$-measurable, $Z_0 > 0$ P-a.s. with $E_P[Z_0] = 1$, and D is adapted and strictly positive with $E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1$ for all k), we can define a probability measure $Q \approx P$ via

$$\frac{\mathrm{d}Q}{\mathrm{d}P} := Z_0 \prod_{j=1}^{T} D_j.$$

Written in terms of D, the Bayes formula (3.2) for $j = k - 1$ becomes

$$E_Q[U_k \,|\, \mathcal{F}_{k-1}] = E_P[D_k U_k \,|\, \mathcal{F}_{k-1}]. \tag{3.3}$$

This shows that the ratios D_k play the role of "one-step conditional densities" of Q with respect to P.

The above parametrisation is very simple and yet very useful when we want to *construct* an *equivalent martingale measure* for a given process S. All we need to find are an $\mathcal{F}_0$-measurable random variable $Z_0 > 0$ P-a.s. with $E_P[Z_0] = 1$ and an adapted strictly positive process $D = (D_k)_{k=1,\ldots,T}$ satisfying $E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1$ for all k (these are the properties required to get an equivalent probability measure Q), and in addition $E_P[D_k(S_k - S_{k-1}) \,|\, \mathcal{F}_{k-1}] = 0$ for all k. Indeed, the latter condition is, in view of (3.3), simply the martingale property of S under the measure Q determined by (Z_0, D). (To be accurate, we also need to make sure that S is Q-integrable, meaning that $E_Q[|S_k|] < \infty$ for all k; this amounts to the integrability requirement that $E_P[Z_k|S_k|] < \infty$ for all k, where $Z_k = Z_0 \prod_{j=1}^{k} D_j$.)

The simplest choice for Z_0 is clearly the constant $Z_0 \equiv 1$; this amounts to saying that Q and P should coincide on $\mathcal{F}_0$. If $\mathcal{F}_0$ is P-trivial (i.e. $P[A] \in \{0, 1\}$ for all $A \in \mathcal{F}_0$) as is often the case, then every $\mathcal{F}_0$-measurable random variable is P-a.s. constant, and then $Z_0 \equiv 1$ is actually the only possible choice (because we must have $E_P[Z_0] = 1$).

Concerning the D_k, not much can be said in this generality because we do not have any specific structure for our model. To get more explicit results, we therefore specialise and consider a setting with *i.i.d. returns* under P; this means that

$$\widetilde{S}^1_k = S^1_0 \prod_{j=1}^{k} Y_j, \qquad \widetilde{S}^0_k = (1+r)^k,$$

where $Y_1, \dots, Y_T$ are > 0 and i.i.d. under P. The filtration we use is generated by $(\widetilde{S}^0, \widetilde{S}^1)$ or equivalently by $\widetilde{S}^1$ or by Y; so $\mathcal{F}_0$ is P-trivial and Y_k is under P independent of $\mathcal{F}_{k-1}$ for each k. The Q-martingale condition for $S^1 = \widetilde{S}^1/\widetilde{S}^0$ in multiplicative form is then by (3.3) given by

$$1 = E_Q\left[\frac{S^1_k}{S^1_{k-1}} \,\middle|\, \mathcal{F}_{k-1}\right] = E_Q\left[\frac{\widetilde{S}^1_k/\widetilde{S}^0_k}{\widetilde{S}^1_{k-1}/\widetilde{S}^0_{k-1}} \,\middle|\, \mathcal{F}_{k-1}\right] = E_P\left[\frac{D_k Y_k}{1+r} \,\middle|\, \mathcal{F}_{k-1}\right].$$

Because $S^1 > 0$, this also implies by iteration that $E_Q[|S^1_k|] = E_Q[S^1_k] = E_Q[S^1_0] = S^1_0 < \infty$ so that Q-integrability is automatically included in the martingale condition.

To keep things as simple as possible, we now might try to choose D_k like Y_k independent of $\mathcal{F}_{k-1}$. Then [one can prove that] we must have $D_k = g_k(Y_k)$ for some measurable function g_k, and we have to choose g_k in such a way that we get

$$1 = E_P[D_k \,|\, \mathcal{F}_{k-1}] = E_P[g_k(Y_k)]$$

and

$$1 + r = E_P[D_k Y_k \,|\, \mathcal{F}_{k-1}] = E_P[Y_k g_k(Y_k)].$$

(Note that these calculations both exploit the P-independence of Y_k from $\mathcal{F}_{k-1}$.) If this choice is possible, we can then choose all the $g_k \equiv g_1$, because the Y_k are (assumed) i.i.d. under P and so the distribution of Y_k under P is the same as that of Y_1. To ensure that $D_k > 0$, we can impose $g_k > 0$.

If we find such a function $g_1 > 0$ with $E_P[g_1(Y_1)] = 1$ and $E_P[Y_1 g_1(Y_1)] = 1+r$, setting

$$\frac{\mathrm{d}P}{\mathrm{d}Q} := \prod_{j=1}^{T} g_1(Y_j)$$

defines an EMM Q for $S^1 = \widetilde{S}^1/\widetilde{S}^0$. Moreover, [one can show that] the returns $Y_1, \ldots, Y_T$ are again i.i.d. under Q (but of course not necessarily under an arbitrary EMM Q' for S^1).

Example. We still assume that we have i.i.d. returns under P. If the Y_k are *discrete* random variables taking values $(1+y_j)_{j\in\mathbb{N}}$ with probabilities $P[Y_k = 1+y_j] = p_j$, then g_1 is (for our purposes) determined by its values $g_1(1+y_j)$, and $Q \approx P$ means that we need $q_j := Q[Y_k = 1+y_j] > 0$ for all those j with $p_j > 0$. If we set

$$q_j := p_j g_1(1+y_j),$$

we are thus in more abstract terms looking for q_j having $q_j > 0$ whenever $p_j > 0$ and satisfying

$$1 = E_P[g_1(Y_1)] = \sum_{j\in\mathbb{N}} p_j g_1(1+y_j) = \sum_{j\in\mathbb{N}} q_j$$

and

$$1 + r = E_P[Y_1 g_1(Y_1)] = \sum_{j\in\mathbb{N}} p_j(1+y_j) g_1(1+y_j) = \sum_{j\in\mathbb{N}} q_j(1+y_j) = 1 + \sum_{j\in\mathbb{N}} q_j y_j,$$

or equivalently

$$\sum_{j\in\mathbb{N}} q_j y_j = r.$$

Note that the actual values of the p_j are not relevant here; it only matters which of them are strictly positive.

Example. In the *multinomial model* with parameters $y_1, \ldots, y_m$ and r, the above recipe boils down to finding $q_1, \ldots, q_m > 0$ with $\sum_{j=1}^m q_j = 1$ and $\sum_{j=1}^m q_j y_j = r$. If $m > 2$ and the y_j are as usual all distinct, there is clearly an infinite number of solutions (provided of course that there is at least one).

Example. If we have i.i.d. *lognormal returns*, then $Y_i = e^{\sigma U_i + b}$ with random variables $U_1, \ldots, U_T$ i.i.d. $\sim \mathcal{N}(0,1)$ under P. Instead of $D_i = g_1(Y_i)$, we here try (equivalently) with $D_i = \tilde{g}_1(U_i)$, and more specifically with $D_i = e^{\alpha U_i + \beta}$. Then we have

$$E_P[D_i] = e^{\beta + \frac{1}{2}\alpha^2} = 1 \qquad \text{for } \beta = -\tfrac{1}{2}\alpha^2,$$

and we get

$$E_P[D_i Y_i] = E_P[e^{b+\beta+(\alpha+\sigma)U_i}] = e^{b+\beta+\frac{1}{2}(\alpha+\sigma)^2} = 1 + r$$

for

$$\log(1+r) = b + \beta + \frac{1}{2}(\alpha+\sigma)^2 = b + \frac{1}{2}\sigma^2 + \alpha\sigma,$$

hence

$$\alpha = \frac{1}{\sigma}\left(\log(1+r) - b - \frac{1}{2}\sigma^2\right).$$

So we could for instance take

$$D_k = \exp\left(-\gamma U_k - \frac{1}{2}\gamma^2\right)$$

with

$$\gamma = -\alpha = \frac{b + \frac{1}{2}\sigma^2 - \log(1+r)}{\sigma}.$$

3 Valuation and hedging in complete markets

In Chapter 2, we have characterised those financial market models in finite discrete time that are reasonable in the sense that they do not allow arbitrage. More precisely, we have studied when it is impossible to create money pumps by cleverly combining the basic traded assets (stocks and bank account).

If we now introduce into that market a *new financial instrument* (e.g. an option) and stipulate that this should not create arbitrage opportunities, what can then be said about the *price* of that new instrument? Note that "absence of arbitrage" now takes a different meaning because we consider a different market than before — the basic instruments are now the old stocks, the old bank account, and the new option. Depending on the structure of the stock price process S as well as the structure of the option under consideration, the restrictions on the possible price of the new option can be more or less severe; in the extreme, it can happen that the price of the option is uniquely determined. While this makes things nice and transparent, we should say that this is the exception rather than the rule.

Throughout this chapter, we consider as usual a (discounted) financial market in finite discrete time on $(\Omega, \mathcal{F}, P)$ with $\mathbb{F} = (\mathcal{F}_k)_{k=0,1,\dots,T}$, where discounted asset prices are given by $S^0 \equiv 1$ and $S = (S_k)_{k=0,1,\dots,T}$ with values in $\mathbb{R}^d$. Note that we again express all (discounted) quantities in terms or units of asset 0, and we think of asset 0 as representing money.

3.1 Attainable payoffs

Let us first introduce a general financial instrument of European type.

Definition. A *general European option* or *payoff* or *contingent claim* is a random variable $H \in L^0_+(\mathcal{F}_T)$.

The *interpretation* is that H describes the *net payoff* (in units of asset 0) that the owner of this instrument obtains at time T; so having $H \geq 0$ is natural and also avoids integrability issues. (A bit more generally, one could instead impose that H is bounded

below P-a.s. by some constant.) As H is $\mathcal{F}_T$-measurable, the payoff can depend on the entire information up to time T; and "European" means that the time for the payoff is fixed at the end T.

Remark. We could also deal with an $\mathcal{F}_k$-measurable payoff made at time k; but as $S^0 \equiv 1$, it is financially equivalent whether such a payoff is made at k or at T, because we can use the bank account to transfer money over time without changing it or its value in any way. By using linearity, we could then also deal with *payoff streams* having a payoff at every date k (with, of course, the time k payoff being $\mathcal{F}_k$-measurable, i.e. the payoff stream being an adapted process). However, we do not consider here *American-type* payoffs where the owner of the financial instrument has some additional freedom in choosing the time of the payoff; the theory for that is a bit more complicated. ◇

Example. A *European call option* on asset i with *maturity* T and *strike* K gives its owner the right, but not the obligation, to buy at time T one unit of asset i for the price K, irrespective of what the actual asset price S^i_T then is. Any rational person will make use of (*exercise*) that right if and only if $S^i_T(\omega) > K$, because it is in that, and only in that, situation that the right is more valuable than the asset itself. In that case, in purely monetary terms, the net payoff is then $S^i_T(\omega) - K$, and this is obtained by buying asset i at the low price K and immediately selling it on the market at the high price $S^i_T(\omega)$. In the other case $S^i_T(\omega) \leq K$, the option is clearly worthless — it makes no monetary sense to pay K for one unit of asset i if one can get this on the market for less, namely for $S^i_T(\omega)$. So here we have for the option a net payoff, in monetary terms, of

$$H(\omega) = \max\big(0, S^i_T(\omega) - K\big) = \big(S^i_T(\omega) - K\big)^+.$$

As a random variable, this is clearly nonnegative and $\mathcal{F}_T$-measurable because S^i is adapted. Actually, H here is even simpler because it only depends on the terminal asset price S^i_T; we can write $H = h(S^i_T)$ with the function $h(x) = (x - K)^+$.

Remark. In the above example, and more generally by identifying an option with its net payoff in units of S^0, we are implicitly restricting ourselves to so-called *cash delivery*

of options. However, there might be other contractual agreements. For instance, with a call option with *physical delivery*, one actually obtains at time T in case of exercise the shares or units of the specified asset and has to pay in cash the agreed amount K. If the underlying asset is some commodity like e.g. oil or grain, this distinction becomes quite important. However, we do not discuss this here any further. $\diamond$

Example. If we want to *bet on* a reasonably *stable asset price evolution*, we might be interested in a payoff of the form $H = I_B$ with

$$B = \Big\{ a \leq \min_{i=1,\ldots,d} \min_{k=0,1,\ldots,T} S_k^i < \max_{i=1,\ldots,d} \max_{k=0,1,\ldots,T} S_k^i \leq b \Big\}.$$

In words, this option pays at time T one unit of money if and only if all stocks remain between the levels a and b up to time T. This H is also $\mathcal{F}_T$-measurable, but now depends on the asset price evolution over the whole time range $k = 0, 1, \ldots, T$; it cannot be written as a function of the final stock price S_T alone.

Example. A payoff of the form

$$H = I_A \, g\Big(\frac{1}{T} \sum_{k=1}^{T} S_k^i \Big) \qquad \text{with } A \in \mathcal{F}_T \text{ and a function } g \geq 0$$

gives a payoff which depends on the average price (over time) of asset i, but which is only due in case that a certain event A occurs. In *insurance*, the set A could for instance be the event of the death up to time T of an insured person; then H would describe the payoff from an *index-linked insurance policy*. This is an example where H depends on more than only the basic asset prices. To get interesting examples of this type, we need the filtration $\mathbb{F}$ to be strictly larger than the filtration $\mathbb{F}^S$ generated by asset prices.

The *basic question* studied in this chapter is the following: Given a contingent claim $H \in L^0_+(\mathcal{F}_T)$, how can we assign to H a *value at* any *time* $k < T$ in such a way that this creates no arbitrage opportunities (if the claim is made available for trading at these values)? And having sold H, what can one do to *insure* oneself *against* the *risk* involved in having to pay the random, uncertain amount H at time T?

The *key idea* for answering both questions is very simple. With the help of the basic traded assets S^0 and S, we try to *construct* an *artificial product* that looks as *similar to* H as possible. The value of this product is then known because the product is constructed from the given assets; and this value should by absence of arbitrage be a good approximation for the value of H.

Let us first look at the *ideal case*. Suppose that we can find a self-financing strategy $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ such that $V_T(\varphi) = H$ P-a.s. Then both the strategy φ and just holding H have costs of 0 at all intermediate times $k = 1, \ldots, T-1$ because φ is self-financing, and both have at time T a value of H. To avoid arbitrage, the values of both structures must therefore coincide at time 0 as well, because we can otherwise buy the cheaper and sell the more expensive product to make a riskless profit. (Note that this argument crucially exploits that in finite discrete time, (NA) and (NA′) are equivalent, so that we need not worry about any admissibility condition for the "strategy", in the extended market, of combining two products.) In consequence, the value or price of H at time 0 must be V_0. An analogous argument and conclusion are valid for any time k, where the value or price of H must then be $V_k(\varphi)$.

Definition. A payoff $H \in L^0_+(\mathcal{F}_T)$ is called *attainable* if there exists an admissible self-financing strategy $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ with $V_T(\varphi) = H$ P-a.s. The strategy φ is then said to *replicate* H and is called a *replicating strategy for* H.

Remark. Even in finite discrete time, it is important (and exploited below) that a replicating strategy should be admissible. In continuous or infinite discrete time, this becomes indispensable. ◇

The next result formalises the key idea explained just before the above definition. In addition, it also provides an efficient way of computing the resulting option price.

Theorem 1.1 (Arbitrage-free valuation of attainable payoffs). *Consider a dis-*

counted financial market in finite discrete time and suppose that S is arbitrage-free and $\mathcal{F}_0$ is trivial. Then every attainable payoff H has a unique price process $V^H = (V_k^H)_{k=0,1,\dots,T}$ which admits no arbitrage (in the extended market consisting of 1, S and V^H). It is given by

$$V_k^H = E_Q[H \,|\, \mathcal{F}_k] = V_k(V_0, \vartheta) \qquad \text{for } k = 0, 1, \dots, T,$$

for any equivalent martingale measure Q for S and for any replicating strategy $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ for H.

Proof. By the DMW theorem in Theorem 2.2.1, $\mathbb{P}_{\mathrm{e}}(S)$ is nonempty because S is arbitrage-free; so there is at least one EMM Q. By assumption, H is attainable; so there is at least one replicating strategy φ. Because φ and H provide the same payoff structures, they must by absence of arbitrage in the extended market have the same value processes; so $V^H = V(\varphi)$, and this holds for any replicating φ. Because any such $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ is admissible by definition, $V(\varphi) = V_0 + \vartheta \bullet S = V(V_0, \vartheta)$ is a Q-martingale by Theorem 1.3.3, for any $Q \in \mathbb{P}_{\mathrm{e}}(S)$, and as its final value is $V_T(\varphi) = H$ (P-a.s., hence also Q-a.s.), we get

$$V_k^H = V_k(\varphi) = E_Q[H \,|\, \mathcal{F}_k] \qquad \text{for all } k.$$

More precisely, V_0 is a constant because $\mathcal{F}_0$ is trivial, and φ is admissible so that $V(\varphi)$ is bounded from below. So $\vartheta \bullet S = V(\varphi) - V_0$ is also bounded from below, which justifies the use of Theorem 1.3.3. **q.e.d.**

In terms of *efficiency*, Theorem 1.1 is a substantial achievement. In a first step, we ought to check in any case whether or not the basic model we use for S is arbitrage-free, and that is most easily done by exhibiting or constructing an EMM Q for S. If we then have any attainable payoff, we very simply compute its price process by taking conditional expectations under Q, without having to spend any effort on finding a replication strategy.

However, the above statement is a bit *misleading*. First of all, for hedging purposes, we very often are interested in actually knowing and then also using a replicating strategy. But more fundamentally, how can we decide for a given payoff whether or not it is attainable, without exhibiting or constructing a replicating strategy? Is there a different and maybe simpler way to show the existence of a replicating strategy?

The next result shows how the last question can be answered by again using E(L)MMs for S.

Theorem 1.2 (Characterisation of attainable payoffs). *Consider a discounted financial market in finite discrete time and suppose that S is arbitrage-free and $\mathcal{F}_0$ is trivial. For any payoff $H \in L^0_+(\mathcal{F}_T)$, the following are equivalent:*

1) *H is attainable.*

2) *$\sup_{Q \in I\!P_{\mathrm{e,loc}}(S)} E_Q[H] < \infty$ is attained in some $Q^* \in I\!P_{\mathrm{e,loc}}(S)$, i.e. the supremum is finite and a maximum; in other words, we have $\sup_{Q \in I\!P_{\mathrm{e,loc}}(S)} E_Q[H] = E_{Q^*}[H] < \infty$ for some $Q^* \in I\!P_{\mathrm{e,loc}}(S)$.*

3) *The mapping $I\!P_{\mathrm{e}}(S) \to I\!R$, $Q \mapsto E_Q[H]$ is constant, i.e. H has the same and finite expectation under all EMMs Q for S.*

Proof. While some of the implications are rather straightforward, the full proof, and in particular the implication "2) $\Rightarrow$ 1)", is difficult because it relies on the so-called *optional decomposition theorem*. For the case where prices S are nonnegative, see Föllmer/Schied [9, Remark 7.17 and Theorem 5.32]. The general case is more delicate; the simplification for $S \geq 0$ is due to the fact that the sets $I\!P_{\mathrm{e}}(S)$ and $I\!P_{\mathrm{e,loc}}(S)$ then coincide. A full proof is for instance given in the lecture "Introduction to Mathematical Finance". **q.e.d.**

Remark. For models with continuous or infinite discrete time, the equivalence between 1) and 2) in Theorem 1.2 still holds (with a slightly stronger definition of attainability), but the equivalence between 2) and 3) may (surprisingly!) fail. More precisely, "3) $\Rightarrow$ 2)" remains valid if we replace $I\!P_{\mathrm{e}}$ by $I\!P_{\mathrm{e,loc}}$ in 3), but "2) $\Rightarrow$ 3)" in general only holds if H is bounded; see Delbaen/Schachermayer [4, Chapter 10] for a *counterexample*. $\diamond$

In summary, the approach to valuing and hedging a given payoff H in a financial market in finite discrete time (with $\mathcal{F}_0$ trivial) looks quite simple:

1) Check if S is arbitrage-free by finding at least one ELMM Q for S.

2) Find all ELMMs Q for S.

3) Compute $E_Q[H]$ for all ELMMs Q for S and determine the supremum of $E_Q[H]$ over Q.

4a) If the supremum is finite and a maximum, i.e. attained in some $Q^* \in I\!P_{\mathrm{e,loc}}(S)$, then H is attainable and its price process can be computed as $V_k^H = E_Q[H \,|\, \mathcal{F}_k]$, for any $Q \in I\!P_{\mathrm{e}}(S)$.

4b) If the supremum is not attained (or, equivalently for finite discrete time, there is a pair of EMMs Q_1, Q_2 with $E_{Q_1}[H] \neq E_{Q_2}[H]$), then H is not attainable.

In case 4a), Theorem 1.1 tells us how to *value* H; but if we also want to find a *replicating strategy*, then more work is required.

In case 4b), we are faced with a *genuine problem*: It is impossible to replicate H, so our whole conceptual approach up to here breaks down. We then have the difficult problem of *valuation and hedging* for a *non-attainable payoff*, and there are in the literature several competing approaches to that, all involving in some way the specification of *preferences* or *subjective views* of the option seller.

Remark. Because it involves no preferences, but only the assumption of absence of arbitrage, the valuation from Theorem 1.1 is often also called *risk-neutral valuation*, and an EMM Q for S is called a *risk-neutral measure*. ◇

Warning: In large parts of the literature, the terminology "risk-neutral valuation" is used for computing conditional expectations of a given payoff H under some EMM Q. This is potentially *problematic* for two reasons:

1) $V_k^{H,Q} := E_Q[H \,|\, \mathcal{F}_k]$ typically depends on Q if H is not attainable. So when following that approach, one should at the very least think carefully about which $Q \in I\!P_{\mathrm{e}}(S)$ one uses, and why.

2) If H is not attainable, it is at best not clear how to hedge H in any reasonably safe way, and at worst, this may be impossible to achieve.

Both of these issues are often ignored in the literature; whether this happens intentionally or through ignorance is not always clear. One area where this used to be particularly prominent is credit risk. One can of course argue that having some approach to obtain a valuation is better than nothing; but a value which has substantial arbitrariness and perhaps no clear risk management outlook should certainly be treated with care and respect.

3.2 Complete markets

As we have seen in Theorem 1.1, absence of arbitrage is already enough to value or price any attainable payoff.

Definition. A financial market model (in finite discrete time) is called *complete* if every payoff $H \in L^0_+(\mathcal{F}_T)$ is attainable. Otherwise it is called *incomplete*.

An obvious corollary of Theorem 1.1 is then

Theorem 2.1 (**Valuation and hedging in complete markets**). *Consider a discounted financial market model in finite discrete time and suppose that $\mathcal{F}_0$ is trivial and S is arbitrage-free and complete. Then for every payoff $H \in L^0_+(\mathcal{F}_T)$, there is a unique price process $V^H = (V^H_k)_{k=0,1,\dots,T}$ which admits no arbitrage. It is given by*

$$V^H_k = E_Q[H \,|\, \mathcal{F}_k] = V_k(V_0, \vartheta) \qquad \text{for } k = 0, 1, \dots, T$$

for any EMM Q for S and any replicating strategy $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ for H.

While Theorem 2.1 looks very nice, it raises the important *question* of how to recognise a complete market, because completeness is a statement about *all* payoffs $H \in L^0_+(\mathcal{F}_T)$. But very fortunately, there is a very simple criterion — and it should be no surprise by now that this again involves EMMs Q.

Theorem 2.2. *Consider a discounted financial market model in finite discrete time and assume that S is arbitrage-free, $\mathcal{F}_0$ is trivial and $\mathcal{F}_T = \mathcal{F}$. Then S is complete if and only if there is a unique equivalent martingale measure for S. In brief:*

$$(NA) + \textit{completeness} \quad \Longleftrightarrow \quad \#\big(I\!P_{\mathrm{e}}(S)\big) = 1, \textit{ i.e. } I\!P_{\mathrm{e}}(S) \textit{ is a singleton.}$$

Proof. "$\Longleftarrow$": If $I\!P_{\mathrm{e}}(S)$ contains only one element, then $Q \mapsto E_Q[H]$ is of course constant over $Q \in I\!P_{\mathrm{e}}(S)$ for any $H \in L^0_+(\mathcal{F}_T)$. Hence H is attainable by Theorem 1.2.

[To be accurate and avoid the case that $Q \mapsto E_Q[H] \equiv +\infty$, one also needs to check a priori some integrability issues, namely that $E_Q[H] < \infty$ for at least one $Q \in I\!P_e(S)$; see Föllmer/Schied [9, Theorems 5.30 and 5.26] for details.]

"$\Longrightarrow$": For any $A \in \mathcal{F}_T$, the payoff $H := I_A$ is attainable; so by Theorem 1.1, we have for any pair of EMMs Q_1, Q_2 for S that

$$Q_1[A] = E_{Q_1}[H] = V_0^H = E_{Q_2}[H] = Q_2[A].$$

So Q_1 and Q_2 coincide on $\mathcal{F}_T = \mathcal{F}$, which means that there can be at most one EMM for S. By the DMW theorem in Theorem 2.2.1, there is at least one EMM because S is arbitrage-free, and so the proof is complete. **q.e.d.**

Theorem 2.2 is sometimes called the *second fundamental theorem of asset pricing*. Combining it with the first FTAP in Theorem 2.2.1, we have a very *simple and beautiful description* of discounted financial market models in finite discrete time:

– Existence of an EMM is equivalent to the market being arbitrage-free.

– Uniqueness of the EMM is equivalent to completeness of the market.

For continuous or infinite discrete time, such statements become more subtle to formulate and more difficult to prove.

Remarks. 1) We can see from the proof of Theorem 2.2 where the assumption $\mathcal{F}_T = \mathcal{F}$ is used. But it is also clear from looking at the statement why it is needed; after all, completeness is only an assertion about $\mathcal{F}_T$-measurable quantities.

2) One can show that if a financial market in finite discrete time is complete, then $\mathcal{F}_T$ must be *finite*; see Föllmer/Schied [9, Theorem 5.38]. In effect, finiteness of $\mathcal{F}_T$ means that Ω can also be taken finite. This shows that while it makes the theory nice and simple, completeness is also a very restrictive property — complete financial markets in finite discrete time are effectively given by finite tree models. $\diamond$

Example. The multinomial model with a bank account and one stock ($d = 1$) is incomplete whenever $m > 2$, i.e. as soon as there is some node in the tree which allows

more than two possible stock price evolutions. This follows from Theorem 2.2 because in that situation, there are infinitely many EMMs; see Section 2.3.

Example. Consider any model with $d = 1$ (one risky asset) and i.i.d. returns $Y_1, \dots, Y_T$ under P. If Y_1 has a density (e.g. if we have lognormal returns), then S is incomplete. This is because $\mathcal{F}_1$ (and hence also $\mathcal{F}_T$) must be infinite for Y_1 to have a density. Alternatively, one can easily construct different EMMs if there is at least one. [$\rightarrow$ *Exercise*]

3.3 Example: The binomial model

In this section, we briefly illustrate how the preceding theory works out in the *binomial* or *Cox–Ross–Rubinstein model*. We recall that this model is described by parameters $p \in (0,1)$ and $u > r > d > -1$; then we have $\widetilde{S}^0_k = (1+r)^k$ and $\widetilde{S}^1_k = S^1_0 \prod_{j=1}^k Y_j$ with $S^1_0 > 0$ and $Y_1, \dots, Y_T$ i.i.d. under P taking values $1+u$ or $1+d$ with probability p or $1-p$, respectively. The filtration $\mathbb{F}$ is generated by $\widetilde{S} = (\widetilde{S}^0, \widetilde{S}^1)$ or equivalently by $\widetilde{S}^1$ or by Y. Note that $\mathcal{F}_0$ is then trivial because $\widetilde{S}^0_0 = 1$ and $\widetilde{S}^1_0 = S^1_0$ is a constant. We also take $\mathcal{F} = \mathcal{F}_T$; this is even an automatic conclusion if we construct the model on the canonical path space as in Section 1.4.

We already know from Corollary 2.2.3 that this model is *arbitrage-free* and has a *unique EMM* for $S^1 = \widetilde{S}^1/\widetilde{S}^0$. Hence S^1 is *complete* by Theorem 2.2, and so every $H \in L^0_+(\mathcal{F}_T)$ is *attainable*, with a *price process* given by

$$V^H_k = E_{Q^*}[H \,|\, \mathcal{F}_k] \qquad \text{for } k = 0, 1, \dots, T,$$

where Q^* is the unique EMM for S^1. We also recall from Corollary 2.2.3 that the Y_j are under Q^* again i.i.d., but with

$$Q^*[Y_1 = 1+u] = q^* := \frac{r-d}{u-d} \in (0,1).$$

All the above quantities S^1, H, V^H are discounted with $\widetilde{S}^0$, i.e. expressed in units of asset 0. The *undiscounted quantities* are the stock price $\widetilde{S}^1 = S^1 \widetilde{S}^0$, the payoff $\widetilde{H} := H \widetilde{S}^0_T$ and its price process $\tilde{V}^{\widetilde{H}}$ with $\tilde{V}^{\widetilde{H}}_k := V^H_k \widetilde{S}^0_k$ for $k = 0, 1, \dots, T$. Putting together all we know then yields

Corollary 3.1. *In the binomial model with $u > r > d$, the undiscounted arbitrage-free price process of any undiscounted payoff $\widetilde{H} \in L^0_+(\mathcal{F}_T)$ is given by*

$$\tilde{V}^{\widetilde{H}}_k = \widetilde{S}^0_k E_{Q^*}\left[\frac{\widetilde{H}}{\widetilde{S}^0_T} \,\middle|\, \mathcal{F}_k\right] = E_{Q^*}\left[\widetilde{H}\frac{\widetilde{S}^0_k}{\widetilde{S}^0_T} \,\middle|\, \mathcal{F}_k\right] = \frac{\widetilde{S}^0_k}{\widetilde{S}^0_T} E_{Q^*}[\widetilde{H} \,|\, \mathcal{F}_k] \qquad \text{for } k = 0, 1, \dots, T.$$

Example. For a *European call option* on $\widetilde{S}^1$ with maturity T and undiscounted strike

$\widetilde{K}$, we have

$$\widetilde{H} = (\widetilde{S}_T^1 - \widetilde{K})^+ = (\widetilde{S}_T^1 - \widetilde{K}) I_{\{\widetilde{S}_T^1 > \widetilde{K}\}}.$$

Now

$$\{\widetilde{S}_T^1 > \widetilde{K}\} = \left\{ \widetilde{S}_k^1 \prod_{j=k+1}^{T} Y_j > \widetilde{K} \right\} = \left\{ \sum_{j=k+1}^{T} \log Y_j > \log(\widetilde{K}/\widetilde{S}_k^1) \right\}.$$

If we define

$$W_j := I_{\{Y_j = 1+u\}} = \begin{cases} 1 & \text{if } Y_j = 1+u, \\ 0 & \text{if } Y_j = 1+d, \end{cases}$$

then $W_1, \ldots, W_T$ are under Q^* independent 0-1 experiments with success parameter q^*, so that their sum has under Q^* a *binomial distribution*. Moreover, using the fact that $\log Y_j = W_j \log(1+u) + (1-W_j)\log(1+d) = W_j \log \frac{1+u}{1+d} + \log(1+d)$ gives

$$\sum_{j=k+1}^{T} \log Y_j = W_{k,T} \log \frac{1+u}{1+d} + (T-k)\log(1+d),$$

where $W_{k,T} := \sum_{j=k+1}^{T} W_j \sim \mathrm{Bin}(T-k, q^*)$ is independent of $\mathcal{F}_k$ under Q^*. So we get

$$\{\widetilde{S}_T^1 > \widetilde{K}\} = \left\{ W_{k,T} \log \frac{1+u}{1+d} > \log \frac{\widetilde{K}}{\widetilde{S}_k^1} - (T-k)\log(1+d) \right\}$$

and therefore

$$Q^*[\widetilde{S}_T^1 > \widetilde{K} \,|\, \mathcal{F}_k] = Q^* \left[W_{k,T} > \frac{\log \frac{\widetilde{K}}{s} - (T-k)\log(1+d)}{\log \frac{1+u}{1+d}} \right] \Bigg|_{s=\widetilde{S}_k^1},$$

because $W_{k,T}$ is independent of $\mathcal{F}_k$ under Q^* and $\widetilde{S}_k^1$ is $\mathcal{F}_k$-measurable. The above probability can be computed explicitly because $W_{k,T}$ has a binomial distribution; and as

$$E_{Q^*}[\widetilde{H} \,|\, \mathcal{F}_k] = E_{Q^*}\left[\widetilde{S}_T^1 I_{\{\widetilde{S}_T^1 > \widetilde{K}\}} \,\middle|\, \mathcal{F}_k\right] - \widetilde{K} Q^*[\widetilde{S}_T^1 > \widetilde{K} \,|\, \mathcal{F}_k],$$

we already have the second half of the so-called binomial call pricing formula.

For the first term, one can either use explicit (and lengthy) computations or more elegantly a so-called *change of numeraire* to obtain that

$$
\begin{aligned}
(3.1) \qquad E_{Q^*}\big[\widetilde{S}^1_T I_{\{\widetilde{S}^1_T>\widetilde{K}\}} \,\big|\, \mathcal{F}_k\big] &= \widetilde{S}^1_k \frac{\widetilde{S}^0_T}{\widetilde{S}^0_k}\frac{\widetilde{S}^0_k}{\widetilde{S}^1_k} E_{Q^*}\left[\frac{\widetilde{S}^1_T}{\widetilde{S}^0_T} I_{\{\widetilde{S}^1_T>\widetilde{K}\}} \,\middle|\, \mathcal{F}_k\right] \\
&= \widetilde{S}^1_k \frac{\widetilde{S}^0_T}{\widetilde{S}^0_k} Q^{**}[\widetilde{S}^1_T > \widetilde{K} \,|\, \mathcal{F}_k] \\
&= \widetilde{S}^1_k \frac{\widetilde{S}^0_T}{\widetilde{S}^0_k} Q^{**}\left[W_{k,T} > \frac{\log\frac{\widetilde{K}}{s} - (T-k)\log(1+d)}{\log\frac{1+u}{1+d}}\right]\Bigg|_{s=\widetilde{S}^1_k},
\end{aligned}
$$

where $W_{k,T}$ under Q^{**} is $\mathrm{Bin}(T-k, q^{**})$-distributed with

$$
q^{**} := q^* \frac{1+u}{1+r}, \qquad \text{hence } 1-q^{**} = (1-q^*)\frac{1+d}{1+r}.
$$

Indeed, because $\widetilde{S}^1/\widetilde{S}^0 = S^1$ is under Q^* a positive martingale, one can use it to define via $\mathrm{d}Q^{**}/\,\mathrm{d}Q^* := S^1_T/S^1_0$ a probability measure $Q^{**} \approx Q^*$ on $\mathcal{F}_T$; then the Q^*-martingale S^1/S^1_0 starting at 1 is by construction the density process $Z^{Q^{**};Q^*}$ of Q^{**} with respect to Q^*, and the second equality in (3.1) is due to the Bayes formula (2.3.2) in Lemma 2.3.1. One then easily verifies [$\to$ *exercise*] that Q^{**} is the unique probability measure equivalent to P on $\mathcal{F}_T$ such that $\widetilde{S}^0/\widetilde{S}^1 = 1/S^1$ becomes a Q^{**}-martingale, and one can also check that $Y_1, \ldots, Y_T$ are under Q^{**} i.i.d. with $Q^{**}[Y_1 = 1+u] = q^{**}$. Indeed, this is not really surprising — by Lemma 2.3.1, 3), the process $1/S^1$ is a Q^{**}-martingale because the product $Z^{Q^{**};Q^*}(1/S^1) = (S^1/S^1_0)(1/S^1) \equiv 1/S^1_0$ is obviously a Q^*-martingale, and $1/S^1$ has a binomial structure exactly like S^1 itself. The measure Q^{**} is sometimes called *dual martingale measure*.

So all in all, we obtain the fairly simple formula

$$
(3.2) \qquad \tilde{V}^{\widetilde{H}}_k = \widetilde{S}^1_k Q^{**}[W_{k,T} > x] - \widetilde{K}\frac{\widetilde{S}^0_k}{\widetilde{S}^0_T} Q^*[W_{k,T} > x]
$$

with

$$
(3.3) \qquad x = \frac{\log\frac{\widetilde{K}}{s} - (T-k)\log(1+d)}{\log\frac{1+u}{1+d}}, \qquad \text{for } s = \widetilde{S}^1_k,
$$

and where $W_{k,T}$ has a binomial distribution with parameter $T-k$ and with q^* under Q^*, respectively with q^{**} under Q^{**}. This *binomial call pricing formula* is the *discrete analogue of* the famous *Black–Scholes formula*.

For a *general payoff* $\widetilde{H}$, the discounted price process V^H is by its construction a Q^*-martingale with final value H, so that $V_T^H = H$ and

$$V_{k-1}^H = E_{Q^*}[V_k^H \,|\, \mathcal{F}_{k-1}] \qquad \text{for } k = 1, \ldots, T.$$

This provides a very *simple recursive algorithm* by using that the filtration $\mathbb{F}$ in the binomial model has the structure of a *(binary) tree*. Indeed, if we fix some node (corresponding to some atom) at time $k-1$ (respectively of $\mathcal{F}_{k-1}$) and denote by v_{k-1} the value of V_{k-1}^H there (on that atom), then there are only two possible successor nodes (atoms of $\mathcal{F}_k$) and V_k^H can only take two values there, say v_k^u and v_k^d. The Q^*-martingale property then says that

$$v_{k-1} = q^* v_k^u + (1-q^*) v_k^d,$$

because the one-step transition probabilities of Q^* are the same throughout the tree and given by $q^*, 1-q^*$. In undiscounted terms, we have

$$\frac{\tilde{V}_{k-1}^{\widetilde{H}}}{\widetilde{S}_{k-1}^0} = E_{Q^*}\left[\frac{\tilde{V}_k^{\widetilde{H}}}{\widetilde{S}_k^0} \,\middle|\, \mathcal{F}_{k-1} \right]$$

or

$$\tilde{V}_{k-1}^{\widetilde{H}} = \frac{1}{1+r} E_{Q^*}[\tilde{V}_k^{\widetilde{H}} \,|\, \mathcal{F}_{k-1}],$$

which translates at the level of node values to the *recursion*

$$\tilde{v}_{k-1} = \frac{1}{1+r}\big(q^* \tilde{v}_k^u + (1-q^*)\tilde{v}_k^d\big). \tag{3.4}$$

The *terminal condition* $V_T^H = H$ or $\tilde{V}_T^{\widetilde{H}} = \widetilde{H}$ means that the values v_T or $\tilde{v}_T$ at the terminal nodes are given by the values of $\widetilde{H}$ there. Note that for a general (hence typically path-dependent) payoff $\widetilde{H}$, we have to work with the *full, non-recombining tree* and all its 2^T terminal nodes.

To work out the *replicating strategy*, also for a general payoff H, we recall from Theorem 1.1 that

$$V_k^H = V_k(V_0, \vartheta) = V_0 + \sum_{j=1}^{k} \vartheta_j \Delta S_j^1 \qquad \text{for } k = 0, 1, \ldots, T.$$

For the *increments*, this means that

$$\Delta V_k^H = V_k^H - V_{k-1}^H = \vartheta_k \Delta S_k^1 = \vartheta_k (S_k^1 - S_{k-1}^1). \tag{3.5}$$

Now let us look again at some fixed node at time $k-1$ (atom of $\mathcal{F}_{k-1}$). Because ϑ is predictable, ϑ_k is $\mathcal{F}_{k-1}$-measurable and so the value of ϑ_k is already known at time $k-1$, hence in that node (on that atom), and it cannot change as we move forward to time k. If we denote as before by v_{k-1} the value of V_{k-1}^H in the chosen node (on the chosen atom) at time $k-1$ and by s_{k-1} the value of S_{k-1}^1 there, we know that v_{k-1} evolves to either v_k^u or v_k^d, and s_{k-1} evolves to $s_k^u = s_{k-1}\frac{1+u}{1+r}$ or $s_k^d = s_{k-1}\frac{1+d}{1+r}$, respectively, in the next step. But the relation (3.5) between increments must hold in all nodes (on all atoms) and at all times; so if ξ_k denotes the value of ϑ_k in the chosen node (on the chosen atom) at time $k-1$, we obtain the *two equations*

$$v_k^u - v_{k-1} = \xi_k (s_k^u - s_{k-1}),$$
$$v_k^d - v_{k-1} = \xi_k (s_k^d - s_{k-1}).$$

Note that we have the same ξ_k in both equations because the value of ϑ_k cannot change as we go from time $k-1$ to time k. The above two equations are readily solved to give

$$\xi_k = \frac{v_k^u - v_k^d}{s_k^u - s_k^d} = \frac{v_k^u - v_k^d}{\frac{u-d}{1+r} s_{k-1}}. \tag{3.6}$$

Again, the right-hand side is known at time $k = T$ because we know that $V_T^H = H$. So both the price process V^H and the hedging strategy ϑ can be computed in parallel while working backward through the tree.

If the payoff $\widetilde{H}$ is like the call option of the *simple path-independent form* $\widetilde{H} = \tilde{h}(\widetilde{S}_T^1)$ for some function $\tilde{h}$, then the above formulas and computation scheme simplify considerably.

Indeed one can show by backward induction that

$$\tilde{V}_k^{\tilde{H}} = \tilde{v}(k, \widetilde{S}_k^1) \qquad \text{for } k = 0, 1, \ldots, T$$

and

$$\vartheta_k = \tilde{\xi}(k, \widetilde{S}_{k-1}^1) \qquad \text{for } k = 1, \ldots, T$$

with *functions* $\tilde{v}(k, s)$ and $\tilde{\xi}(k, s)$ that are given by the recursion (compare (3.4))

$$\tilde{v}(k-1, s) = \frac{1}{1+r}\Big(q^* \tilde{v}\big(k, s(1+u)\big) + (1-q^*)\tilde{v}\big(k, s(1+d)\big)\Big)$$

with terminal condition

$$\tilde{v}(T, s) = \tilde{h}(s)$$

and, from (3.6) multiplied in both numerator and denominator by $\widetilde{S}_k^0 = (1+r)^k$, by

$$\tilde{\xi}(k, s) = \frac{\tilde{v}(k, s(1+u)) - \tilde{v}(k, s(1+d))}{(u-d)s}. \tag{3.7}$$

In particular, it is here enough to do all the computations in the *simplified, recombining tree* because neither $\tilde{V}^{\tilde{H}}$ nor ϑ have any path-dependence, but only depend on current values of $\widetilde{S}^1$. So instead of 2^T terminal nodes for all the trajectories ω, we need here only $T+1$ terminal nodes, for all the possible values of $\widetilde{S}_T^1$. The corresponding tree is therefore also massively smaller, and so are computation times and storage requirements.
[It is a very good [→ *exercise*] to either derive the above relations for the path-independent case directly or deduce them from the preceding general results. In both cases, one uses a *backward induction* argument.]

4 Basics about Brownian motion

The continuous-time analogue (and limit, in an appropriate sense) of the Cox–Ross–Rubinstein binomial model is the Black–Scholes model of geometric Brownian motion. To be able to study this later in Chapter 7, we collect in this chapter some basic facts and results about *Brownian motion*. This is the stochastic process driving the Black–Scholes model; but it is of fundamental importance in many other areas as well. Very loosely, one can think of Brownian motion as a dynamic version of the normal distribution, with a comparable status as an object of central significance.

Throughout this chapter, we work on a probability space $(\Omega, \mathcal{F}, P)$ which is tacitly assumed to be big and rich enough for our purposes. In particular, Ω cannot be finite or countable. We also work with a *filtration* $\mathbb{F} = (\mathcal{F}_t)$ in continuous time; this is like in discrete time a family of σ-fields $\mathcal{F}_t \subseteq \mathcal{F}$ with $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$. The time parameter runs either through $t \in [0, T]$ with a fixed time horizon $T \in (0, \infty)$ or through $t \in [0, \infty)$. In the latter case, we define

$$\mathcal{F}_\infty := \bigvee_{t \geq 0} \mathcal{F}_t := \sigma\Big(\bigcup_{t \geq 0} \mathcal{F}_t \Big).$$

For technical reasons, we should also assume (or make sure, if we construct the filtration in some way) that $\mathbb{F}$ satisfies the so-called *usual conditions* of being right-continuous and P-complete, but we do not dwell on this technical mathematical issue in more detail.

4.1 Definition and first properties

Definition. A *Brownian motion with respect to P and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$* is a (real-valued) stochastic process $W = (W_t)_{t \geq 0}$ which is adapted to $\mathbb{F}$, starts at 0 (i.e. $W_0 = 0$ P-a.s.) and satisfies the following properties:

(BM1) For $s \leq t$, the *increment* $W_t - W_s$ is independent (under P) of $\mathcal{F}_s$ with (under P) a normal distribution $\mathcal{N}(0, t - s)$.

(BM2) W has *continuous trajectories*, meaning that for P-almost all $\omega \in \Omega$, the function $t \mapsto W_t(\omega)$ on $[0, \infty)$ is continuous.

Remarks. **1)** One can prove that Brownian motion exists, but this is a nontrivial mathematical result. See the course on "Brownian Motion and Stochastic Calculus" (in short BMSC) for more details.

2) The letter W is used in honour of Norbert Wiener who gave the first rigorous proof of the existence of Brownian motion in 1923. It is historically interesting to note, however, that Brownian motion was already introduced and used considerably earlier in both finance and physics — by Louis Bachelier in his PhD thesis in 1900 for finance and by Albert Einstein in 1905 for physics.

3) Brownian motion in $\mathbb{R}^m$ is simply an adapted $\mathbb{R}^m$-valued stochastic process null at 0 with (BM2) and such that (BM1) holds with $\mathcal{N}(0, t-s)$ replaced by $\mathcal{N}(0, (t-s)I_{m\times m})$, where $I_{m\times m}$ denotes the $m \times m$ identity matrix. This is equivalent to saying that the m components are all real-valued Brownian motions and independent (as processes). $\diamond$

There is also a definition of Brownian motion (BM for short) without any filtration $\mathbb{F}$. This is a (real-valued) stochastic process $W = (W_t)_{t\geq 0}$ which starts at 0, satisfies (BM2) and instead of (BM1) the following property:

(BM1′) For any $n \in \mathbb{N}$ and any times $0 = t_0 < t_1 < \cdots < t_n < \infty$, the increments $W_{t_i} - W_{t_{i-1}}$, $i = 1, \ldots, n$, are independent (under P) and we have (under P) that $W_{t_i} - W_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})$, or $\sim \mathcal{N}(0, (t_i - t_{i-1})I_{m\times m})$ if W is $\mathbb{R}^m$-valued.

Instead of (BM1′), one also says (in words) that W has *independent stationary increments* with a (specific) normal distribution.

The two definitions of BM are equivalent if one chooses as $\mathbb{F}$ the filtration $\mathbb{F}^W$ generated by W (and made right-continuous and P-complete). This (like many other subsequent results and facts) needs a proof, which we do not give. More details can be found in the lecture notes on "Brownian Motion and Stochastic Calculus".

There are several *transformations* that produce a new Brownian motion from a given one, and this can in turn be used to prove results about BM. More precisely:

Proposition 1.1. *Suppose $W = (W_t)_{t\geq 0}$ is a BM. Then:*

1) $W^1 := -W$ *is a BM.*

2) $W_t^2 := W_{t+T} - W_T$, $t \geq 0$, *is a BM for any* $T \in (0, \infty)$ (restarting at a fixed time T).

3) $W_t^3 := cW_{\frac{t}{c^2}}$, $t \geq 0$, *is a BM for any* $c \in \mathbb{R}$, $c \neq 0$ (rescaling in space and time).

4) $W_t^4 := W_{T-t} - W_T$, $0 \leq t \leq T$, *is a BM on* $[0, T]$ *for any* $T \in (0, \infty)$ (time-reversal).

5) *The process* W_t^5, $t \geq 0$, *defined by*

$$W_t^5 := \begin{cases} tW_{\frac{1}{t}} & \text{for } t > 0 \\ 0 & \text{for } t = 0 \end{cases}$$

is a BM (inversion of small and large times).

(Note that we always use here the definition of BM without an exogenous filtration.)

While parts 1)–4) of Proposition 1.1 are easy to prove, part 5) is a bit more tricky. However, it is also very useful because it relates the asymptotic behaviour of BM as $t \to \infty$ to the behaviour of BM close to time 0, and vice versa.

The next result gives some information about how *trajectories of BM* behave.

Proposition 1.2. *Suppose* $W = (W_t)_{t \geq 0}$ *is a BM. Then:*

1) Law of large numbers: $\lim\limits_{t \to \infty} \dfrac{W_t}{t} = 0$ *P-a.s., i.e. BM grows more slowly than linearly as* $t \to \infty$.

2) (Global) Law of the iterated logarithm (LIL): *With* $\psi_{\text{glob}}(t) := \sqrt{2t \log(\log t)}$, *we have*

$$\left.\begin{matrix} \limsup\limits_{t \to \infty} \\ \liminf\limits_{t \to \infty} \end{matrix}\right\} \frac{W_t}{\psi_{\text{glob}}(t)} = \begin{cases} +1 \\ -1 \end{cases} \qquad P\text{-a.s.},$$

i.e., for P-almost all ω, *the function* $t \mapsto W_t(\omega)$ *for* $t \to \infty$ *oscillates precisely between* $t \mapsto \pm\psi_{\text{glob}}(t)$.

3) (Local) Law of the iterated logarithm (LIL): *With* $\psi_{\mathrm{loc}}(h) := \sqrt{2h\log(\log\frac{1}{h})}$, *we have for every* $t \geq 0$

$$\left.\begin{array}{l}\limsup\limits_{h\searrow 0}\\ \liminf\limits_{h\searrow 0}\end{array}\right\} \frac{W_{t+h}-W_t}{\psi_{\mathrm{loc}}(h)} = \left\{\begin{array}{l}+1\\ -1\end{array}\right. \qquad P\text{-a.s.},$$

i.e., for P-almost all ω, *to the right of* t, *the trajectory* $u \mapsto W_u(\omega)$ *around the level* $W_t(\omega)$ *oscillates precisely between* $h \mapsto \pm\psi_{\mathrm{loc}}(h)$.

One immediate consequence of 2) and 3) is that BM crosses the level 0 (or, with a bit more effort for the proof, any level a) infinitely many times — and once it is at that level, it even manages to achieve infinitely many crossings in an arbitrarily short amount of time. This is already a first indication of the amazingly strong activity of BM.

We remark that part 1) of Proposition 1.2 is easily proved by using part 5) of Proposition 1.1. Moreover, part 2) follows directly from part 3) via part 5) of Proposition 1.1, and for proving part 3), it is enough to take $t = 0$, by part 2) of Proposition 1.1, and to prove the lim sup result, by part 1) of Proposition 1.1. But then the easy reductions stop and the proof becomes difficult.

The oscillation results in Proposition 1.2 already make it clear that the trajectories of BM behave rather wildly. Another result in that direction is

Proposition 1.3. *Suppose* $W = (W_t)_{t\geq 0}$ *is a BM. Then for P-almost all* $\omega \in \Omega$, *the function* $t \mapsto W_t(\omega)$ *from* $[0,\infty)$ *to* $\mathbb{R}$ *is continuous, but nowhere differentiable.*

The deeper reason behind the wild behaviour of Brownian trajectories, and the key to understanding stochastic calculus and Itô's formula for BM, is that Brownian trajectories are *continuous functions having* a *nonzero quadratic variation*. Heuristically, this can be seen as follows. By definition, Brownian motion increments $W_{t+h} - W_t$ have a normal

distribution $\mathcal{N}(0,h)$, which implies they are symmetric around 0 with variance h so that roughly, "$W_{t+h} - W_t \approx \pm\sqrt{h}$ with probability $\frac{1}{2}$ each". In very loose and purely formal terms, this means that infinitesimal increments "$\mathrm{d}W_t = W_t - W_{t-\mathrm{d}t}$" of BM have the property that

$$\text{“}(\mathrm{d}W_t)^2 = \mathrm{d}t\text{”}.$$

While this is very helpful for an *intuitive understanding*, we emphasise that it is *purely formal* and must not be used for rigorous mathematical arguments. A more precise description is as follows.

Call a *partition* of $[0,\infty)$ any set $\Pi \subseteq [0,\infty)$ of time points with $0 \in \Pi$ and such that $\Pi \cap [0,T]$ is finite for all $T \in [0,\infty)$. This implies that Π is at most countable and can be ordered increasingly as $\Pi = \{0 = t_0 < t_1 < \cdots < t_m < \cdots < \infty\}$. The *mesh size* of Π is then defined as $|\Pi| := \sup\{t_i - t_{i-1} : t_{i-1}, t_i \in \Pi\}$, i.e. the size of the biggest time-step in Π. For any partition Π of $[0,\infty)$, any function $g : [0,\infty) \to \mathbb{R}$ and any $p > 0$, we first define the *p-variation of g on $[0,T]$ along Π* as

$$V_T^p(g,\Pi) := \sum_{t_i \in \Pi} |g(t_i \wedge T) - g(t_{i-1} \wedge T)|^p.$$

One can then define the *p-variation of g on $[0,T]$* as

$$V_T^p(g) := \sup_{\Pi} V_T^p(g,\Pi),$$

where the supremum is taken over all partitions Π of $[0,\infty)$. For a sequence $(\Pi_n)_{n\in\mathbb{N}}$ of partitions of $[0,\infty)$ with $\lim_{n\to\infty} |\Pi_n| = 0$, one can also define the *p-variation of g on $[0,T]$ along $(\Pi_n)_{n\in\mathbb{N}}$* as

$$\lim_{n\to\infty} V_T^p(g,\Pi_n),$$

provided that the limit exists.

With the above notations, a function g is of *finite variation* or has *finite 1-variation* if $V_T^1(g) < \infty$ for every $T \in (0,\infty)$. The interpretation is that the graph of g has finite length on any time interval. More precisely, if we define the *arc length* of (the graph of) g on the interval $[0,T]$ as

$$\sup_{\Pi} \sum_{t_i \in \Pi} \sqrt{(t_i \wedge T - t_{i-1} \wedge T)^2 + \big(g(t_i \wedge T) - g(t_{i-1} \wedge T)\big)^2},$$

with the supremum again taken over all partitions Π of $[0,\infty)$, then g has finite variation on $[0,T]$ if and only if it has finite arc length on $[0,T]$. This can be checked by using the inequality $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for $a, b \ge 0$.

Any monotonic (increasing or decreasing) function is clearly of finite variation, because the absolute values above disappear and we get a telescoping sum. Moreover, one can show that *any function of finite variation can be written as the difference of two increasing functions (and vice versa).*

Now let us return to Brownian motion, taking $p = 2$ and as g one trajectory $W_\bullet(\omega)$. Then

$$Q_T^\Pi := \sum_{t_i \in \Pi} (W_{t_i \wedge T} - W_{t_{i-1} \wedge T})^2 = V_T^2(W_\bullet, \Pi)$$

is the *sum* up to time T *of* the *squared increments* of BM along Π. With the above formal intuition "$(\mathrm{d}W_t)^2 = \mathrm{d}t$", we then expect, at least for $|\Pi|$ very small so that time points are close together, that $(W_{t_i \wedge T} - W_{t_{i-1} \wedge T})^2 \approx t_i \wedge T - t_{i-1} \wedge T$ and hence

$$Q_T^\Pi \approx \sum_{t_i \in \Pi} (t_i \wedge T - t_{i-1} \wedge T) = T \qquad \text{for } |\Pi| \text{ small.}$$

Even if the above reasoning is only heuristic, the result surprisingly is correct:

Theorem 1.4. *Suppose $W = (W_t)_{t \ge 0}$ is a BM. For any sequence $(\Pi_n)_{n \in \mathbb{N}}$ of partitions of $[0,\infty)$ which is refining (i.e. $\Pi_n \subseteq \Pi_{n+1}$ for all n) and satisfies $\lim_{n\to\infty} |\Pi_n| = 0$, we have*

$$P\Big[\lim_{n\to\infty} Q_t^{\Pi_n} = t \text{ for every } t \ge 0\Big] = 1.$$

We express this by saying that along $(\Pi_n)_{n \in \mathbb{N}}$, the Brownian motion W has (with probability 1) quadratic variation *t on $[0,t]$ for every $t \ge 0$, and we write $\langle W \rangle_t = t$. (We sometimes also say, with a certain abuse of terminology, that P-almost all trajectories $W_\bullet(\omega) : [0,\infty) \to \mathbb{R}$ of BM have quadratic variation t on $[0,t]$, for each $t \ge 0$.)*

Remark 1.5. 1) It is a very nice and useful [→ *exercise*] in analysis to prove that *every continuous function f which has nonzero quadratic variation along a sequence (Π_n) as*

above must have infinite variation, i.e. unbounded oscillations. (This will come up again later in Section 6.1.) More generally, if $\lim_{n\to\infty} V_T^q(f,\Pi_n) > 0$ for some $q > 0$, then $\lim_{n\to\infty} V_T^p(f,\Pi_n) = +\infty$ for any p with $0 < p < q$, and if $\lim_{n\to\infty} V_T^p(f,\Pi_n) < \infty$ for some $p > 0$, then $\lim_{n\to\infty} V_T^q(f,\Pi_n) = 0$ for all $q > p$. We also recall that a classical result due to Lebesgue says that any function of finite variation is almost everywhere differentiable. So Proposition 1.3 implies that Brownian trajectories must have infinite variation, and Theorem 1.4 makes this even quantitative.

2) Caution: The comment in 1) is only true for *continuous* functions. With RCLL functions, this breaks down in general.

3) It is important in Theorem 1.4 that the partitions Π_n do not depend on the trajectory $W_\bullet(\omega)$, but are fixed a priori. One can show for P-almost all trajectories $W_\bullet(\omega)$, the (true) quadratic variation of $W_\bullet(\omega)$ is $+\infty$.

4) There is an extension of Theorem 1.4 to general local martingales M instead of Brownian motion W. But then the limit, called $[M]_t$, of the sequence $(Q_t^{\Pi_n}(M))_{n\in\mathbb{N}}$ is not t, but some ($\mathcal{F}_t$-measurable) random variable, and the convergence holds not P-almost surely, but only in probability. (Alternatively, one can obtain P-a.s. convergence along a sequence of partitions, but then this cannot be chosen, but is only shown to exist.) Moreover, $t \mapsto [M]_t(\omega)$ is then always increasing (for P-almost all ω), but only continuous if M itself has continuous trajectories. Finally, as for Brownian motion, the limit does not depend on the sequence $(\Pi_n)_{n\in\mathbb{N}}$ of partitions. $\diamond$

4.2 Martingale properties and results

There are many martingales which are naturally associated to Brownian motion, and this is useful in many different contexts. We present here just a small sample that will be used or useful later.

As in discrete time, a *martingale* with respect to P and $\mathbb{F}$ is a (real-valued) stochastic process $M = (M_t)$ such that M is adapted to $\mathbb{F}$, M is P-integrable in the sense that each M_t is in $L^1(P)$, and the *martingale property* holds: for $s \leq t$, we have

$$E[M_t \,|\, \mathcal{F}_s] = M_s \qquad P\text{-a.s.} \tag{2.1}$$

If we have in (2.1) the inequality "$\leq$" instead of "$=$", then M is a *supermartingale*; if we have "$\geq$", then M is a *submartingale*. Of course, $\mathbb{F} = (\mathcal{F}_t)$ and $M = (M_t)$ should have the same time index set.

Remark 2.1. Because our filtration satisfies the usual conditions, a general result from the theory of stochastic processes says that any martingale has a version with nice (*RCLL*, i.e. *right-continuous with left limits*, to be precise) *trajectories*. We can and do therefore always assume that our martingales have nice trajectories in that sense, and this is important for some of the subsequent results. We shall point this out more explicitly when it is used. ◇

Again exactly like in discrete time, a *stopping time* with respect to $\mathbb{F}$ is a mapping $\tau : \Omega \to [0, \infty]$ such that $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$. One of the standard examples is the first time that some adapted right-continuous process X (e.g. Brownian motion W) hits an open set B (e.g. (a, ∞)), i.e.

$$\tau := \inf\{t \geq 0 : X_t \in B\} \qquad \big(= \inf\{t \geq 0 : W_t > a\}, \text{ for } X = W \text{ and } B = (a, \infty)\big).$$

We remark that checking the stopping time property above uses that the filtration is right-continuous; and we mention that τ above is still a stopping time if B is allowed to be a Borel set, but the proof of this apparently minor extension is surprisingly difficult.

One of the most useful properties of martingales is that the martingale property (2.1) and its consequences very often extend to the case where the fixed times $s \leq t$ are replaced

by stopping times $\sigma \leq \tau$. "Very often" means under additional conditions, as we shall see presently. To make sense of (2.1) for σ and τ, we also first need to define, for a stopping time σ, the σ-field of *events observable up to time* σ as

$$\mathcal{F}_\sigma := \big\{A \in \mathcal{F} : A \cap \{\sigma \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\big\}.$$

(One must and can check that $\mathcal{F}_\sigma$ is a σ-field, and that one has $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$ for $\sigma \leq \tau$.) We also need to define M_τ, the value of M at the stopping time τ, by

$$(M_\tau)(\omega) := M_{\tau(\omega)}(\omega).$$

Note that this implicitly assumes that we have a random variable M_∞, because τ can take the value $+\infty$. One can then also prove that if τ is a stopping time and M is an adapted process with RC trajectories, then M_τ is $\mathcal{F}_\tau$-measurable (as one intuitively expects). Finally, we also recall the stopped process $M^\tau = (M_t^\tau)_{t\geq 0}$ which is defined by $M_t^\tau := M_{t\wedge\tau}$ for all $t \geq 0$. Again, if M is adapted with RC trajectories and τ is a stopping time, then also M^τ is adapted and has RC trajectories.

After the above preliminaries, we now have

Theorem 2.2 (**Stopping theorem**). *Suppose that $M = (M_t)_{t\geq 0}$ is a $(P, \mathbb{F})$-martingale with RC trajectories, and σ, τ are $\mathbb{F}$-stopping times with $\sigma \leq \tau$. If either τ is bounded by some $T \in (0, \infty)$ or M is uniformly integrable, then M_τ, M_σ are both in $L^1(P)$ and*

$$E[M_\tau \,|\, \mathcal{F}_\sigma] = M_\sigma \qquad P\text{-a.s.} \tag{2.2}$$

Two frequent *applications* of Theorem 2.2 are the following:

1) For any RC martingale M and any stopping time τ, we have $E[M_{\tau\wedge t} \,|\, \mathcal{F}_s] = M_{\tau\wedge s}$ for $s \leq t$, i.e., the stopped process $M^\tau = (M_t^\tau)_{t\geq 0} = (M_{t\wedge\tau})_{t\geq 0}$ is again a martingale (because we have $E[M_t^\tau \,|\, \mathcal{F}_s] = M_s^\tau$).

[Because not necessarily $s \leq \tau \wedge t$, this needs a little bit of extra work.]

2) If M is an RC martingale and τ is any stopping time, then we always have for any $t \geq 0$ that $E[M_{\tau\wedge t}] = E[M_0]$. If either τ is bounded or M is uniformly integrable, then we also obtain $E[M_\tau] = E[M_0]$.

For future use, let us also recall the notion of a local martingale null at 0, now in continuous time. An adapted process $X = (X_t)_{t\geq 0}$ null at 0 (i.e. with $X_0 = 0$) is called a *local martingale null at* 0 (with respect to P and $\mathbb{F}$) if there exists a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ increasing to ∞ such that for each $n \in \mathbb{N}$, the stopped process $X^{\tau_n} = (X_{t\wedge\tau_n})_{t\geq 0}$ is a $(P, \mathbb{F})$-martingale. We then call $(\tau_n)_{n\in\mathbb{N}}$ a *localising sequence*. (If X is defined on $[0, T]$ for some $T \in (0, \infty)$, the requirement for a localising sequence is that (τ_n) increases to T *stationarily*, i.e. $\tau_n \nearrow T$ P-a.s. and $P[\tau_n < T] \to 0$ as $n \to \infty$.)

The next result presents a number of martingales directly related to Brownian motion.

Proposition 2.3. *Suppose $W = (W_t)_{t\geq 0}$ is a $(P, \mathbb{F})$-Brownian motion. Then the following processes are all $(P, \mathbb{F})$-martingales:*

1) *W itself.*

2) *$W_t^2 - t$, $t \geq 0$.*

3) *$e^{\alpha W_t - \frac{1}{2}\alpha^2 t}$, $t \geq 0$, for any $\alpha \in \mathbb{R}$.*

Proof. We do this argument (in part) because it illustrates how to work with the properties of BM. For each of the above processes, adaptedness is obvious, and integrability is also clear because each W_t has a normal distribution and hence all exponential moments. Finally, as $W_t - W_s$ is independent of $\mathcal{F}_s$ and $\sim \mathcal{N}(0, t - s)$, we get 1) from

$$E[W_t - W_s \,|\, \mathcal{F}_s] = E[W_t - W_s] = 0.$$

Using this with $W_t^2 - W_s^2 = (W_t - W_s)^2 + 2W_s(W_t - W_s)$ and $\mathcal{F}_s$-measurability of W_s then gives

$$\begin{aligned} E[W_t^2 - W_s^2 \,|\, \mathcal{F}_s] &= E[(W_t - W_s)^2 \,|\, \mathcal{F}_s] \\ &= E[(W_t - W_s)^2] = \mathrm{Var}[W_t - W_s] = t - s, \end{aligned}$$

hence 2). Finally, setting $M_t := e^{\alpha W_t - \frac{1}{2}\alpha^2 t}$ yields

$$\begin{aligned} E\left[\frac{M_t}{M_s}\,\middle|\,\mathcal{F}_s\right] &= E\big[e^{\alpha(W_t - W_s) - \frac{1}{2}\alpha^2(t-s)}\,\big|\,\mathcal{F}_s\big] \\ &= e^{-\frac{1}{2}\alpha^2(t-s)} E\big[e^{\alpha(W_t - W_s)}\big] = 1 \end{aligned}$$

because $E[e^Z] = e^{\mu+\frac{1}{2}\sigma^2}$ for $Z \sim \mathcal{N}(\mu, \sigma^2)$. So we have 3) as well. **q.e.d.**

Example. To illustrate that the conditions in Theorem 2.2 are really needed, consider a Brownian motion W and the stopping time

$$\tau := \inf\{t \geq 0 : W_t > 1\}.$$

Due to the law of the iterated logarithm in part 2) of Proposition 1.2, we have $\tau < \infty$ P-a.s., and because W has continuous trajectories, we get $W_\tau = 1$ P-a.s. For $\sigma = 0$, if (2.2) were valid for W and τ, σ, we should get by taking expectations that

$$1 = E[W_\tau] = E[W_\sigma] = E[W_0] = 0,$$

which is clearly false. So τ *cannot be bounded* by a constant (in fact, one can even show that $E[\tau] = +\infty$), and W is a martingale, but *not uniformly integrable*. Finally, we also see that (2.2) is not true in general (i.e. without assumptions on M and/or τ).

One useful application of the above martingale results is the computation of the Laplace transforms of certain *hitting times*. More precisely, let $W = (W_t)_{t\geq 0}$ be a Brownian motion and define for $a > 0$, $b > 0$ the stopping times

$$\tau_a := \inf\{t \geq 0 : W_t > a\},$$
$$\sigma_{a,b} := \inf\{t \geq 0 : W_t > a + bt\}.$$

Note that $\tau_a < \infty$ P-a.s. by the (global) law of the iterated logarithm in part 2) of Proposition 1.2, whereas $\sigma_{a,b}$ can be $+\infty$ with positive probability (see below).

Proposition 2.4. *Let W be a BM and $a > 0$, $b > 0$. Then for any $\lambda > 0$, we have*

$$E[e^{-\lambda\tau_a}] = e^{-a\sqrt{2\lambda}} \tag{2.3}$$

and

$$E[e^{-\lambda\sigma_{a,b}}] = E\big[e^{-\lambda\sigma_{a,b}} I_{\{\sigma_{a,b}<\infty\}}\big] = e^{-a(b+\sqrt{b^2+2\lambda})}. \tag{2.4}$$

Proof. We give this argument because it illustrates how to use the preceding martingale results. First of all, take $\alpha > 0$ and define $M_t := \exp(\alpha W_t - \frac{1}{2}\alpha^2 t)$, $t \geq 0$. Then M is a martingale by part 3) of Proposition 2.3, and hence so is the stopped process M^τ by (the first comment after) Theorem 2.2, for $\tau \in \{\tau_a, \sigma_{a,b}\}$. This implies (as in the second comment after Theorem 2.2) that

$$1 = E[M_0] = E[M_{\tau\wedge t}] = E\big[e^{\alpha W_{\tau\wedge t} - \frac{1}{2}\alpha^2(\tau\wedge t)}\big]$$

for all t, and we now want to let $t \to \infty$.

For $\tau = \tau_a$, we have $W_{\tau_a\wedge t} \leq a$ and therefore $M_{\tau_a\wedge t}$ is bounded uniformly in t and ω (by $e^{\alpha a}$); so dominated convergence yields for $t \to \infty$ that

$$\begin{aligned} 1 &= \lim_{t\to\infty} E\big[e^{\alpha W_{\tau_a\wedge t} - \frac{1}{2}\alpha^2(\tau_a\wedge t)}\big] \\ &= E\Big[\lim_{t\to\infty} e^{\alpha W_{\tau_a\wedge t} - \frac{1}{2}\alpha^2(\tau_a\wedge t)}\Big] \\ &= E\big[e^{\alpha W_{\tau_a} - \frac{1}{2}\alpha^2\tau_a}\big] \\ &= e^{\alpha a} E\big[e^{-\frac{1}{2}\alpha^2\tau_a}\big] \end{aligned}$$

because $\tau_a < \infty$ P-a.s., and so $\alpha := \sqrt{2\lambda}$ gives (2.3).

For $\tau = \sigma_{a,b}$, we have $W_{\sigma_{a,b}\wedge t} \leq a + b(\sigma_{a,b} \wedge t)$ so that

$$M_{\sigma_{a,b}\wedge t} \leq \exp\left(\alpha a + \Big(\alpha b - \frac{1}{2}\alpha^2\Big)(\sigma_{a,b} \wedge t)\right)$$

is bounded uniformly in t and ω (by $e^{\alpha a}$) for $\alpha b < \frac{1}{2}\alpha^2$, i.e. for $\alpha > 2b$. Moreover, $\alpha b - \frac{1}{2}\alpha^2 < 0$ implies that on the set $\{\sigma_{a,b} = +\infty\}$, we have both $M_{\sigma_{a,b}\wedge t} \to 0$ as $t \to \infty$ and $e^{(\alpha b - \frac{1}{2}\alpha^2)\sigma_{a,b}} = 0$. Therefore we get in the same way as above via dominated convergence that

$$1 = e^{\alpha a} E\big[e^{(\alpha b - \frac{1}{2}\alpha^2)\sigma_{a,b}} I_{\{\sigma_{a,b}<\infty\}}\big] = e^{\alpha a} E\big[e^{(\alpha b - \frac{1}{2}\alpha^2)\sigma_{a,b}}\big].$$

Then (2.4) follows for $\alpha := b + \sqrt{b^2 + 2\lambda}$, which gives by a straightforward computation that $\alpha b - \frac{1}{2}\alpha^2 = \alpha(b - \frac{1}{2}\alpha) = -\lambda < 0$. **q.e.d.**

Remark. If we let $\lambda \searrow 0$ in (2.4), we obtain $P[\sigma_{a,b} < \infty] = e^{-2ab}$ so that indeed $P[\sigma_{a,b} = +\infty] = 1 - e^{-2ab} > 0$. ◇

For a general random variable $U \geq 0$, the function $\lambda \mapsto E[e^{-\lambda U}]$ for $\lambda > 0$ is called the *Laplace transform* of U. Its general importance in probability theory is that it uniquely determines the distribution of U.

In mathematical finance, both τ_a and $\sigma_{a,b}$ come up in connection with a number of so-called *exotic options*. In particular, they are important for *barrier options* whose payoff depends on whether or not a (upper or lower) level has been reached by a given time. When computing prices of such options in the Black–Scholes model, one almost immediately encounters the Laplace transforms from Proposition 2.4. For more details, see for instance Dana/Jeanblanc [3, Chapter 9].

4.3 Markovian properties

We have already seen in part 2) of Proposition 1.1 that for any fixed time $T \in (0, \infty)$, the process

$$W_{t+T} - W_T,\ t \geq 0, \qquad \text{is again a BM} \tag{3.1}$$

if $(W_t)_{t\geq 0}$ is a Brownian motion. This means that if we restart a BM from level 0 at some fixed time, it behaves exactly as if it had only just started. Moreover, one can show that the independence of increments of BM implies that

$$W_{t+T} - W_T,\ t \geq 0, \qquad \text{is independent of } \mathcal{F}_T^0, \tag{3.2}$$

where $\mathcal{F}_T^0 = \sigma(W_s, s \leq T)$ is the σ-field generated by BM up to time T. Intuitively, this means that BM at any fixed time T simply forgets its past up to time T (with the only possible exception that it remembers its current position W_T at time T), and starts afresh.

One consequence of (3.1) and (3.2) is the following. Suppose that at some fixed time T, we are interested in the behaviour of W after time T and try to predict this on the basis of the past of W up to time T, where "prediction" is done in the sense of a conditional expectation. Then we may as well forget about the past and look only at the current value W_T at time T. A bit more precisely, we can express this, for functions $g \geq 0$ applied to the part of BM after time T, as

$$E[g(W_u, u \geq T) \,|\, \sigma(W_s, s \leq T)] = E[g(W_u, u \geq T) \,|\, \sigma(W_T)]. \tag{3.3}$$

This is called the *Markov property* of BM, and it is already very useful in many situations.

Exactly as with martingales, we suspect that it might be interesting and helpful if one could in (3.3) replace the fixed time $T \in (0, \infty)$ by a stopping time τ. Note, however, that quite apart from the difficulties of writing down an analogue of (3.3) for a random time $\tau(\omega)$, it is even not clear whether this should then be true, because after all, τ itself can explicitly depend on the past behaviour of BM. Nevertheless, it turns out that such a result is true; one says that BM even has the *strong Markov property*.

Because a precise analogue of (3.3) for a stopping time becomes a bit technical, we formulate things a bit differently. If we denote almost as above by $\mathbb{F}^W$ the filtration generated by W (and made right-continuous, to be accurate), and if τ is a stopping time with respect to $\mathbb{F}^W$ and such that $\tau < \infty$ P-a.s., then

$$W_{t+\tau} - W_\tau,\ t \geq 0, \qquad \text{is again a BM and independent of } \mathcal{F}_\tau^W.$$

Of course, this includes (3.1) and (3.2) as special cases, and one can easily believe that it is even more useful than (3.3). However, the proof is too difficult to be given here.

5 Stochastic integration

From the discrete-time theory developed in Chapters 1–3, we know that the trading gains or losses from a self-financing strategy $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ are described by the *stochastic integral*

$$G(\vartheta) = \vartheta \bullet S = \int \vartheta \, \mathrm{d}S = \sum_j \vartheta_j^{\mathrm{tr}} \Delta S_j = \sum_j \vartheta_j^{\mathrm{tr}} (S_j - S_{j-1}).$$

To be able to develop an analogous theory in continuous time, we therefore need to understand how to define, and how to work with, a continuous-time stochastic integral process $\int \vartheta \, \mathrm{d}S$. From classical integration theory, the obvious idea is to start with approximating Riemann sums of the form $\sum \vartheta_{\tilde{t}_i}^{\mathrm{tr}} (S_{t_{i+1}} - S_{t_i})$, with $\tilde{t}_i$ lying between t_i and t_{i+1}, and then pass to the limit in a suitable sense. The simplest idea for that would be to fix ω, look at the trajectories $t \mapsto S_t(\omega)$ and $t \mapsto \vartheta_t(\omega)$ and take limits of

$$\sum \vartheta_{\tilde{t}_i}(\omega) \big(S_{t_{i+1}}(\omega) - S_{t_i}(\omega) \big)$$

like in courses on measure and integration theory. But unfortunately, this works well (i.e., for many integrands ϑ) only if the function $t \mapsto S_t(\omega)$ is of *finite variation* — and this would immediately exclude as integrator a process like Brownian motion which does not have this property. So one must use a different approach, and this will be explained in this chapter. For an amplification (and proof) of the above point that "naive stochastic integration is impossible", we refer to Protter [13, Section I.8]; the idea originally goes back to C. Stricker.

Remarks. 1) To avoid misunderstandings later, let us clarify that defining stochastic integrals as above in a pathwise manner (i.e. ω by ω) may well be possible if the integrator S and the integrand ϑ match up nicely enough, even if $t \mapsto S_t(\omega)$ is not of finite variation. We shall see this later in the context of Itô's formula, where ϑ has the form $\vartheta_t = g(S_{t-})$ for some C^1-function g. But if we want to fix S and allow many ϑ without imposing undue restrictions, an ω-wise approach leads to problems.

2) In classical integration theory, it does not matter in which point $\tilde{t}_i \in [t_i, t_{i+1}]$ one evaluates the integrand when defining the Riemann approximation. For stochastic

integrals, this is different — choosing the left endpoint $\tilde{t}_i = t_i$ leads to the *Itô integral*, the right endpoint $\tilde{t}_i = t_{i+1}$ yields the *forward integral*, and the midpoint choice $\tilde{t}_i = \frac{1}{2}(t_i + t_{i+1})$ produces the *Stratonovich integral*. However, for applications in finance, it is clear that one must choose $\tilde{t}_i = t_i$ (and hence the Itô integral) because the strategy must be decided before the price move. ◇

5.1 The basic construction

Our goal in this section is to construct a stochastic integral process $H \bullet M = \int H \, \mathrm{d}M$ when M is a (real-valued) local martingale null at 0 and H is a (real-valued) predictable process with a suitable integrability property (relative to M). In Section 5.3 below, we also explain how to extend this from local martingales to semimartingales; but the key step and the main work happen in the martingale case.

Remark. For simplicity, we take both M and H to be real-valued. It is reasonably straightforward, although somewhat technical, to extend the theory from this section to M and H that are both $\mathbb{R}^d$-valued, and we comment on the necessary changes a bit later. We then also point out some of the pitfalls that one has to avoid in that context. ◇

Throughout this chapter, we work on a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfying the usual conditions of right-continuity and P-completeness. If needed, we define $\mathcal{F}_\infty := \bigvee_{t \geq 0} \mathcal{F}_t$. We also fix a (real-valued) *local martingale* $M = (M_t)_{t \geq 0}$ null at 0 (as defined before Proposition 4.2.3) and having RCLL (right-continuous with left limits) trajectories. (The latter property, as pointed out earlier in Remark 4.2.1, is not a restriction; we can always find an RCLL version of M thanks to the usual conditions on $\mathbb{F}$.) Because we want to define stochastic integrals $\int_a^b H \, \mathrm{d}M$ and these are always over half-open intervals of the form $(a, b]$ with $0 \leq a < b \leq \infty$, the value of M at 0 is irrelevant and it is enough to look at processes $H = (H_t)$ defined for $t > 0$. This will simplify some definitions. For any process $Y = (Y_t)_{t \geq 0}$ with RCLL trajectories, we denote by $\Delta Y_t := Y_t - Y_{t-} := Y_t - \lim_{s \to t, s < t} Y_s$ the jump of Y at time $t > 0$.

The *simplest example* to be kept in mind is when $M = W$ is a Brownian motion. From Proposition 4.2.3, we know that both W itself and $(W_t^2 - t)_{t\geq 0}$ are then martingales, and by Theorem 4.1.4, the quantity t we subtract from W_t^2 is the quadratic variation of W, which can be obtained as a pathwise limit of sums of squared increments of W. As already mentioned in Remark 4.1.5, a similar result is true for a general local martingale M, and this is the key for constructing stochastic integrals.

Theorem 1.1. *For any local martingale $M = (M_t)_{t\geq 0}$ null at 0, there exists a unique adapted increasing RCLL process $[M] = ([M]_t)_{t\geq 0}$ null at 0 with $\Delta[M] = (\Delta M)^2$ and having the property that $M^2 - [M]$ is also a local martingale. This process $[M]$ can be obtained as the quadratic variation of M in the following sense: There exists a sequence $(\Pi_n)_{n\in\mathbb{N}}$ of partitions of $[0,\infty)$ with $|\Pi_n| \to 0$ as $n \to \infty$ such that*

$$P\left[[M]_t(\omega) = \lim_{n\to\infty} \sum_{t_i\in\Pi_n} \big(M_{t_i\wedge t}(\omega) - M_{t_{i-1}\wedge t}(\omega)\big)^2 \text{ for all } t \geq 0\right] = 1.$$

We call $[M]$ the optional quadratic variation *or* square bracket *process of M.*

If M satisfies $\sup_{0\leq s\leq T} |M_s| \in L^2$ for some $T > 0$ (and hence is in particular a square-integrable martingale on $[0,T]$), then $[M]$ is integrable on $[0,T]$ (i.e. $[M]_T \in L^1$) and $M^2 - [M]$ is a martingale on $[0,T]$.

Proof. See Protter [13, Section II.6] or Dellacherie/Meyer [5, Theorem VII.42] or Jacod/Shiryaev [11, Section I.4c].

Remarks. 1) Recall from Theorem 1.4 that Brownian motion W has $[W]_t = t$.

2) Note that $[M]$ has paths of finite variation. So one can easily define integrals $\int \ldots \mathrm{d}[M]$ in a pathwise manner as usual Lebesgue–Stieltjes integrals. This does not need any new theory.

3) The sequence $(\Pi_n)_{n\in\mathbb{N}}$ of partitions in Theorem 1.4 of course depends on M. $\diamond$

For two local martingales M, N null at 0, we define the (optional) *covariation process*

$[M, N]$ by polarisation, i.e.

$$[M, N] := \frac{1}{4}([M + N] - [M - N]).$$

From the characterisation of $[M]$ in Theorem 1.1, it follows easily that the operation $[\,\cdot\,,\,\cdot\,]$ is bilinear, and also that $[M, N]$ is the unique adapted RCLL process B null at 0, of finite variation with $\Delta B = \Delta M \Delta N$ and such that the difference $MN - B$ is again a local martingale.

Remark 1.2. 1) If M is a square-integrable martingale, then $[M]$ is integrable and therefore, by the general theory of stochastic processes, admits a so-called *(predictable) compensator* or *dual predictable projection*: There exists a unique increasing predictable integrable process $\langle M \rangle = (\langle M \rangle_t)_{t \geq 0}$ null at 0 such that $[M] - \langle M \rangle$, and therefore also $M^2 - \langle M \rangle = M^2 - [M] + [M] - \langle M \rangle$, is a martingale. The process $\langle M \rangle$ is called the *sharp bracket* (or sometimes the *predictable variance*) process of M. See Dellacherie/Meyer [5, Theorem VI.65 and Definition VI.77] or Jacod/Shiryaev [11, Theorem I.3.17]. Note that we still need to define what "predictable" means in continuous time.

2) Once we know what localisation means (see the end of this section for more details), we can easily extend the results in 1). It is enough if M is a locally square-integrable local martingale; then $\langle M \rangle$ is also locally integrable, and then both $[M] - \langle M \rangle$ and $M^2 - \langle M \rangle$ are local martingales.

3) We already point out here that any adapted process which is continuous is automatically locally bounded (see later for the definition) and therefore also locally square-integrable. Again, we refer to the end of this section for more details.

4) If M is *continuous*, then so is $[M]$, because $\Delta[M] = (\Delta M)^2 = 0$. This implies then also that $[M] = \langle M \rangle$. In particular, for a Brownian motion W, we have $[W]_t = \langle W \rangle_t = t$ for all $t \geq 0$.

5) If both M and N are locally square-integrable (e.g. if they are continuous), we also get $\langle M, N \rangle$ via polarisation.

6) If M is $\mathbb{R}^d$-valued, then $[M]$ becomes a $d \times d$-matrix-valued process with entries $[M]^{ik} = [M^i, M^k]$. To work with that, one needs to establish more properties. The same applies to $\langle M \rangle$, if it exists.

7) The key difference between $[M]$ and $\langle M \rangle$ is that $[M]$ exists for any local martingale M null at 0, whereas the existence of $\langle M \rangle$ requires some extra local integrability of M. $\diamond$

Definition. We denote by $b\mathcal{E}$ the set of all *bounded elementary processes* of the form

$$H = \sum_{i=0}^{n-1} h_i I_{(t_i, t_{i+1}]}$$

with $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \cdots < t_n < \infty$ and each h_i a bounded (real-valued) $\mathcal{F}_{t_i}$-measurable random variable. For any stochastic process $X = (X_t)_{t \geq 0}$, we then define the *stochastic integral* $\int H \, \mathrm{d}X$ of $H \in b\mathcal{E}$ by

$$\int_0^t H_s \, \mathrm{d}X_s := H \bullet X_t := \sum_{i=0}^{n-1} h_i (X_{t_{i+1} \wedge t} - X_{t_i \wedge t}) \qquad \text{for } t \geq 0.$$

Note that if X is RCLL, then so is $\int H \, \mathrm{d}X = H \bullet X$.

If X and H are both $\mathbb{R}^d$-valued, the integral is still real-valued, and we simply replace products by scalar products everywhere. But then Lemma 1.3 below looks more complicated.

Lemma 1.3. *Suppose that M is a square-integrable martingale (i.e., M is a martingale with $M_t \in L^2$ for all $t \geq 0$, or equivalently with $\sup_{0 \leq s \leq T} |M_s| \in L^2$ for all $T > 0$). For every $H \in b\mathcal{E}$, the stochastic integral process $H \bullet M = \int H \, \mathrm{d}M$ is then also a square-integrable martingale, and we have $[H \bullet M] = \int H^2 \, \mathrm{d}[M]$ and the* isometry property

$$\begin{aligned} E[(H \bullet M_\infty)^2] &= E\left[\left(\int_0^\infty H_s \, \mathrm{d}M_s \right)^2 \right] \\ &= E\left[\sum_{i=0}^{n-1} h_i^2 ([M]_{t_{i+1}} - [M]_{t_i}) \right] \\ &= E\left[\int_0^\infty H_s^2 \, \mathrm{d}[M]_s \right]. \end{aligned} \tag{1.1}$$

Note that the last $\mathrm{d}[M]$-integral can be defined ω by ω via classical measure and integration theory, because $t \mapsto [M]_t(\omega)$ is increasing and hence of finite variation. But of course it is here also just a finite sum, because H has such a simple form.

Proof of Lemma 1.3. Adaptedness of $H \bullet M$ is clear, and so is square-integrability because H is bounded and each $H \bullet M_t$ is just a finite sum. Moreover, H is identically 0 after t_n so that both infinite integrals actually end at t_n. We first argue the martingale property, for simplicity only for $s = t_i$, $t = t_{i+1}$. [$\to$ *Exercise:* Prove this in detail for arbitrary $s \le t$.] Indeed, by first using that h_i is $\mathcal{F}_{t_i}$-measurable and bounded, and then that M is a martingale, we get

$$E[H \bullet M_t - H \bullet M_s \,|\, \mathcal{F}_s] = E[h_i(M_{t_{i+1}} - M_{t_i}) \,|\, \mathcal{F}_{t_i}] = h_i E[M_{t_{i+1}} - M_{t_i} \,|\, \mathcal{F}_{t_i}] = 0.$$

Next, it is easy to check [$\to$ *exercise*] for any square-integrable martingale N that

$$E[N_t^2 - N_s^2 \,|\, \mathcal{F}_s] = E[(N_t - N_s)^2 \,|\, \mathcal{F}_s] \qquad \text{for } s \le t.$$

Applying this once to $H \bullet M$ and once to M yields

$$\begin{aligned} E[(H \bullet M_{t_{i+1}})^2 - (H \bullet M_{t_i})^2 \,|\, \mathcal{F}_{t_i}] &= E[(H \bullet M_{t_{i+1}} - H \bullet M_{t_i})^2 \,|\, \mathcal{F}_{t_i}] \\ &= E[h_i^2 (M_{t_{i+1}} - M_{t_i})^2 \,|\, \mathcal{F}_{t_i}] \\ &= h_i^2 E[M_{t_{i+1}}^2 - M_{t_i}^2 \,|\, \mathcal{F}_{t_i}] \\ &= h_i^2 E\big[[M]_{t_{i+1}} - [M]_{t_i} \,\big|\, \mathcal{F}_{t_i}\big] \\ &= E\big[h_i^2([M]_{t_{i+1}} - [M]_{t_i}) \,\big|\, \mathcal{F}_{t_i}\big] \\ &= E\big[H^2 \bullet [M]_{t_{i+1}} - H^2 \bullet [M]_{t_i} \,\big|\, \mathcal{F}_{t_i}\big], \end{aligned}$$

where we have used twice that h_i is $\mathcal{F}_{t_i}$-measurable and bounded, and in the fourth step also that $M^2 - [M]$ is a martingale. Summing up and taking expectations then gives (1.1). Moreover, it is not very difficult to argue that

$$\Delta\bigg(\int H^2 \,\mathrm{d}[M] \bigg) = H^2 \Delta[M] = H^2 (\Delta M)^2 = \big(\Delta(H \bullet M)\big)^2$$

for $H \in b\mathcal{E}$, by exploiting that H is piecewise constant and $\Delta[M] = (\Delta M)^2$. In view of Theorem 1.1 and the uniqueness there, the combination of these two properties can also

be formulated as saying that

$$[H \bullet M] = \left[\int H \, \mathrm{d}M \right] = \int H^2 \, \mathrm{d}[M] = H^2 \bullet [M] \qquad \text{for } H \in b\mathcal{E}.$$

This completes the proof. **q.e.d.**

Remark. The argument in the proof of Lemma 1.3 actually shows that the process $(H \bullet M)^2 - \int H^2 \, \mathrm{d}[M]$ is a martingale. [→ *Exercise:* Prove this in detail.] See also Remark 1.2. ◇

Our goal is now to extend the above results from $H \in b\mathcal{E}$ to a larger class of integrands. To that end, it is useful to view stochastic processes as random variables on the *product space* $\overline{\Omega} := \Omega \times (0, \infty)$. (Recall that the values at 0 are irrelevant for stochastic integrals.) We define the *predictable σ-field* $\mathcal{P}$ on $\overline{\Omega}$ as the σ-field generated by all adapted left-continuous processes, and we call a stochastic process $H = (H_t)_{t>0}$ *predictable* if it is $\mathcal{P}$-measurable when viewed as a mapping $H : \overline{\Omega} \to \mathbb{R}$. As a consequence, every $H \in b\mathcal{E}$ is then predictable as it is adapted and left-continuous. We also define the (possibly infinite) measure $P_M := P \otimes [M]$ on $(\overline{\Omega}, \mathcal{P})$ by setting

$$\int_{\overline{\Omega}} Y \, \mathrm{d}P_M := E_M[Y] := E\left[\int_0^\infty Y_s(\omega) \, \mathrm{d}[M]_s(\omega) \right]$$

for $Y \geq 0$ predictable; the inner integral is defined ω-wise as a Lebesgue–Stieltjes integral because $t \mapsto [M]_t(\omega)$ is increasing, null at 0 and RCLL and so can be viewed as the distribution function of a (possibly infinite) ω-dependent measure on $(0, \infty)$. (Actually, one could even allow Y to be product-measurable here.) Note that $P_M = P \otimes [M]$ is not a product measure in general because unlike $\langle W \rangle_t = t$, the quadratic variation $[M]$ of a general local martingale M depends on both t and ω. Finally, we introduce the space

$$\begin{aligned} L^2(M) &:= L^2(M, P) := L^2(\overline{\Omega}, \mathcal{P}, P_M) \\ &= \Big\{ \text{all (equivalence classes of) predictable } H = (H_t)_{t>0} \text{ such that} \\ &\qquad \|H\|_{L^2(M)} := (E_M[H^2])^{\frac{1}{2}} = \left(E\left[\int_0^\infty H_s^2 \, \mathrm{d}[M]_s \right] \right)^{\frac{1}{2}} < \infty \Big\}. \end{aligned}$$

(As usual, taking equivalence classes means that we identify H and H' if they agree P_M-a.e. on $\overline{\Omega}$ or, equivalently, if $E[\int_0^\infty (H_s - H'_s)^2 \,\mathrm{d}[M]_s] = 0$.)

With the above notations, we can restate the first half of Lemma 1.3 as follows:

For a fixed square-integrable martingale M, the mapping $H \mapsto H \bullet M$ is linear and goes from $b\mathcal{E}$ to the space $\mathcal{M}_0^2$ of all RCLL martingales $N = (N_t)_{t\geq 0}$ null at 0 which satisfy $\sup_{t\geq 0} E[N_t^2] < \infty$.

The last assertion is true because each $H \bullet M$ remains constant after some t_n given by $H \in b\mathcal{E}$, and because Doob's inequality gives for any martingale N and any $t \geq 0$ that

$$E\Big[\sup_{0\leq s\leq t} |N_s|^2\Big] \leq 4E[N_t^2].$$

Now the martingale convergence theorem implies that each $N \in \mathcal{M}_0^2$ admits a limit $N_\infty = \lim_{t\to\infty} N_t$ P-a.s., and we have $N_\infty \in L^2$ by Fatou's lemma, and the process $(N_t)_{0\leq t\leq\infty}$ defined up to ∞, i.e. on the *closed* interval $[0,\infty]$, is still a martingale. Moreover, Doob's maximal inequality implies that two martingales N and N' which have the same final value, i.e. $N_\infty = N'_\infty$ P-a.s., must coincide. Therefore we can identify $N \in \mathcal{M}_0^2$ with its limit $N_\infty \in L^2(\mathcal{F}_\infty, P)$, and so $\mathcal{M}_0^2$ becomes a *Hilbert space* with the norm

$$\|N\|_{\mathcal{M}_0^2} = \|N_\infty\|_{L^2} = (E[N_\infty^2])^{\frac{1}{2}}$$

and the scalar product

$$(N, N')_{\mathcal{M}_0^2} = (N_\infty, N'_\infty)_{L^2} = E[N_\infty N'_\infty].$$

Rephrasing Lemma 1.3 once again, we see that

the mapping $H \mapsto H \bullet M$ from $b\mathcal{E}$ to $\mathcal{M}_0^2$ is linear *and an* isometry

because (1.1) says that for $H \in b\mathcal{E}$,

$$\|H \bullet M\|_{\mathcal{M}_0^2} = \big(E[(H \bullet M_\infty)^2]\big)^{\frac{1}{2}} = \left(E\left[\int_0^\infty H_s^2 \,\mathrm{d}[M]_s\right]\right)^{\frac{1}{2}} = \|H\|_{L^2(M)}. \tag{1.2}$$

By general principles, this mapping can therefore be uniquely extended to the *closure of* $b\mathcal{E}$ *in* $L^2(M)$; in other words, we can define a stochastic integral process $H \bullet M$ for every H that can be approximated, with respect to the norm $\|\cdot\|_{L^2(M)}$, by processes from $b\mathcal{E}$, and the resulting $H \bullet M$ is again a martingale in $\mathcal{M}_0^2$ and still satisfies the isometry property (1.2).

(The argument behind these general principles is quite standard. If $(H^n)_{n\in\mathbb{N}}$ is a sequence of predictable processes converging to H with respect to $\|\cdot\|_{L^2(M)}$, then (H^n) is also a Cauchy sequence with respect to $\|\cdot\|_{L^2(M)}$. If all the H^n are in $b\mathcal{E}$, then the stochastic integral process $H^n \bullet M$ is well defined and in $\mathcal{M}_0^2$ for each n by Lemma 1.3. Moreover, by the isometry property in Lemma 1.3 for integrands in $b\mathcal{E}$, the sequence $(H^n \bullet M)_{n\in\mathbb{N}}$ is then also a Cauchy sequence in $\mathcal{M}_0^2$, and because $\mathcal{M}_0^2$ is a Hilbert space, hence complete, that Cauchy sequence must have a limit which is again in $\mathcal{M}_0^2$. This limit is then defined to be the stochastic integral $H \bullet M$ of H with respect to M. That the isometry property extends to the limit is also standard.)

The crucial question now is of course how we can describe the closure of $b\mathcal{E}$ and especially how big it is — the bigger the better, because we then have many integrands.

Proposition 1.4. *Suppose that M is in $\mathcal{M}_0^2$. Then:*

1) *$b\mathcal{E}$ is dense in $L^2(M)$, i.e. the closure of $b\mathcal{E}$ in $L^2(M)$ is $L^2(M)$. In other words, every $H \in L^2(M)$ can be written as a limit, with respect to the norm $\|\cdot\|_{L^2(M)}$, of a sequence $(H^n)_{n\in\mathbb{N}}$ in $b\mathcal{E}$.*

2) *For every $H \in L^2(M)$, the stochastic integral process $H \bullet M = \int H \,\mathrm{d}M$ is well defined, in $\mathcal{M}_0^2$ and satisfies* (1.2).

Proof. Assertion 1) uses a martingale approximation argument on $\overline{\Omega}$ which we do not

give here. However, we point out that the assumption $M \in \mathcal{M}_0^2$ is used to ensure that P_M is a finite measure. Assertion 2) is then clear from the discussion above. **q.e.d.**

By definition, saying that M is in $\mathcal{M}_0^2$ means that M is an RCLL martingale null at 0 with $\sup_{t\geq 0} E[M_t^2] < \infty$. In particular, we then have $E[M_t^2] < \infty$ for every $t \geq 0$ so that every $M \in \mathcal{M}_0^2$ is also a square-integrable martingale. However, the converse is not true; Brownian motion W for example is a martingale and has $E[W_t^2] = t$ so that $\sup_{t\geq 0} E[W_t^2] = +\infty$, which means that BM is not in $\mathcal{M}_0^2$. This makes it clear that we need to extend our approach to stochastic integration further. This can be done via *localisation*.

Definition. We call a local martingale M null at 0 *locally square-integrable* and write $M \in \mathcal{M}_{0,\,\mathrm{loc}}^2$ if there is a sequence of stopping times $\tau_n \nearrow \infty$ P-a.s. such that $M^{\tau_n} \in \mathcal{M}_0^2$ for each n. We say for a predictable process H that $H \in L^2_{\mathrm{loc}}(M)$ if there exists a sequence of stopping times $\tau_n \nearrow \infty$ P-a.s. such that $HI_{[\![0,\tau_n]\!]} \in L^2(M)$ for each n. Here we use the *stochastic interval* notation $[\![0,\tau_n]\!] := \{(\omega,t) \in \overline{\Omega} : 0 < t \leq \tau_n(\omega)\}$.

More generally, if we have a class $\mathcal{C}$ of stochastic processes, we define the *localised class* $\mathcal{C}_{\mathrm{loc}}$ by saying that a process X *is in* $\mathcal{C}_{\mathrm{loc}}$ or that X *is locally in* $\mathcal{C}$ if there exists a sequence of stopping times $\tau_n \nearrow \infty$ P-a.s. such that X^{τ_n} is in $\mathcal{C}$ for each n. If the process we consider is an integrand H, then we have to require instead that $HI_{[\![0,\tau_n]\!]}$ is in $\mathcal{C}$ for each n.

For $M \in \mathcal{M}_{0,\,\mathrm{loc}}^2$ and $H \in L^2_{\mathrm{loc}}(M)$, defining the stochastic integral is straightforward; we simply set

$$H \bullet M := (HI_{[\![0,\tau_n]\!]}) \bullet M^{\tau_n} \qquad \text{on } [\![0,\tau_n]\!]$$

which gives a definition on all of $\overline{\Omega}$, because $\tau_n \nearrow \infty$ so that $[\![0,\tau_n]\!]$ increases to $\overline{\Omega}$. The only point we need to check is that this definition is *consistent*, i.e. that the definition on $[\![0,\tau_{n+1}]\!] \supseteq [\![0,\tau_n]\!]$ does not clash with the definition on $[\![0,\tau_n]\!]$. This can be done by using the (subsequently listed) properties of stochastic integrals, but we do not go into details here. Of course, $H \bullet M$ is then in $\mathcal{M}_{0,\,\mathrm{loc}}^2$.

Remarks. 1) A closer look at the developments so far shows that the *definitions* (but not the preceding results and arguments) for P_M and $L^2(M)$ only need $[M]$; hence one can introduce and use them for any local martingale M, due to Theorem 1.1.

2) One can also define a stochastic integral process $H \bullet M$ for $H \in L^2_{\mathrm{loc}}(M)$ when M is a general local martingale, but this requires substantially more theory. For more details, see Dellacherie/Meyer [5, Theorem VIII.37].

3) If M is $\mathbb{R}^d$-valued with components M^i that all are local martingales null at 0, one can also define the so-called *vector stochastic integral* $H \bullet M$ for $\mathbb{R}^d$-valued predictable processes in a suitable space $L^2_{\mathrm{loc}}(M)$; the result is then a real-valued process. Details can be found in Jacod/Shiryaev [11, Sections III.4a and III.6a]. However, one *warning* is indicated: $L^2_{\mathrm{loc}}(M)$ is not obtained by just asking that each component H^i should be in $L^2_{\mathrm{loc}}(M^i)$ and then setting $H \bullet M = \sum_i H^i \bullet M^i$. In fact, it can happen that $H \bullet M$ is well defined whereas the individual $H^i \bullet M^i$ are not. So the intuition for the multidimensional case is that

$$\text{“} \int H \, \mathrm{d}M = \int \sum_i H^i \, \mathrm{d}M^i \neq \sum_i \int H^i \, \mathrm{d}M^i \text{”},$$

as we have already pointed out in Remark 1.2.2.

4) One can extend the stochastic integral even further to more general integrands in a space called $L(M)$, but this becomes technical and also has a nontrivial *pitfall*: There are (real-valued) local martingales M and predictable integrands H such that the stochastic integral process $\int H \, \mathrm{d}M$ is well defined, but not a local martingale (!). This is in marked contrast to discrete time; see Theorem 1.3.1. We remark, however, that this can only happen if M has jumps. $\diamond$

To end this section on a positive note, let us consider the case where M is a *continuous* local martingale null at 0, briefly written as $M \in \mathcal{M}^c_{0,\mathrm{loc}}$. This includes in particular the case of a Brownian motion W. Then M is in $\mathcal{M}^2_{0,\,\mathrm{loc}}$ because it is even *locally bounded*: For the stopping times

$$\tau_n := \inf\{t \geq 0 : |M_t| > n\} \nearrow \infty \qquad P\text{-a.s.},$$

we have by continuity that $|M^{\tau_n}| \leq n$ for each n, because

$$\left|M_t^{\tau_n}\right| = |M_{t\wedge\tau_n}| = \begin{cases} |M_t| \leq n & \text{if } t < \tau_n, \\ |M_{\tau_n}| = n & \text{if } t \geq \tau_n. \end{cases}$$

(Note that continuity of M is only used to obtain the equality $|M_{\tau_n}| = n$; everything else works just as well if M is only assumed to be adapted and RCLL.) The set $L^2_{\mathrm{loc}}(M)$ of nice integrands for M can here be explicitly described as

$$L^2_{\mathrm{loc}}(M) = \Bigg\{ \text{all predictable processes } H = (H_t)_{t>0} \text{ such that} \\ \int_0^t H_s^2 \,\mathrm{d}[M]_s = \int_0^t H_s^2 \,\mathrm{d}\langle M\rangle_s < \infty \ P\text{-a.s. for each } t \geq 0 \Bigg\}.$$

Finally, the resulting stochastic integral $H \bullet M = \int H \,\mathrm{d}M$ is then (as we shall see from the properties in Section 5.2 below) also a continuous local martingale, and of course null at 0.

5.2 Properties

As with usual integrals, one very rarely computes a stochastic integral by passing to the limit from some approximation. One works with stochastic integrals by using a set of *rules* and *properties*. These are listed in this section, without proofs.

- *(Local) Martingale properties:*
  - If M is a local martingale and $H \in L^2_{\mathrm{loc}}(M)$, then $\int H \,\mathrm{d}M$ is a local martingale in $\mathcal{M}^2_{0,\,\mathrm{loc}}$. If $H \in L^2(M)$, then $\int H \,\mathrm{d}M$ is even a martingale in $\mathcal{M}^2_0$.
  - If M is a local martingale and H is predictable and locally bounded (which means that there are stopping times $\tau_n \nearrow \infty$ P-a.s. such that $H I_{[\![0,\tau_n]\!]}$ is bounded by a constant c_n, say, for each $n \in \mathbb{N}$), then $\int H \,\mathrm{d}M$ is a local martingale.
  - If M is a martingale in $\mathcal{M}^2_0$ and H is predictable and bounded, then $\int H \,\mathrm{d}M$ is again a martingale in $\mathcal{M}^2_0$.
  - *Warning*: If M is a martingale and H is predictable and bounded, then $\int H \,\mathrm{d}M$ need not be a martingale; this is in striking contrast to the situation in discrete time.

- *Linearity:*
  - If M is a local martingale and H, H' are in $L^2_{\mathrm{loc}}(M)$ and $a, b \in \mathbb{R}$, then also $aH + bH'$ is in $L^2_{\mathrm{loc}}(M)$ and

$$(aH + bH')\bullet M = (aH)\bullet M + (bH')\bullet M = a(H\bullet M) + b(H'\bullet M).$$

- *Associativity:*
  - If M is a local martingale and $H \in L^2_{\mathrm{loc}}(M)$, we already know that $H\bullet M$ is again a local martingale. Then a predictable process K is in $L^2_{\mathrm{loc}}(H\bullet M)$ if and only if the product KH is in $L^2_{\mathrm{loc}}(M)$, and then

$$K\bullet(H\bullet M) = (KH)\bullet M,$$

i.e.

$$\int K \,\mathrm{d}\Big(\int H \,\mathrm{d}M \Big) = \int KH \,\mathrm{d}M.$$

- *Behaviour under stopping:*
  - Suppose that M is a local martingale, $H \in L^2_{\mathrm{loc}}(M)$ and τ is a stopping time. Then M^τ is a local martingale by the stopping theorem, H is in $L^2_{\mathrm{loc}}(M^\tau)$, $HI_{[\![0,\tau]\!]}$ is in $L^2_{\mathrm{loc}}(M)$, and we have

$$(H \bullet M)^\tau = H \bullet (M^\tau) = (HI_{[\![0,\tau]\!]}) \bullet M = (HI_{[\![0,\tau]\!]}) \bullet (M^\tau).$$

    In words: A stopped stochastic integral is computed by either first stopping the integrator and then integrating, or setting the integrand equal to 0 after the stopping time and then integrating, or combining the two.

- *Quadratic variation and covariation:*
  - Suppose that M, N are local martingales, $H \in L^2_{\mathrm{loc}}(M)$ and $K \in L^2_{\mathrm{loc}}(N)$. Then

$$\Big[\int H \,\mathrm{d}M, N \Big] = \int H \,\mathrm{d}[M, N]$$

    and

$$\Big[\int H \,\mathrm{d}M, \int K \,\mathrm{d}N \Big] = \int HK \,\mathrm{d}[M, N].$$

    In words: The covariation process of two stochastic integrals is obtained by integrating the product of the integrands with respect to the covariation process of the integrators.

  - In particular, $[\int H \,\mathrm{d}M] = \int H^2 \,\mathrm{d}[M]$. (We have seen this already for $H \in b\mathcal{E}$ in Lemma 1.3.)

- *Jumps:*
  - Suppose M is a local martingale and $H \in L^2_{\mathrm{loc}}(M)$. Then we already know that $H \bullet M$ is in $\mathcal{M}^2_{0,\,\mathrm{loc}}$ and therefore RCLL. Its jumps are given by

$$\Delta\left(\int H \,\mathrm{d}M\right)_t = H_t \,\Delta M_t \qquad \text{for } t > 0,$$

    where $\Delta Y_t := Y_t - Y_{t-}$ again denotes the jump at time t of a process Y with trajectories which are RCLL (right-continuous and having left limits).

Example. To illustrate why the direct use of the definitions is complicated, let us compute the stochastic integral $\int W \,\mathrm{d}W$ for a Brownian motion W. This is well defined because $M := W$ is in $\mathcal{M}^2_{0,\,\mathrm{loc}}$ (it is even continuous) and $H := W$ is predictable and locally bounded, because it is adapted and continuous.

Because

$$2W_{t_i}(W_{t_{i+1}} - W_{t_i}) = W^2_{t_{i+1}} - W^2_{t_i} - (W_{t_{i+1}} - W_{t_i})^2$$

by elementary algebra, we obtain by summing up that

$$\sum_{t_i \in \Pi_n} W_{t_i \wedge t}(W_{t_{i+1} \wedge t} - W_{t_i \wedge t}) = \frac{1}{2}(W_t^2 - W_0^2) - \frac{1}{2}\sum_{t_i \in \Pi_n} (W_{t_{i+1} \wedge t} - W_{t_i \wedge t})^2.$$

If the mesh size $|\Pi_n|$ of the partition sequence (Π_n) goes to 0, then the sum on the right-hand side converges P-a.s. to t by Theorem 4.1.4, if the partitions are also refining. We therefore expect to obtain

$$\int_0^t W_s \,\mathrm{d}W_s = \frac{1}{2}W_t^2 - \frac{1}{2}t,$$

and we shall see later from Itô's formula that this is indeed correct. Note that we should expect the first term $\frac{1}{2}W_t^2$ from classical calculus (where we have $\int_0^x y \,\mathrm{d}y = \frac{1}{2}x^2$); the second-order correction term $\frac{1}{2}t$ appears due to the quadratic variation of Brownian trajectories.

Exercise: Prove directly (without using the above result) that the stochastic integral process $\int W \,\mathrm{d}W$ is a martingale, but not in $\mathcal{M}^2_0$.

Exercise: Compute the Stratonovich integral and the backward integral for $\int W \, \mathrm{d}W$, and analyse their properties.

Exercise: Prove that if H is predictable and bounded, then $\int H \, \mathrm{d}W$ is a square-integrable martingale.

Exercise: For any local martingale M null at 0 and any stopping time τ, prove that we have $[M]^\tau = [M^\tau]$.

5.3 Extension to semimartingales

So far, we have seen two ideas for constructing stochastic integrals $\int H \, \mathrm{d}X$ of some process H with respect to another process X:

a) In Section 5.1, we have taken for $X = M$ a *local martingale* null at 0 and for H a process in $L^2_{\mathrm{loc}}(M)$; this means that H must be predictable and possess some integrability.

b) If $X = A$ has trajectories $t \mapsto A_t(\omega)$ that are of *finite variation*, we can classically define $\int H_s(\omega) \, \mathrm{d}A_s(\omega)$ for each ω (pathwise) as a Lebesgue–Stieltjes integral. This requires some measurability and integrability for $s \mapsto H_s(\omega)$.

Because integration is a linear operation, the obvious and easy idea for an extension is therefore to look at processes that are sums of the above two types, because we can then define an integral with respect to the sum as the sum of the two integrals.

Definition. A *semimartingale* is a stochastic process $X = (X_t)_{t \geq 0}$ that can be decomposed as $X = X_0 + M + A$, where M is a local martingale null at 0 and A is an adapted RCLL process null at 0 and having trajectories of finite variation. A semimartingale X is called *special* if there exists such a decomposition where A is in addition predictable.

Remark 3.1. 1) If X is a special semimartingale, the decomposition with A predictable is *unique* and called the *canonical decomposition*. The uniqueness result is based on the useful fact that *any local martingale which is predictable and of finite variation must be constant*.

2) If X is a *continuous* semimartingale, both M and A can be chosen continuous as well. Therefore X is special because A is then predictable, as it is adapted and continuous.

3) If X is a semimartingale, we define its *optional quadratic variation* or *square bracket* process $[X] = ([X]_t)_{t \geq 0}$ via

$$[X] := [M] + 2[M, A] + [A] := [M] + 2 \sum \Delta M \Delta A + \sum (\Delta A)^2.$$

One can show that this is well defined and does not depend on the chosen decomposition of X. Moreover, $[X]$ can also be obtained as a quadratic variation similarly as in Theorem 1.1; see Section 6.1 below for more details. However, $X^2 - [X]$ is no longer a local martingale, but only a semimartingale in general. $\diamond$

If X is a semimartingale, we can define a stochastic integral $H \bullet X = \int H \, \mathrm{d}X$ at least for any process H which is predictable and locally bounded. We simply set

$$H \bullet X := H \bullet M + H \bullet A,$$

where $H \bullet M$ is as in Section 5.1 and $H \bullet A$ is defined ω-wise as a Lebesgue–Stieltjes integral. Of course one still needs to check that this is well defined (e.g. without ambiguity if X has several decompositions), but this can be done; see for instance Dellacherie/Meyer [5, Section VIII.1] or Jacod/Shiryaev [11, Section I.4d].

The resulting stochastic integral then has all the *properties* from Section 5.2 except those that rest in an essential way on the (local) martingale property; so the isometry property for example is of course lost. But we still have, for H predictable and locally bounded:

- $H \bullet X$ is a *semimartingale*.
- If X is special with canonical decomposition $X = X_0 + M + A$, then $H \bullet X$ is also special, with canonical decomposition $H \bullet X = H \bullet M + H \bullet A$.
  [This uses the non-obvious fact that if A is predictable and of finite variation and H is predictable and locally bounded, the pathwise defined integral $H \bullet A$ can be chosen to be predictable again.]
- *linearity:* same formula as before.
- *associativity:* same formula as before.
- *behaviour under stopping:* same formula as before.
- *quadratic variation and covariation:* same formula as before.

– *jumps:* same formula as before.

– If X is *continuous*, then so is $H \bullet X$; this is clear from $\Delta(H \bullet X) = H\Delta X = 0$.

In addition, there is also a sort of *dominated convergence theorem*: If H^n, $n \in \mathbb{N}$, are predictable processes with $H^n \to 0$ pointwise on $\overline{\Omega}$ and $|H^n| \leq |H|$ for some locally bounded H, then $H^n \bullet X \to 0$ uniformly on compacts in probability, which means that

$$\sup_{0 \leq s \leq t} |H^n \bullet X_s| \longrightarrow 0 \qquad \text{in probability as } n \to \infty, \text{ for every } t \geq 0. \tag{3.1}$$

This can also be viewed as a *continuity property* of the stochastic integral operator $H \mapsto H \bullet X$, because (pointwise and locally bounded) convergence of (H^n) implies convergence of $(H^n \bullet X)$, in the ucp sense of (3.1).

From the whole approach above, the definition of a semimartingale looks completely ad hoc and rather artificial. But it turns out that this concept is in fact very natural and has a number of very good properties:

1) If X is a semimartingale and f is a C^2-function, then $f(X)$ is again a semimartingale. This will follow from *Itô's formula*, which even gives an explicit expression for $f(X)$.

2) If X is a semimartingale with respect to P and R is a probability measure equivalent to P, then X is still a semimartingale with respect to R. This will follow from *Girsanov's theorem*, which even gives a decomposition of X under R.

3) If X is any adapted process with RC trajectories, we can always define the (elementary) stochastic integral $H \bullet X$ for processes H in $b\mathcal{E}$. If X is such that this mapping on $b\mathcal{E}$ also has the continuity property (3.1) for any sequence $(H^n)_{n \in \mathbb{N}}$ in $b\mathcal{E}$ converging pointwise to 0 and with $|H^n| \leq 1$ for all n, then X must in fact be a semimartingale. This deep result is due to *Bichteler and Dellacherie* and shows that semimartingales are a *natural class of integrators*.

One direct consequence of 2) for finance is that semimartingales are the natural processes to model discounted asset prices in financial markets. In fact, the fundamental

theorem of asset pricing (in a suitably general version for continuous-time models) essentially says that a suitably arbitrage-free model should be such that S is a local martingale (or more generally a σ-martingale) under some $Q \approx P$. But then S is a Q-semimartingale and thus by 2) also a P-semimartingale.

Put differently, the above result implies that if we start with any model where S is *not a semimartingale*, there will be *arbitrage* of some kind. Things become different if one includes transaction costs; but in frictionless markets, one must be careful about this issue.

Remark. We have explained so far how to obtain a stochastic integral $H \bullet X$ for semimartingales X and locally bounded predictable H. The Bichteler–Dellacherie result shows that one cannot go beyond semimartingales without a serious loss; but because not every predictable process is locally bounded, one can ask if, for a given semimartingale X, there are more possible integrands H for X. This leads to the notion and definition of the class $L(X)$ of *X-integrable* processes; but the development of this requires rather advanced results and techniques from stochastic calculus, and so we cannot go into details here. See Dellacherie/Meyer [5, Section VIII.3] or Jacod/Shiryaev [11, Section III.6]. Alternatively, this is usually presented in the course "Mathematical Finance". $\diamond$

6 Stochastic calculus

Our goal in this chapter is to provide the basic tools, results and techniques for working with stochastic processes and especially stochastic integrals in continuous time. This will be used in the next chapter when we discuss continuous-time option pricing and in particular the famous Black–Scholes formula.

Throughout this chapter, we work on a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathbb{F} = (\mathcal{F}_t)$ satisfying the usual conditions of right-continuity and P-completeness. For all local martingales, we then can and tacitly do choose a version with RCLL trajectories. For the time parameter t, we have either $t \in [0, T]$ with a fixed time horizon $T \in (0, \infty)$ or $t \geq 0$. In the latter case, we set

$$\mathcal{F}_\infty := \bigvee_{t \geq 0} \mathcal{F}_t := \sigma\Big(\bigcup_{t \geq 0} \mathcal{F}_t \Big).$$

6.1 Itô's formula

The question to be addressed in this section is very simple. If X is a semimartingale and f is some (suitable) function, what can we say about the stochastic process $f(X)$? What kind of process is it, and what does it look like in more detail?

In the simplest case, let $x : [0, \infty) \to \mathbb{R}$ be a function $t \mapsto x(t)$ and think of x as a typical trajectory $t \mapsto X_t(\omega)$ of X. The *classical chain rule from analysis* then says that if x is in C^1 (i.e. continuously differentiable) and $f : \mathbb{R} \to \mathbb{R}$ is in C^1, the composition $f \circ x : [0, \infty) \to \mathbb{R}$, $t \mapsto f(x(t))$ is again in C^1 and its derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ x)(t) = \frac{\mathrm{d}f}{\mathrm{d}x}\big(x(t)\big)\, \frac{\mathrm{d}x}{\mathrm{d}t}(t),$$

or more compactly

$$(f \circ x)^{\cdot}\,(t) = f'\big(x(t)\big)\, \dot{x}(t),$$

where the dot $\dot{}$ denotes the derivative with respect to t and the prime $'$ is the derivative with respect to x. In formal differential notation, we can rewrite this as

$$\mathrm{d}(f \circ x)(t) = f'\big(x(t)\big)\, \mathrm{d}x(t), \tag{1.1}$$

or in integral form

$$f\big(x(t)\big) - f\big(x(0)\big) = \int_0^t f'\big(x(s)\big)\,\mathrm{d}x(s). \tag{1.2}$$

In this last form, the chain rule can be extended to the case where f is in C^1 and x is *continuous* and of *finite variation*.

Unfortunately, this classical result does not help us a lot. For one thing, X might have only RCLL instead of continuous trajectories. This is still solvable if X has trajectories of finite variation. But even if X is continuous, we cannot hope that its trajectories are of finite variation, as the example of X being a Brownian motion clearly demonstrates. So we need a different result, namely a chain rule for functions having a nonzero quadratic variation.

Let us now connect the above idea to semimartingales. Recall that a *semimartingale* is a stochastic process of the form $X = X_0 + M + A$, where M is a local martingale null at 0 and A is an adapted process null at 0 with RCLL trajectories of finite variation. For any such A and any fixed, i.e. nonrandom, sequence $(\Pi_n)_{n\in\mathbb{N}}$ of partitions of $[0,\infty)$ with $\lim_{n\to\infty}|\Pi_n| = 0$, the *quadratic variation* of A along $(\Pi_n)_{n\in\mathbb{N}}$ is given by the sum of the squared jumps of A, i.e.

$$[A]_t = \lim_{n\to\infty} \sum_{t_i\in\Pi_n} (A_{t_{i+1}\wedge t} - A_{t_i\wedge t})^2 = \sum_{0<s\le t} (\Delta A_s)^2 = \sum_{0<s\le t} (A_s - A_{s-})^2 \qquad \text{for } t \ge 0.$$

By polarisation, we then obtain for any semimartingale Y that

$$[A,Y]_t = \sum_{0<s\le t} \Delta A_s \Delta Y_s \qquad \text{for } t \ge 0.$$

So the quadratic variation of a general semimartingale $X = X_0 + M + A$ has the form

$$[X] = [M+A] = [M] + [A] + 2[M,A] = [M] + \sum_{0<s\le\cdot} (\Delta A_s)^2 + 2\sum_{0<s\le\cdot} \Delta M_s \Delta A_s.$$

This partly repeats Remark 5.3.1. If A is continuous, we obtain that $[X] = [M]$, even if X (hence M) is only RCLL.

Now suppose that X is a *continuous* semimartingale. As already pointed out in Remark 5.3.1, the processes M and A can then also be chosen continuous. A simple result from analysis [→ *exercise*] says that

(1.3) any continuous function of finite variation has zero quadratic variation along any sequence $(\Pi_n)_{n\in\mathbb{N}}$ of partitions of $[0,\infty)$ whose mesh size $|\Pi_n|$ goes to 0 as $n \to \infty$.

(Note that this is a variant of the result already mentioned in Remark 4.1.5 in Chapter 4.) So if the semimartingale X is continuous, then its (unique) finite variation part A has zero quadratic variation, and its (unique) local martingale part M has quadratic variation $[M] = \langle M \rangle$; see Remark 5.1.2 in Chapter 5. The covariation of M and A is thus also zero by Cauchy–Schwarz. A continuous semimartingale X with canonical decomposition $X = X_0 + M + A$ therefore has the quadratic variation $[X] = \langle X \rangle = [M] = \langle M \rangle$ which is again continuous.

Now let us return to the transformation $f(X)$ of a semimartingale X by a function f. In the simplest case, the answer to our basic question in this section looks as follows.

Theorem 1.1 (Itô's formula I). *Suppose $X = (X_t)_{t\geq 0}$ is a continuous real-valued semimartingale and $f : \mathbb{R} \to \mathbb{R}$ is in C^2. Then $f(X) = (f(X_t))_{t\geq 0}$ is again a continuous (real-valued) semimartingale, and we explicitly have P-a.s.*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}\langle X\rangle_s \tag{1.4}$$

for all $t \geq 0$.

Remarks. 1) Not only the result is important, but also the basic idea for its proof.

2) The $\mathrm{d}X$-integral in (1.4) is a stochastic integral; it is well defined because X is a semimartingale and $f'(X)$ is adapted and continuous, hence predictable and locally bounded. The $\mathrm{d}\langle X\rangle$-integral is a classical Lebesgue–Stieltjes integral because $\langle X \rangle$ has

increasing trajectories; it is also well defined because $f''(X)$ is also predictable and locally bounded.

3) In purely formal *differential notation*, (1.4) is usually written more compactly as

$$\mathrm{d}f(X_t) = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)\,\mathrm{d}\langle X\rangle_t = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)\,\mathrm{d}\langle M\rangle_t, \tag{1.5}$$

using that $\langle X\rangle = \langle M\rangle$.

4) Comparing (1.1), (1.2) to (1.5), (1.4) shows that we have in comparison to the classical chain rule an extra *second-order term* coming from the *quadratic variation* of X (or here more precisely from the quadratic variation of the martingale part M of X). This is the important point to remember, and it also shows up in the proof.

5) One can view Itô's formula and its proof as a purely analytical result which provides an *extension of the chain rule* for $f \circ x$ to functions x that have a nonzero quadratic variation. This has been pointed out and developed by Hans Föllmer [8]. Not surprisingly, relaxing the assumptions on x then requires stronger assumptions on f than in the classical case (C^2 instead of C^1).

6) To see the financial relevance of Itô's formula, think of X as some underlying financial asset and of $Y = f(X)$ as a new product obtained from the underlying by a possibly nonlinear transformation f. Then (1.4) or (1.5) show us how the product reacts to changes in the underlying. The important message of Theorem 1.1 is then that when using stochastic models (for X), *a simple linear approximation is not good enough; one must also account for the second-order behaviour of* X. ◇

Proof of Theorem 1.1. The easiest way to remember both the result and its proof for the case where X is continuous is via the following *quick and dirty* argument: "A Taylor expansion at the infinitesimal level gives

$$\mathrm{d}f(X_t) = f(X_t) - f(X_{t-\mathrm{d}t}) = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)(\,\mathrm{d}X_t)^2,$$

and $(\mathrm{d}X_t)^2 = (X_t - X_{t-\mathrm{d}t})^2 = \langle X\rangle_t - \langle X\rangle_{t-\mathrm{d}t} = \mathrm{d}\langle X\rangle_t$." Note, however, that this reasoning is purely formal and does not constitute a correct proof. (For example, it does not explain why we stop at the second and not at another higher order in the expansion.)

To make the above idea rigorous, we write for non-infinitesimal increments

$$f(X_{t_{i+1}\wedge t}) - f(X_{t_i}) = f'(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i}) + \frac{1}{2}f''(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i})^2 + R_i,$$

where R_i stands for the error term in the Taylor expansion and the t_i come from a partition Π_n of $[0,\infty)$. Now we sum over the $t_i \le t$ and obtain on the left-hand side a telescoping sum which equals $f(X_t) - f(X_0)$. When we study the terms on the right-hand side, we first recall the convergence

$$Q_t^{\Pi_n} := \sum_{t_i\in\Pi_n,\, t_i\le t} (X_{t_{i+1}\wedge t} - X_{t_i})^2 \longrightarrow \langle X\rangle_t \qquad \text{as } |\Pi_n| \to 0$$

from Theorem 5.1.1; see also Remark 5.3.1. This implies firstly by a weak convergence argument that

$$\frac{1}{2}\sum_{t_i\in\Pi_n,\, t_i\le t} f''(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i})^2 \longrightarrow \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}\langle X\rangle_s,$$

and secondly by a careful estimate that

$$\sum_{t_i\in\Pi_n,\, t_i\le t} |R_i| \longrightarrow 0.$$

(This is exactly the point where the mathematical analysis shows why the second order is the correct order of expansion.) As a consequence, the sums

$$\sum_{t_i\in\Pi_n,\, t_i\le t} f'(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i})$$

must also converge, and the dominated convergence theorem for stochastic integrals then implies that the limit is $\int_0^t f'(X_s)\,\mathrm{d}X_s$. **q.e.d.**

Example. For $X = W$ a Brownian motion and $f(x) = x^2$, we obtain $f'(x) = 2x$, $f''(x) \equiv 2$ and therefore

$$W_t^2 = W_0^2 + \int_0^t 2W_s\,\mathrm{d}W_s + \frac{1}{2}\int_0^t 2\,\mathrm{d}\langle W\rangle_s.$$

Using $W_0 = 0$ and the fact that BM has quadratic variation $\langle W \rangle_t = t$, hence $\mathrm{d}\langle W \rangle_s = \mathrm{d}s$, gives

$$W_t^2 = 2\int_0^t W_s \,\mathrm{d}W_s + \int_0^t \mathrm{d}s = 2\int_0^t W_s \,\mathrm{d}W_s + t$$

or rewritten

$$\int_0^t W_s \,\mathrm{d}W_s = \frac{1}{2}W_t^2 - \frac{1}{2}t.$$

This ties up with the example we have seen in Section 5.2.

Before moving on to more examples, we need an extension of Theorem 1.1.

Theorem 1.2 (**Itô's formula II**). *Suppose $X = (X_t)_{t\geq 0}$ is a general $\mathbb{R}^d$-valued semimartingale and $f : \mathbb{R}^d \to \mathbb{R}$ is in C^2. Then $f(X) = (f(X_t))_{t\geq 0}$ is again a (real-valued) semimartingale, and we explicitly have P-a.s. for all $t \geq 0$*

1) *if X has continuous trajectories:*

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s)\,\mathrm{d}\langle X^i, X^j\rangle_s, \tag{1.6}$$

or in more compact notation, using subscripts to denote partial derivatives,

$$\mathrm{d}f(X_t) = \sum_{i=1}^d f_{x^i}(X_t)\,\mathrm{d}X_t^i + \frac{1}{2}\sum_{i,j=1}^d f_{x^i x^j}(X_t)\,\mathrm{d}\langle X^i, X^j\rangle_t.$$

2) *if $d = 1$ (so that X is real-valued, but not necessarily continuous):*

$$\begin{aligned} f(X_t) = f(X_0) &+ \int_0^t f'(X_{s-})\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_{s-})\,\mathrm{d}[X]_s \\ &+ \sum_{0<s\leq t}\left(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s - \frac{1}{2}f''(X_{s-})(\Delta X_s)^2\right). \end{aligned} \tag{1.7}$$

Proof. See Protter [13, Section II.7]. **q.e.d.**

Remark. There is of course also a version of Itô's formula for general $\mathbb{R}^d$-valued semimartingales (which contains both 1) and 2) as special cases). It looks similar to part 2) of Theorem 1.2, but has in addition sums like in part 1), with $\langle \cdot, \cdot \rangle$ replaced by $[\cdot, \cdot]$. And of course one could also write (1.7) in differential form. $\diamond$

If X is continuous, one frequently useful *simplification* of (1.6) arises if one or several of the components of X are of finite variation. If X^k, say, is of finite variation, then we know from (1.3) that $\langle X^k \rangle \equiv 0$ and hence also $\langle X^i, X^k \rangle \equiv 0$ for all i by Cauchy–Schwarz. (Recall that we have already used such an argument before Theorem 1.1.) This implies that all the second-order terms containing X^k will vanish; hence we do not need all the corresponding partial derivatives, and so we can also relax the assumptions on f in that regard.

Example 1.3. The CRR binomial model can be written as

$$\frac{\widetilde{S}^0_k - \widetilde{S}^0_{k-1}}{\widetilde{S}^0_{k-1}} = r,$$

$$\frac{\widetilde{S}^1_k - \widetilde{S}^1_{k-1}}{\widetilde{S}^1_{k-1}} = Y_k - 1 =: R_k = E[R_k] + (R_k - E[R_k]).$$

Note that the terms in brackets above has expectation 0 and a variance which depends on the distribution of the R_k. Passing from time steps of size 1 to $\mathrm{d}t$ and noting that Brownian increments have expectation 0 like the term $R_k - E[R_k]$, a continuous-time analogue would be of the form

$$\frac{\mathrm{d}\widetilde{S}^0_t}{\widetilde{S}^0_t} = r\,\mathrm{d}t, \tag{1.8}$$

$$\frac{\mathrm{d}\widetilde{S}^1_t}{\widetilde{S}^1_t} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{1.9}$$

(More accurately, we should put $\mathrm{d}\widetilde{S}^0_t/\widetilde{S}^0_{t-}$ and $\mathrm{d}\widetilde{S}^1_t/\widetilde{S}^1_{t-}$. But as both $\widetilde{S}^0$ and $\widetilde{S}^1$ turn out to be continuous, the difference does not matter.)

Of course, the equation (1.8) for $\widetilde{S}^0$ is just a very simple ordinary differential equation (ODE), whose solution for the starting value $\widetilde{S}^0_0 = 1$ is $\widetilde{S}^0_t = e^{rt}$. The equation (1.9) for

$\widetilde{S}^1$ is a *stochastic differential equation (SDE)*, and its solution is given by the *geometric Brownian motion (GBM)*

$$\widetilde{S}^1_t = \widetilde{S}^1_0 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right) \qquad \text{for } t \geq 0. \tag{1.10}$$

Note the possibly surprising term $-\frac{1}{2}\sigma^2$. To see that this is indeed a solution, we write

$$\widetilde{S}^1_t = f(W_t, t) \qquad \text{with } f(x,t) = \widetilde{S}^1_0\, e^{\sigma x + (\mu - \frac{1}{2}\sigma^2)t}.$$

We now apply Itô's formula (1.6) for $d = 2$ to $X_t = (W_t, t)$. As the second component $X^{(2)}_t = t$ is continuous and increasing, it has finite variation; so (1.6) simplifies and we only need the derivatives

$$\begin{aligned} f_x &= \frac{\partial f}{\partial x} = \sigma f, \\ f_t &= \frac{\partial f}{\partial t} = \left(\mu - \frac{1}{2}\sigma^2\right) f, \\ f_{xx} &= \frac{\partial^2 f}{\partial x^2} = \sigma^2 f. \end{aligned}$$

Then we get, by using that $\langle W\rangle_t = t$ and $f(W_t, t) = \widetilde{S}^1_t$, that

$$\begin{aligned} \mathrm{d}\widetilde{S}^1_t &= f_x(W_t,t)\,\mathrm{d}W_t + f_t(W_t,t)\,\mathrm{d}t + \frac{1}{2} f_{xx}(W_t,t)\,\mathrm{d}\langle W\rangle_t \\ &= \sigma \widetilde{S}^1_t\,\mathrm{d}W_t + \left(\mu - \frac{1}{2}\sigma^2\right)\widetilde{S}^1_t\,\mathrm{d}t + \frac{1}{2}\sigma^2 \widetilde{S}^1_t\,\mathrm{d}t \\ &= \widetilde{S}^1_t(\sigma\,\mathrm{d}W_t + \mu\,\mathrm{d}t), \end{aligned}$$

exactly as claimed. Note that we did not argue (as one should and can) that the above explicit process in (1.10) is the only solution of (1.9).

Example. If $X = (X_t)_{t\geq 0}$ is a continuous real-valued semimartingale null at 0, then

$$Z_t := e^{X_t - \frac{1}{2}\langle X\rangle_t} \qquad \text{for } t \geq 0 \tag{1.11}$$

is the unique solution of the SDE

$$\mathrm{d}Z_t = Z_t\,\mathrm{d}X_t, \qquad Z_0 = 1.$$

Put differently, this means that Z satisfies

$$Z_t = 1 + \int_0^t Z_s \, \mathrm{d}X_s \qquad \text{for all } t \geq 0, \; P\text{-a.s.}$$

Checking that the above Z does satisfy the above SDE, as well as proving uniqueness of the solution, is a good [→ *exercise*] in the use of Itô's formula.

Definition. For a general real-valued semimartingale X null at 0, the *stochastic exponential* of X is defined as the unique solution Z of the SDE

$$\mathrm{d}Z_t = Z_{t-} \, \mathrm{d}X_t, \qquad Z_0 = 1,$$

i.e.,

$$Z_t = 1 + \int_0^t Z_{s-} \, \mathrm{d}X_s \qquad \text{for all } t \geq 0, \; P\text{-a.s.},$$

and it is denoted by $\mathcal{E}(X) := Z$.

From the preceding example, we have the explicit formula $\mathcal{E}(X) = \exp(X - \frac{1}{2}\langle X \rangle)$ when X is continuous and null at 0. For general X, an explicit formula is given in Protter [13, Theorem II.37]. Note that $Z = \mathcal{E}(X)$ can become 0 or negative when X has jumps; in fact, the properties of jumps of stochastic integrals yield

$$Z_t - Z_{t-} = \Delta Z_t = \Delta\left(1 + \int Z_{s-} \, \mathrm{d}X_s\right) = Z_{t-} \Delta X_t,$$

and this shows that $Z_t = Z_{t-}(1 + \Delta X_t)$ so that $Z = \mathcal{E}(X)$ changes sign between $t-$ and t whenever $1 + \Delta X_t < 0$, i.e. when X has a jump $\Delta X_t < -1$.

Example 1.4. Suppose W is a Brownian motion, $T \in (0, \infty)$ is fixed and $h : \mathbb{R} \to \mathbb{R}$ is a measurable function with $h(W_T) \in L^1$. Then clearly

$$M_t := E[h(W_T) \,|\, \mathcal{F}_t] \qquad \text{for } 0 \leq t \leq T$$

is a martingale. But writing

$$M_t = E[h(W_t + W_T - W_t) \,|\, \mathcal{F}_t]$$

and using that W_t is $\mathcal{F}_t$-measurable and $W_T - W_t$ is independent of $\mathcal{F}_t$ and $\sim \mathcal{N}(0, T-t)$ shows that we also have

$$M_t = E[h(x + W_T - W_t)]\big|_{x=W_t} = f(W_t, t)$$

with

$$f(x,t) = E[h(x + W_T - W_t)] = \int_{-\infty}^{\infty} h(x+y) \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{y^2}{2(T-t)}} \, \mathrm{d}y.$$

So $f(\,\cdot\,, t)$, as a function of x for fixed $t < T$, is the convolution of h with a function in C^∞ and therefore also C^∞ with respect to x, and $f(x, \,\cdot\,)$ is clearly in C^1 with respect to t as long as $t < T$. Therefore Itô's formula may be applied and gives

$$M_t = M_0 + \int_0^t f_x(W_s, s) \, \mathrm{d}W_s + \int_0^t \left(f_t + \frac{1}{2} f_{xx} \right)(W_s, s) \, \mathrm{d}s \qquad \text{for } 0 \le t < T. \tag{1.12}$$

Now one can check by laborious analysis that the function $f(x,t)$ satisfies the partial differential equation (PDE) $f_t + \frac{1}{2} f_{xx} = 0$; or one can use the fact that the canonical decomposition of a special semimartingale (like the martingale M) is unique. (Alternatively, one can use that any continuous local martingale of finite variation is constant.) Any of these leads to the conclusion that the $\mathrm{d}s$-integral in (1.12) must vanish identically because it is continuous and adapted, hence predictable, and of finite variation like any $\mathrm{d}s$-integral. By letting $t \nearrow T$ in (1.12), we therefore obtain the representation

$$h(W_T) = M_T = M_0 + \int_0^T f_x(W_s, s) \, \mathrm{d}W_s$$

of the random variable $h(W_T)$ as an initial value M_0 plus a stochastic integral with respect to the Brownian motion W. A more general result in that direction is given in Section 6.3.

Example. An *Itô process* is a stochastic process of the form

$$X_t = X_0 + \int_0^t \mu_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s \qquad \text{for } t \ge 0$$

for some Brownian motion W, where μ and σ are predictable processes satisfying appropriate integrability conditions (e.g. $\int_0^T (|\mu_s| + |\sigma_s|^2) \, \mathrm{d}s < \infty$ P-a.s. for every $T < \infty$).

More generally, X, μ, W could be vector-valued and σ could be matrix-valued, of course all with appropriate dimensions. For any C^2-function f, the process $f(X)$ is then again an Itô process, and Itô's formula gives

$$f(X_t) = f(X_0) + \int_0^t \left(f'(X_s)\mu_s + \frac{1}{2} f''(X_s)\sigma_s^2 \right) \mathrm{d}s + \int_0^t f'(X_s)\sigma_s \, \mathrm{d}W_s.$$

This is another good [→ *exercise*] for using Itô's formula.

Example. For any two real-valued (RCLL) semimartingales X and Y, the *product rule* is obtained by applying Itô's formula with the function $f(x, y) = xy$. The result says that

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_{s-} \, \mathrm{d}X_s + \int_0^t X_{s-} \, \mathrm{d}Y_s + [X, Y]_t$$

or compactly in differential notation

$$\mathrm{d}(XY) = Y_- \, \mathrm{d}X + X_- \, \mathrm{d}Y + \mathrm{d}[X, Y].$$

If both X and Y are continuous, this yields

$$\mathrm{d}(XY) = Y \, \mathrm{d}X + X \, \mathrm{d}Y + \mathrm{d}\langle X, Y \rangle.$$

Example. Let $W = (W_t)_{t \geq 0}$ be a Brownian motion, $a < 0 < b$ and

$$\tau_{a,b} := \inf \left\{ t \geq 0 : W_t > b \text{ or } W_t < a \right\}$$

the first time that BM leaves the interval $[a, b]$ around 0. Then classical results about the *ruin problem for Brownian motion* say that

$$E[\tau_{a,b}] = |a| \, b \qquad (\text{so that } \tau_{a,b} < \infty \ P\text{-a.s.})$$

and

$$P[W_{\tau_{a,b}} = b] = \frac{|a|}{b - a} = 1 - P[W_{\tau_{a,b}} = a]. \tag{1.13}$$

It is also known, or can be computed from (1.13), that $E[W_{\tau_{a,b}}] = 0$.

In order to compute the *covariance* of $\tau_{a,b}$ and $W_{\tau_{a,b}}$, we start with the function $f(x,t) = -\frac{1}{3}x^3 + tx$. Then clearly $f_t + \frac{1}{2}f_{xx} \equiv 0$ so that Itô's formula shows that

$$M_t := f(W_t, t) = 0 + \int_0^t f_x(W_s, s)\,\mathrm{d}W_s$$

is like W a continuous local martingale, and so is then the stopped process $M^{\tau_{a,b}}$. But

$$M_t^{\tau_{a,b}} = M_{t\wedge\tau_{a,b}} = -\frac{1}{3}\big(W_t^{\tau_{a,b}}\big)^3 + (t \wedge \tau_{a,b})W_t^{\tau_{a,b}}$$

is bounded by a constant for $t \le T$ as $|W^{\tau_{a,b}}| \le \max(|a|, b)$, and so $M^{\tau_{a,b}}$ is a martingale on $[0,T]$ for each $T < \infty$. This directly implies that

$$0 = E\big[M_0^{\tau_{a,b}}\big] = E\big[M_T^{\tau_{a,b}}\big] = -\frac{1}{3}E\big[W^3_{\tau_{a,b}\wedge T}\big] + E\big[(\tau_{a,b}\wedge T)W_{\tau_{a,b}\wedge T}\big],$$

and letting $T \to \infty$ yields by dominated convergence, also using $\tau_{a,b} \in L^1$, that

$$0 = -\frac{1}{3}E\big[W^3_{\tau_{a,b}}\big] + E\big[\tau_{a,b}W_{\tau_{a,b}}\big].$$

Hence we find

$$\mathrm{Cov}\big(\tau_{a,b}, W_{\tau_{a,b}}\big) = E\big[\tau_{a,b}W_{\tau_{a,b}}\big] = \frac{1}{3}E\big[W^3_{\tau_{a,b}}\big] = \frac{1}{3}|a|b(b-|a|),$$

where the last equality is obtained by computing with the known (two-point) distribution of $W_{\tau_{a,b}}$ given in (1.13).

6.2 Girsanov's theorem

In Section 6.1, we have seen that the family of semimartingales is *invariant* under a transformation by a C^2-function, i.e., $f(X)$ is a semimartingale whenever X is a semimartingale and $f \in C^2$. In this section, our goal is to show that the class of semimartingales is also invariant under a change to an equivalent probability measure.

Suppose we have P and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. Assuming that $Q \approx P$ on $\mathcal{F}$ (or $\mathcal{F}_\infty$) can be too restrictive; so we fix $T \in (0,\infty)$ and assume only that $Q \approx P$ on $\mathcal{F}_T$. If we have this for every $T < \infty$, we call Q and P *locally equivalent* and write $Q \overset{\text{loc}}{\approx} P$. For an infinite horizon, this is usually strictly weaker than $Q \approx P$. (Also, one must be careful with the filtration and the usual conditions, but we do not discuss these technical issues.)

To start, fix $T \in (0,\infty)$ for simplicity and suppose that $Q \approx P$ on $\mathcal{F}_T$. Denote by

$$Z_t := Z_t^{Q;P} := E_P\left[\frac{\mathrm{d}Q|_{\mathcal{F}_T}}{\mathrm{d}P|_{\mathcal{F}_T}}\,\middle|\,\mathcal{F}_t\right] \qquad \text{for } 0 \leq t \leq T \tag{2.1}$$

the *density process* of Q with respect to P on $[0,T]$, choosing an RCLL version of this P-martingale on $[0,T]$. Because $Q \approx P$ on $\mathcal{F}_T$, we have $Z > 0$ on $[0,T]$, meaning that $P[Z_t > 0, \forall t \in [0,T]] = 1$, and because Z is a P-(super)martingale, we even have $\inf_{0\leq t\leq T} Z_t > 0$ P-a.s. by the so-called *minimum principle for supermartingales*; see Dellacherie/Meyer [5, Theorem VI.17]. This implies that also $Z_- > 0$ on $[0,T]$ so that $1/Z_-$ is well defined and adapted and left-continuous, hence also predictable and locally bounded.

In perfect analogy to Lemma 2.3.1, we now have

Lemma 2.1. *Suppose that $Q \approx P$ on $\mathcal{F}_T$ and define $Z = Z^{Q;P}$ as in* (2.1)*. Then:*

1) *For $s \leq t \leq T$ and every U_t which is $\mathcal{F}_t$-measurable and either ≥ 0 or in $L^1(Q)$, we have the* Bayes formula

$$E_Q[U_t \,|\, \mathcal{F}_s] = \frac{1}{Z_s} E_P[Z_t U_t \,|\, \mathcal{F}_s] \qquad Q\text{-a.s.}$$

2) *An adapted process $Y = (Y_t)_{0\leq t\leq T}$ is a (local) Q-martingale on $[0,T]$ if and only if the product ZY is a (local) P-martingale on $[0,T]$.*

Of course, if $Q \overset{\text{loc}}{\approx} P$, we can use Lemma 2.1 for any $T < \infty$ and hence obtain a statement for processes $Y = (Y_t)_{t\geq 0}$ on $[0, \infty)$. One consequence of part 2) of Lemma 2.1 (with $Y := 1/Z$) is also that $\frac{1}{Z}$ is a Q-martingale, more precisely on $[0, T]$ if $Q \approx P$ on $\mathcal{F}_T$, or even on $[0, \infty)$ if $Q \overset{\text{loc}}{\approx} P$. Furthermore, it is easy to check that $\frac{1}{Z}$ is the density process of P with respect to Q (again on $[0, T]$ or on $[0, \infty)$, respectively).

The next result now proves the announced basic result.

Theorem 2.2 (**Girsanov**). *Suppose that $Q \overset{\text{loc}}{\approx} P$ with density process Z. If M is a local P-martingale null at 0, then*

$$\widetilde{M} := M - \int \frac{1}{Z} \, \mathrm{d}[Z, M]$$

is a local Q-martingale null at 0. In particular, every P-semimartingale is also a Q-semimartingale (and vice versa, by symmetry).

Proof. The second assertion is very easy to prove from the first; we simply write

$$X = X_0 + M + A = X_0 + \widetilde{M} + \left(A + \int \frac{1}{Z} \, \mathrm{d}[Z, M] \right) = X_0 + \widetilde{M} + \widetilde{A}$$

and observe that $\widetilde{A} := A + \int \frac{1}{Z} \, \mathrm{d}[Z, M]$ is of finite variation. Note that $\int \frac{1}{Z} \, \mathrm{d}[Z, M]$ is defined pathwise because $[Z, M]$ is of finite variation; so this requires no stochastic integration, nor predictability of the integrand.

For proving the first assertion, note that the definition of the optional covariation process implies that the difference $ZM - [Z, M]$ is a local P-martingale like M and Z. (To argue this in an alternative manner, we could use the product rule which gives $ZM - [Z, M] = \int Z_- \, \mathrm{d}M + \int M_- \, \mathrm{d}Z$, which is a local P-martingale like M and Z.) So by Lemma 2.1,

$$M - \frac{1}{Z}[Z, M] \text{ is a local } Q\text{-martingale.}$$

Using the product rule gives

$$\frac{1}{Z}[Z,M] = \int [Z,M]_- \,\mathrm{d}\left(\frac{1}{Z}\right) + \int \frac{1}{Z_-}\,\mathrm{d}[Z,M] + \left[\frac{1}{Z},[Z,M]\right]. \tag{2.2}$$

Because $[Z,M]$ is of finite variation, the last term equals

$$\left[\frac{1}{Z},[Z,M]\right] = \sum \Delta\left(\frac{1}{Z}\right)\Delta[Z,M] = \int \Delta\left(\frac{1}{Z}\right)\,\mathrm{d}[Z,M]$$

so that the last two terms in (2.2) add up to $\int \frac{1}{Z}\,\mathrm{d}[Z,M]$. Because $\frac{1}{Z}$ is a local Q-martingale, so is the stochastic integral $\int [Z,M]_-\,\mathrm{d}\left(\frac{1}{Z}\right)$ because its integrand is locally bounded. So we obtain

$$\widetilde{M} = M - \int \frac{1}{Z}\,\mathrm{d}[Z,M] = \left(M - \frac{1}{Z}[Z,M]\right) - \int [Z,M]_-\,\mathrm{d}\left(\frac{1}{Z}\right),$$

and we see that this is a local Q-martingale. **q.e.d.**

In many situations, it is more convenient to do computations not in terms of Z, but rather with its so-called stochastic logarithm. Suppose in general that Y is a semimartingale with $Y_- > 0$ (on $[0,T]$ or $[0,\infty)$, respectively). Then we can define a semimartingale null at 0 by $L := \int \frac{1}{Y_-}\,\mathrm{d}Y$, we have $\mathrm{d}Y = Y_-\,\mathrm{d}L$ by construction, and so we obtain

$$Y = Y_0\,\mathcal{E}(L) > 0 \qquad \text{with a semimartingale } L \text{ null at } 0.$$

It is also clear that L is continuous if and only if Y is continuous, and that L is a local P-martingale if and only if Y is a local P-martingale. This L is called the *stochastic logarithm* of Y. Note that because of the quadratic variation, we do not have $L = \log Y$, not even if Y is continuous; see the explicit formula (1.11) in Section 6.1.

In the situation here, Z is a P-martingale > 0, hence has $Z_- > 0$ as discussed above, and so applying the above with $Y := Z$ yields $Z = Z_0\,\mathcal{E}(L)$, where L is like Z a local P-martingale.

Theorem 2.3 (**Girsanov, continuous version**). *Suppose that* $Q \overset{\text{loc}}{\approx} P$ *with a density process* Z *which is* continuous. *Write* $Z = Z_0\,\mathcal{E}(L)$. *If* M *is a local* P*-martingale null at*

0, *then*

$$\widetilde{M} := M - [L, M] = M - \langle L, M \rangle$$

is a local Q-martingale null at 0.

More specifically, if W is a P-Brownian motion, then $\widetilde{W}$ is a Q-Brownian motion. In particular, if $L = \int \nu \, \mathrm{d}W$ for some $\nu \in L^2_{\mathrm{loc}}(W)$, then

$$\widetilde{W} = W - \left\langle \int \nu \, \mathrm{d}W, W \right\rangle = W - \int \nu_s \, \mathrm{d}s$$

so that the P-Brownian motion $W = \widetilde{W} + \int \nu_s \, \mathrm{d}s$ becomes under Q a Brownian motion with (instantaneous) drift ν.

Proof. Because $Z = Z_0 \, \mathcal{E}(L)$ satisfies $\mathrm{d}Z = Z_- \, \mathrm{d}L$, we have $[Z, M] = \int Z_- \, \mathrm{d}[L, M]$ and hence $\int \frac{1}{Z} \, \mathrm{d}[Z, M] = \int \frac{Z_-}{Z} \, \mathrm{d}[L, M] = [L, M]$ by continuity of Z. So the first assertion follows directly from Theorem 2.2, and $[L, M] = \langle L, M \rangle$ because L is continuous like Z.

The assertion for $\widetilde{W}$ needs some extra work as it relies on the so-called *Lévy characterisation of Brownian motion* that we have not discussed here. **q.e.d.**

In all the above discussions, we have assumed that Q is already given and have then studied its effect on given processes. But in mathematical finance, we often want to proceed the other way round: We start with a process $S = (S_t)_{0 \le t \le T}$ of discounted asset prices and want to find or construct some $Q \approx P$ on $\mathcal{F}_T$ such that S becomes a local Q-martingale. Let us now see how we can tackle this problem by reverse-engineering the preceding theory. We begin very generally and successively become more specific. Moreover, the goal here is not to remember a specific result, but rather to understand how to approach the problem in a systematic way.

We start with a local P-martingale L null at 0 and define $Z := \mathcal{E}(L)$ so that Z is like L a local P-martingale, with $Z_0 = 1$. If we also have $\Delta L > -1$ (and this holds of course in particular if L is continuous), then we have in addition $Z > 0$. This uses that $\Delta Z = Z_- \Delta L$ so that $Z = Z_-(1 + \Delta L)$, which implies that Z never changes sign as long as $\Delta L > -1$.

Suppose now that Z is a *true* P-martingale on $[0,T]$; this amounts to imposing suitable extra conditions on L. Then we can define a probability measure $Q \approx P$ on $\mathcal{F}_T$ by setting $\mathrm{d}Q := Z_T \,\mathrm{d}P$, and the density process of Q with respect to P on $[0,T]$ is then by construction the P-martingale Z. In particular, if L is continuous, also Z is continuous.

In a bit more detail, $Z = \mathcal{E}(L)$ is in the present situation a local P-martingale > 0 on $[0,T]$ and therefore a P-supermartingale starting at 1. So $t \mapsto E[Z_t]$ is decreasing, and one can easily check that Z is a P-martingale on $[0,T]$ if and only if $t \mapsto E[Z_t]$ is identically 1 on $[0,T]$, or also if and only if $E[Z_T] = 1$. However, expressing this directly in terms of L is more tricky, and one has only sufficient conditions on L that ensure $E[\mathcal{E}(L)_T] = 1$. The most famous of these is the *Novikov condition*: If L is a continuous local martingale null at 0 and $E[e^{\frac{1}{2}\langle L\rangle_T}] < \infty$, then $Z = \mathcal{E}(L)$ is a martingale on $[0,T]$.

Now start with an $\mathbb{R}^d$-valued process $S = (S_t)_{0\le t\le T}$ and suppose that S is a P-semimartingale. For each i, the coordinate S^i can then (in general non-uniquely) be written as

$$S^i = S^i_0 + M^i + A^i$$

with a local P-martingale M^i and an adapted process A^i of finite variation, both null at 0. By Theorem 2.2,

$$\widetilde{M}^i = M^i - \int \frac{1}{Z}\,\mathrm{d}[Z, M^i]$$

is then a local Q-martingale, and of course we have

$$S^i = S^i_0 + \widetilde{M}^i + \left(A^i + \int \frac{1}{Z}\,\mathrm{d}[Z, M^i]\right) = S^i_0 + \widetilde{M}^i + \widetilde{A}^i.$$

So S^i is a local Q-martingale (or, equivalently, Q is an ELMM for S^i) if and only if

$$\widetilde{A}^i = A^i + \int \frac{1}{Z}\,\mathrm{d}[Z, M^i] \quad \text{ is a local } Q\text{-martingale.}$$

One sufficient condition is obviously that

$$A^i + \int \frac{1}{Z}\,\mathrm{d}[Z, M^i] \equiv 0. \tag{2.3}$$

This should be viewed as a condition on Z or, equivalently, on L. In general, because $\mathrm{d}Z = Z_- \,\mathrm{d}L$, we have

$$[Z, M^i] = \int Z_- \,\mathrm{d}[L, M^i]$$

and $\Delta Z = Z_- \Delta L$, hence

$$Z = Z_- + \Delta Z = Z_-(1 + \Delta L)$$

and so

$$\frac{Z_-}{Z} = \frac{1}{1 + \Delta L}.$$

So in terms of L, the sufficient condition (2.3) can be written as

$$A^i + \int \frac{1}{1 + \Delta L} \,\mathrm{d}[L, M^i] \equiv 0.$$

If L is continuous, this simplifies further to

$$A^i + \langle L, M^i \rangle \equiv 0;$$

this could alternatively also be derived directly from Theorem 2.3. As a condition on L in terms of M and A, this is fairly explicit. Note that this is actually a system of d conditions (one for each S^i) imposed on a single process L.

In Chapter 7, we shall see how the above ideas can be used to construct explicitly an equivalent martingale measure in the Black–Scholes model of geometric Brownian motion for S. But before that, we study in the next section how local martingales L can (or must) look if we impose more structure on the underlying filtration $\mathbb{F}$.

Remark. Instead of using Theorem 2.2, we could also argue more directly. Suppose again that $Z = \mathcal{E}(L)$ is a true P-martingale > 0 on $[0, T]$, and define $Q \approx P$ on $\mathcal{F}_T$ by $\mathrm{d}Q := Z_T \,\mathrm{d}P$. By Lemma 2.1, S is then a local Q-martingale if and only if ZS is a local P-martingale, and therefore we compute, using the product rule and $\mathrm{d}Z = Z_- \,\mathrm{d}L$,

$$\mathrm{d}(ZS^i) = S^i_- \,\mathrm{d}Z + Z_- \,\mathrm{d}S^i + \mathrm{d}[Z, S^i] = S^i_- \,\mathrm{d}Z + Z_- \,\mathrm{d}M^i + Z_-(\mathrm{d}A^i + \mathrm{d}[L, S^i]).$$

Because both Z and M^i, and hence also their stochastic integrals above, are local P-martingales, we see that Q is an ELMM for S^i if and only if $A^i + [L, S^i]$ is a local P-martingale. A sufficient condition for this is that

$$A^i + [L, S^i] \equiv 0.$$

If L is continuous or if S^i is continuous, this again simplifies to

$$A^i + \langle L, M^i \rangle \equiv 0,$$

because then $[L, A^i] = \sum \Delta L \, \Delta A^i \equiv 0$. $\diamond$

6.3 Itô's representation theorem

Our goal in this section is to describe all martingales that can exist in a filtration $\mathbb{F}$ under the assumption that $\mathbb{F}$ is *generated by a Brownian motion* W. This deep structural result goes back to Kiyosi Itô and is the mathematical explanation for the completeness of the Black–Scholes model that we shall see in the next chapter.

We start with a Brownian motion $W = (W_t)_{t\geq 0}$ in $\mathbb{R}^m$ defined on a probability space $(\Omega, \mathcal{F}, P)$ without an a priori filtration. We define

$$\mathcal{F}_t^0 := \sigma(W_s, s \leq t) \qquad \text{for } t \geq 0,$$
$$\mathcal{F}_\infty^0 := \sigma(W_s, s \geq 0),$$

and construct the filtration $\mathbb{F}^W = (\mathcal{F}_t^W)_{0\leq t\leq\infty}$ by adding to each $\mathcal{F}_t^0$ the class $\mathcal{N}$ of all subsets of P-nullsets in $\mathcal{F}_\infty^0$ to obtain $\mathcal{F}_t^W = \mathcal{F}_t^0 \vee \mathcal{N}$. This so-called *$P$-augmented filtration* $\mathbb{F}^W$ is then P-complete (in $(\Omega, \mathcal{F}_\infty^0, P)$, to be accurate) by construction, and one can show, by using the strong Markov property of Brownian motion, that $\mathbb{F}^W$ is also automatically right-continuous (so that it satisfies the usual conditions). We usually call $\mathbb{F}^W$, slightly misleadingly, the filtration generated by W. One can show that W is also a Brownian motion with respect to $\mathbb{F}^W$; the key point is to argue that $W_t - W_s$ is still independent of $\mathcal{F}_s^W \supseteq \mathcal{F}_s^0$, even though $\mathcal{F}_s^W$ contains some sets from $\mathcal{F}_\infty^0$. If one works on $[0, T]$, one replaces ∞ by T; then $\mathcal{F}_\infty^0$ is not needed separately because we use the P-nullsets from the "last" σ-field $\mathcal{F}_T^0$.

Theorem 3.1 (Itô's representation theorem). *Suppose that $W = (W_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^m$. Then every random variable $H \in L^1(\mathcal{F}_\infty^W, P)$ has a unique representation as*

$$H = E[H] + \int_0^\infty \psi_s \, \mathrm{d}W_s \qquad P\text{-a.s.}$$

for an $\mathbb{R}^m$-valued integrand $\psi \in L^2_{\mathrm{loc}}(W)$ with the additional property that $\int \psi \, \mathrm{d}W$ is a $(P, \mathbb{F}^W)$-martingale on the closed *interval $[0, \infty]$ (and therefore uniformly integrable).*

Remark. The assumptions on H say that H is integrable and $\mathcal{F}_\infty^W$-measurable. The latter

means intuitively that $H(\omega)$ can depend in a measurable way on the entire trajectory $W_\bullet(\omega)$ of Brownian motion, but not on any other source of randomness. $\diamond$

Corollary 3.2. *Suppose the filtration $\mathbb{F} = \mathbb{F}^W$ is generated by a Brownian motion W in $\mathbb{R}^m$. Then:*

1) *Every (real-valued) local $(P, \mathbb{F}^W)$-martingale L is of the form $L = L_0 + \int \gamma \, \mathrm{d}W$ for some $\mathbb{R}^m$-valued process $\gamma \in L^2_{\mathrm{loc}}(W)$.*

2) *Every local $(P, \mathbb{F}^W)$-martingale is continuous.*

Proof. For a localizing sequence $(\tau_k)_{k\in\mathbb{N}}$, each $(L - L_0)^{\tau_k}$ is a uniformly integrable martingale N^k, say, and therefore of the form

$$N^k_t = E[N^k_\infty \,|\, \mathcal{F}^W_t] \qquad \text{for } 0 \le t \le \infty,$$

for some $N^k_\infty \in L^1(\mathcal{F}^W_\infty, P)$. So Theorem 3.1 and the martingale property of $\int \psi^k \, \mathrm{d}W$ give that $N^k = \int \psi^k \, \mathrm{d}W$ (note that $N^k_0 = 0$). In particular, $N^k = (L - L_0)^{\tau_k}$ is continuous, which means that L is continuous on $[\![0, \tau_k]\!]$. As $\tau_k \nearrow \infty$, L is continuous, and γ is obtained by piecing together the ψ^k via $\gamma := \psi^k$ on $[\![0, \tau_k]\!]$. **q.e.d.**

While the above results are remarkable, the next result is bizarre. Note that in its formulation, the filtration $\mathbb{F}$ is even allowed to be general; but of course we could also take $\mathbb{F} = \mathbb{F}^W$.

Theorem 3.3 (Dudley). *Suppose $W = (W_t)_{t\ge 0}$ is a Brownian motion with respect to P and $\mathbb{F} = (\mathcal{F}_t)_{t\ge 0}$. As usual, set*

$$\mathcal{F}_\infty := \bigvee_{t\ge 0} \mathcal{F}_t = \sigma\Big(\bigcup_{t\ge 0} \mathcal{F}_t\Big).$$

Then every $\mathcal{F}_\infty$-measurable random variable H with $|H| < \infty$ P-a.s. (for example every $H \in L^1(\mathcal{F}_\infty, P)$) can be written as

$$H = \int_0^\infty \psi_s \, \mathrm{d}W_s \qquad P\text{-a.s.}$$

for some integrand $\psi \in L^2_{\mathrm{loc}}(W)$.

Note that there is no constant in the representation of H in Theorem 3.3. Note also that we could for instance take for H a constant and represent this as a stochastic integral of Brownian motion. This makes it almost immediately clear that the integrand ψ in Theorem 3.3 cannot be nice. In fact:

1) In Theorem 3.3, the stochastic integral process $\int \psi \, \mathrm{d}W$ is of course a *local martingale*, and can even be a martingale on $[0, \infty)$, but it is in general *not a martingale on* $[0, \infty]$; if it were, it would have constant expectation 0 up to $+\infty$, which would imply that $E[H] = 0$.

2) In Theorem 3.3, the representation by ψ is *not unique*. In fact, one can easily construct some bounded predictable $\bar{\psi}$ with $0 < \int_0^\infty \bar{\psi}_s^2 \, \mathrm{d}s < \infty$ P-a.s. (so that $\bar{\psi} \not\equiv 0$ and $\bar{\psi} \in L^2_{\mathrm{loc}}(W)$), but nevertheless $\int_0^\infty \bar{\psi}_s \, \mathrm{d}W_s = 0$ P-a.s. Of course, ψ and $\psi + \bar{\psi}$ then represent the same H, but they are different in a nontrivial way.
[*Exercise:* Try to find such a $\bar{\psi}$ — it is not very difficult.]

3) In terms of finance, the integrands ψ appearing in Theorem 3.3 are not nice at all. For one thing, $\int \psi \, \mathrm{d}W$ cannot be bounded from below in general. Indeed, if it were, then $\int \psi \, \mathrm{d}W$ would be a local martingale uniformly bounded from below, hence a supermartingale, and this would imply that we must have $E[H] \leq 0$. Moreover, the representation $1 = \int_0^\infty \psi_s \, \mathrm{d}W_s$ looks suspiciously like creating the riskless payoff 1 out of zero initial capital with a self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \psi)$, which would be arbitrage. (But of course, that φ is not admissible, as we have just argued.)

Remark. It is not important for the above results that we work on the infinite interval $[0, \infty]$ or $[0, \infty)$; everything could be done equally well on $[0, T]$ for any $T \in (0, \infty)$. $\diamond$

7 The Black–Scholes formula

Our goal in this final chapter is to combine the modelling and financial ideas from the discrete-time setting with the continuous-time techniques from stochastic calculus. We introduce and study a simple continuous-time financial market model and show how this allows us to derive the celebrated Black–Scholes formula together with the underlying methodology. We emphasise that the latter is much more important than the formula itself, for obvious reasons.

7.1 The Black–Scholes model

The *Black–Scholes model* or *Samuelson model* is the continuous-time analogue of the Cox–Ross–Rubinstein binomial model we have seen at length in earlier chapters. Like the latter, it is too simple to be realistic, but still very popular because it allows many explicit calculations and results. It also serves as a basic starting point or reference model.

To set up the model, we start with a fixed time horizon $T \in (0, \infty)$ and a probability space $(\Omega, \mathcal{F}, P)$ on which there is a Brownian motion $W = (W_t)_{0 \le t \le T}$. We take as *filtration* $\mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}$ the one generated by W and augmented as in Section 6.3 by the P-nullsets from $\mathcal{F}_T^0 := \sigma(W_s, s \le T)$ so that $\mathbb{F} = \mathbb{F}^W$ satisfies the usual conditions under P. We shall see soon that this choice of filtration is important.

The *financial market model* has two basic traded assets: a *bank account* with constant continuously compounded *interest rate* $r \in \mathbb{R}$, and a *risky asset* (usually called *stock*) having two parameters $\mu \in \mathbb{R}$ and $\sigma > 0$. Undiscounted prices are given by

$$\widetilde{S}_t^0 = e^{rt}, \tag{1.1}$$

$$\widetilde{S}_t^1 = S_0^1 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right) \tag{1.2}$$

with a constant $S_0^1 > 0$. Applying Itô's formula easily yields

$$\mathrm{d}\widetilde{S}_t^0 = \widetilde{S}_t^0 r \,\mathrm{d}t, \tag{1.3}$$

$$\mathrm{d}\widetilde{S}_t^1 = \widetilde{S}_t^1 \mu \,\mathrm{d}t + \widetilde{S}_t^1 \sigma \,\mathrm{d}W_t, \tag{1.4}$$

which can be rewritten as

$$\frac{\mathrm{d}\widetilde{S}^0_t}{\widetilde{S}^0_t} = r\,\mathrm{d}t, \tag{1.5}$$

$$\frac{\mathrm{d}\widetilde{S}^1_t}{\widetilde{S}^1_t} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{1.6}$$

This means that the bank account has a *relative price change* $(\widetilde{S}^0_t - \widetilde{S}^0_{t-\mathrm{d}t})/\widetilde{S}^0_{t-\mathrm{d}t}$ over a short time period $(t-\mathrm{d}t, t]$ of $r\,\mathrm{d}t$; so r is the growth rate of the bank account. In the same way, the relative price change of the stock has a part $\mu\,\mathrm{d}t$ giving a growth at rate μ, and a second part $\sigma\,\mathrm{d}W_t$ "with mean 0 and variance $\sigma^2\,\mathrm{d}t$" that causes random fluctuations. We call μ the *drift* (rate) and σ the (instantaneous) *volatility* of $\widetilde{S}^1$. The formulation (1.5), (1.6) also makes it clear why this model is the continuous-time analogue of the CRR binomial model; see Example 6.1.3 in Section 6.1 for a more detailed discussion. (Because $\widetilde{S}^0$ and $\widetilde{S}^1$ are both continuous, we can replace $\widetilde{S}^0_{t-\mathrm{d}t}$ and $\widetilde{S}^1_{t-\mathrm{d}t}$ in the denominators above by $\widetilde{S}^0_t$ and $\widetilde{S}^1_t$, respectively.)

As usual, we pass to quantities *discounted* with $\widetilde{S}^0$; so we have $S^0 = \widetilde{S}^0/\widetilde{S}^0 \equiv 1$, and $S^1 = \widetilde{S}^1/\widetilde{S}^0$ is by (1.1) and (1.2) given by

$$S^1_t = S^1_0 \exp\left(\sigma W_t + \left(\mu - r - \frac{1}{2}\sigma^2\right)t\right). \tag{1.7}$$

Either from (1.7) or from (1.3), (1.4), we obtain via Itô's formula that S^1 solves the SDE

$$\mathrm{d}S^1_t = S^1_t\big((\mu - r)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\big). \tag{1.8}$$

For later use, we observe that this gives

$$\mathrm{d}\langle S^1\rangle_t = (S^1_t)^2\sigma^2\,\mathrm{d}\langle W\rangle_t = (S^1_t)^2\sigma^2\,\mathrm{d}t \tag{1.9}$$

for the *quadratic variation* of S^1, because $\langle W\rangle_t = t$.

Remark 1.1. Because the cofficients μ, r, σ are all constant and $\sigma > 0$, the undiscounted prices $(\widetilde{S}^0, \widetilde{S}^1)$, the discounted prices (S^0, S^1), the discounted stock price S^1 alone, and the Brownian motion W all generate the same filtration. This means that there is here

no compromise between mathematical convenience (the filtration $\mathbb{F}$ is generated by W) and financial modelling (the filtration is generated by information about prices). $\diamond$

As in discrete time, we should like to have an *equivalent martingale measure* for the discounted stock price process S^1. To get an idea how to find this, we rewrite (1.8) as

$$\mathrm{d}S_t^1 = S_t^1 \sigma \left(\mathrm{d}W_t + \frac{\mu - r}{\sigma} \, \mathrm{d}t \right) = S_t^1 \sigma \, \mathrm{d}W_t^*, \tag{1.10}$$

with $W^* = (W_t^*)_{0 \le t \le T}$ defined by

$$W_t^* := W_t + \frac{\mu - r}{\sigma} t = W_t + \int_0^t \lambda \, \mathrm{d}s \qquad \text{for } 0 \le t \le T.$$

The quantity

$$\lambda := \frac{\mu - r}{\sigma}$$

is often called the instantaneous *market price of risk* or infinitesimal *Sharpe ratio* of S^1. By looking at Girsanov's theorem in the form of Theorem 6.2.3, we see that W^* is a Brownian motion on $[0, T]$ under the probability measure Q^* given by

$$\frac{\mathrm{d}Q^*}{\mathrm{d}P} := \mathcal{E}\left(-\int \lambda \, \mathrm{d}W \right)_T = \exp\left(-\lambda W_T - \frac{1}{2} \lambda^2 T \right) \qquad \text{on } \mathcal{F}_T,$$

whose density process with respect to P is

$$Z_t^{Q^*;P} = Z_t^* = \mathcal{E}\left(-\int \lambda \, \mathrm{d}W \right)_t = \exp\left(-\lambda W_t - \frac{1}{2} \lambda^2 t \right) \qquad \text{for } 0 \le t \le T.$$

By (1.10), the stochastic integral process

$$S_t^1 = S_0^1 + \int_0^t S_u^1 \sigma \, \mathrm{d}W_u^*$$

is then a continuous local Q^*-martingale like W^*; it is even a Q^*-martingale because we have the explicit expression

$$S_t^1 = S_1^0 \, \mathcal{E}(\sigma W^*)_t = S_0^1 \exp\left(\sigma W_t^* - \frac{1}{2} \sigma^2 t \right) \tag{1.11}$$

from (1.10) by Itô's formula, and so we can use Proposition 4.2.3 under Q^*.

All in all, then, S^1 admits an equivalent martingale measure, explicitly given by Q^*, and so we expect that S^1 should be "arbitrage-free" in any reasonable sense. However, we cannot make this precise here before defining more carefully what "trading strategy", "self-financing", "arbitrage opportunity" etc. should mean in this context.

Remark. Suppose Q is any probability measure equivalent to P on $\mathcal{F}_T$ and denote its P-density process by $Z^{Q;P} = Z = (Z_t)_{0\le t\le T}$. Then we can write $Z = Z_0\,\mathcal{E}(L)$ as in Section 6.2, where L is a local $(P, \mathbb{F})$-martingale null at 0. But $\mathbb{F}$ is generated by W; so Itô's representation theorem in Corollary 6.3.2 says that

$$L = \int \nu_s \,\mathrm{d}W_s \qquad \text{for some } \nu \in L^2_{\mathrm{loc}}(W)$$

and therefore $\mathrm{d}Z_t = Z_{t-}\,\mathrm{d}L_t = Z_t\nu_t\,\mathrm{d}W_t$ (as Z is automatically continuous like L).

Now suppose in addition that S^1 is a local Q-martingale, i.e. Q is an ELMM for S^1. By the Bayes rule in Lemma 6.2.1, this implies that ZS^1 is a local P-martingale. But the product rule, (1.8) and the rules for computing covariations of stochastic integrals give

$$\begin{aligned}\mathrm{d}(Z_tS^1_t) &= Z_t\,\mathrm{d}S^1_t + S^1_t\,\mathrm{d}Z_t + \mathrm{d}\langle Z, S^1\rangle_t\\ &= Z_tS^1_t(\mu - r)\,\mathrm{d}t + Z_tS^1_t\sigma\,\mathrm{d}W_t + S^1_tZ_t\nu_t\,\mathrm{d}W_t + Z_t\nu_tS^1_t\sigma\,\mathrm{d}\langle W, W\rangle_t\\ &= Z_tS^1_t(\sigma + \nu_t)\,\mathrm{d}W_t + Z_tS^1_t\sigma(\lambda + \nu_t)\,\mathrm{d}t,\end{aligned}$$

using that $\mu - r = \sigma\lambda$. On the left-hand side, we have by assumption a local P-martingale, and on the right-hand side, the $\mathrm{d}W$-integral is also a local P-martingale. Therefore the last term,

$$A_t := \int_0^t Z_sS^1_s\sigma(\lambda + \nu_s)\,\mathrm{d}s \qquad \text{for } 0 \le t \le T,$$

must also be a local P-martingale. But A is adapted and continuous (hence predictable) and of finite variation; so it has quadratic variation 0, hence must be constant, and so its integrand must be 0. This implies that $\nu_s \equiv -\lambda$, because Z, S^1, σ are all > 0, and therefore we get

$$Z = Z_0\,\mathcal{E}(L) = Z_0\,\mathcal{E}\left(\int \nu\,\mathrm{d}W\right) = Z_0\,\mathcal{E}\left(-\int \lambda\,\mathrm{d}W\right).$$

Finally, Z_0 has $E_P[Z_0] = E_P[Z_T] = Q[\Omega] = 1$ and is measurable with respect to $\mathcal{F}_0 = \mathcal{F}_0^W$ which is P-trivial (because W_0 is constant P-a.s.); so $Z_0 = E_P[Z_0] = 1$ and therefore

$$Z = \mathcal{E}\left(-\int \lambda \, \mathrm{d}W\right) = Z^*, \qquad \text{or } Q = Q^*.$$

Thus we have shown that in the Black–Scholes model, there is a *unique equivalent martingale measure*, which is given explicitly by Q^*. So we expect that the Black–Scholes model is not only "arbitrage-free", but also "complete" in a suitable sense. Note that the latter point (as well as the above proof of uniqueness) depends via Itô's representation theorem in a crucial way on the assumption that the filtration $\mathbb{F}$ is generated by W. $\diamond$

Now take any $H \in L^0_+(\mathcal{F}_T)$ and view H as a random *payoff* (in discounted units) due at time T. Recall that $\mathbb{F}$ is generated by W and that $W_t^* = W_t + \lambda t$, $0 \le t \le T$, is a Q^*-Brownian motion. Because λ is deterministic, W and W^* generate the same filtration, and so we can also apply Itô's representation theorem with Q^* and W^* instead of P and W. So if H is also in $L^1(Q^*)$, the Q^*-martingale $V_t^* := E_{Q^*}[H \,|\, \mathcal{F}_t]$, $0 \le t \le T$, can be represented as

$$V_t^* = E_{Q^*}[H] + \int_0^t \psi_s^H \, \mathrm{d}W_s^* \qquad \text{for } 0 \le t \le T,$$

with some unique $\psi^H \in L^2_{\mathrm{loc}}(W^*)$ such that $\int \psi^H \, \mathrm{d}W^*$ is a Q^*-martingale. Recall from (1.10) that

$$\mathrm{d}S_t^1 = S_t^1 \sigma \, \mathrm{d}W_t^*.$$

So if we define for $0 \le t \le T$

$$\vartheta_t^H := \frac{\psi_t^H}{S_t^1 \sigma},$$
$$\eta_t^H := V_t^* - \vartheta_t^H S_t^1$$

(which are both predictable because ψ^H is and S^1, V^* are both adapted and continuous), then we can interpret $\varphi^H = (\vartheta^H, \eta^H)$ as a *trading strategy* whose discounted value process is given by

$$V_t(\varphi^H) = \vartheta_t^H S_t^1 + \eta_t^H S_t^0 = V_t^* \qquad \text{for } 0 \le t \le T,$$

and which is *self-financing* in the (usual) sense that

$$V_t(\varphi^H) = V_t^* = V_0^* + \int_0^t \psi_u^H \,\mathrm{d}W_u^* = V_0(\varphi^H) + \int_0^t \vartheta_u^H \,\mathrm{d}S_u^1 \qquad \text{for } 0 \le t \le T. \tag{1.12}$$

Moreover,

$$V_T(\varphi^H) = V_T^* = H \qquad \text{a.s.}$$

shows that the strategy φ^H *replicates* H, and

$$\int \vartheta^H \,\mathrm{d}S^1 = V(\varphi^H) - V_0(\varphi^H) = V^* - E_{Q^*}[H] \ge -E_{Q^*}[H]$$

(because $V^* \ge 0$, as $H \ge 0$) shows that ϑ^H is admissible (for S^1) in the usual sense.

In summary, then, every $H \in L_+^1(\mathcal{F}_T, Q^*)$ is attainable in the sense that it can be replicated by a dynamic strategy trading in the stock and the bank account in such a way that the strategy is self-financing and admissible, and its value process is a Q^*-martingale. In that sense, we can say that the Black–Scholes model is complete. By analogous arguments as in discrete time, we then also obtain the arbitrage-free value at time t of any payoff $H \in L_+^1(\mathcal{F}_T, Q^*)$ as its conditional expectation

$$V_t^H = V_t^* = E_{Q^*}[H \,|\, \mathcal{F}_t]$$

under the unique equivalent martingale measure Q^* for S^1. This is in perfect parallel to the results we have seen for the CRR binomial model; see Section 3.3.

Remarks. 1) All the above computations and results are in *discounted* units. Of course, we could also go back to undiscounted units.

2) Itô's representation theorem gives the *existence* of a strategy, but does not tell us how it looks. To get more explicit results, additional structure (for the payoff H) and more work is needed. [→ *Exercise*]

3) The SDE (1.8) for discounted prices is

$$\frac{\mathrm{d}S_t^1}{S_t^1} = (\mu - r)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,$$

and this is rather restrictive as μ, r, σ are all constant. An obvious *extension* is to allow the coefficients μ, r, σ to be (suitably integrable) predictable processes, or possibly functionals of S or $\widetilde{S}$. This brings up several issues:

a) If μ, r, σ are specified as functionals of S, it is no longer clear whether there exists a solution of the resulting SDE. This needs a more careful and usually case-based analysis.

b) If μ, r, σ are stochastic processes that depend on extra randomness apart from W, we have to work in a larger filtration and a result like Itô's representation theorem is perhaps no longer available. Typical examples are *stochastic volatility* models where σ usually depends on a second Brownian motion as well, or *credit risk* models where the default of an asset often involves the jump of some process.

c) Even if μ, r, σ are predictable with respect to the filtration $\mathbb{F}$ generated by W, the process $W^* = W + \int \lambda_s \, \mathrm{d}s$ in general does not generate $\mathbb{F}$, but only a smaller filtration. Fortunately, there is still a representation result with respect to W^* and Q^*, but one must work a little to prove this.

4) From the point of view of finance, the *natural filtration* to work with would be the one generated by S or $\widetilde{S}$, i.e. by prices, not by W. From the explicit formulae (1.1), (1.2), one can see that $\widetilde{S}$ and W generate the same filtrations when the coefficients μ, r, σ are deterministic. (This has already been pointed out in Remark 1.1.) But in general (i.e. for more general coefficients), working with the price filtration is rather difficult because it is hard to describe.

5) A closer look at the no-arbitrage argument for valuing H shows that in continuous time, we can only say that the arbitrage-free *seller price process* for the payoff H is given by $V^H = V^*$. The reason is that the strategy φ^H is admissible, but $-\varphi^H$ is not, in general, unless H is in addition bounded from above. In finite discrete time, this phenomenon does not appear because absence of arbitrage for admissible or for general self-financing strategies is the same there. ◇

7.2 Markovian payoffs and PDEs

The presentation in Section 7.1 is often called the *martingale approach* to valuing options, for obvious reasons. If one has more structure for the payoff H (and, in more general models, also for S), an alternative method involves the use of partial differential equations (PDEs) and is thus called the *PDE approach*. We briefly outline some aspects of this here.

Suppose that the (discounted) payoff is of the form $H = h(S^1_T)$ for some measurable function $h \geq 0$ on $\mathbb{R}_+$. We also suppose that H is in $L^1(Q^*)$. One example discussed in detail in the next section is the European call option on $\widetilde{S}^1$ with maturity T and undiscounted strike $\widetilde{K}$; here, $H = (\widetilde{S}^1_T - \widetilde{K})^+/\widetilde{S}^0_T = (S^1_T - \widetilde{K}e^{-rT})^+$ so that the payoff function is $h(x) = (x - \widetilde{K}e^{-rT})^+ =: (x-K)^+$. Our goal, for general h, is to compute the value process V^* and the strategy ϑ^H more explicitly.

We start with the *value process*. Because we have $V^*_t = E_{Q^*}[H \,|\, \mathcal{F}_t] = E_{Q^*}[h(S^1_T) \,|\, \mathcal{F}_t]$, we look at the explicit expression for S^1 in (1.11) and write

$$S^1_T = S^1_t \frac{S^1_T}{S^1_t} = S^1_t \exp\left(\sigma(W^*_T - W^*_t) - \frac{1}{2}\sigma^2(T-t)\right).$$

In the last term, the first factor S^1_t is obviously $\mathcal{F}_t$-measurable. Moreover, W^* is a Q^*-Brownian motion with respect to $\mathbb{F}$, and so in the second factor, $W^*_T - W^*_t$ is under Q^* independent of $\mathcal{F}_t$ and has an $\mathcal{N}(0, T-t)$-distribution. Therefore we get

$$V^*_t = E_{Q^*}[h(S^1_T) \,|\, \mathcal{F}_t] = v(t, S^1_t) \tag{2.1}$$

with the function $v(t,x)$ given, for $Y \sim \mathcal{N}(0,1)$ under Q^*, by

$$\begin{aligned} v(t,x) &= E_{Q^*}\left[h\left(x \exp\left(\sigma(W^*_T - W^*_t) - \frac{1}{2}\sigma^2(T-t)\right)\right)\right] \\ &= E_{Q^*}\left[h\left(xe^{\sigma\sqrt{T-t}\,Y - \frac{1}{2}\sigma^2(T-t)}\right)\right] \\ &= \int_{-\infty}^{\infty} h\left(xe^{\sigma\sqrt{T-t}\,y - \frac{1}{2}\sigma^2(T-t)}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}\,\mathrm{d}y. \end{aligned} \tag{2.2}$$

This already gives a fairly precise structural description of V^*_t as a function of (t and) S^1_t, instead of a general $\mathcal{F}_t$-measurable random variable.

Because we have an explicit formula for the function v as essentially the convolution of h with a very smooth function (the density of a lognormally distributed random variable), one can prove that the function v is *sufficiently smooth* to allow the use of Itô's formula. This gives, writing subscripts in the function v for partial derivatives and using (1.10) and (1.9),

$$\begin{aligned} \mathrm{d}V_t^* &= \mathrm{d}v(t, S_t^1) \\ &= v_t(t, S_t^1)\,\mathrm{d}t + v_x(t, S_t^1)\,\mathrm{d}S_t^1 + \frac{1}{2}v_{xx}(t, S_t^1)\,\mathrm{d}\langle S^1\rangle_t \\ &= v_x(t, S_t^1)\sigma S_t^1\,\mathrm{d}W_t^* + \left(v_t(t, S_t^1) + \frac{1}{2}v_{xx}(t, S_t^1)\sigma^2(S_t^1)^2\right)\mathrm{d}t. \end{aligned} \tag{2.3}$$

But V^* is a local (even a true) Q^*-martingale, by its definition, and so is the integrated $\mathrm{d}W^*$-term on the right-hand side above. Therefore the integrated $\mathrm{d}t$-term on the right-hand side of (2.3) is at the same time continuous and adapted and of finite variation, and a local Q^*-martingale. Hence it must vanish, and so (2.3) and (1.12) yield

$$v_x(t, S_t^1)\,\mathrm{d}S_t^1 = dV_t^* = \vartheta_t^H\,\mathrm{d}S_t^1.$$

In consequence, we obtain the *strategy* explicitly as

$$\vartheta_t^H = \frac{\partial v}{\partial x}(t, S_t^1), \tag{2.4}$$

i.e., as the spatial derivative of v, evaluated along the trajectories of S^1. This is parallel to the result in Section 3.3 for the CRR binomial model; see (3.3.6) or (3.3.7).

A closer look at the above argument also allows us to extract some information about the function v. This is similar to our arguments in Example 6.1.4 for the representation of the random variable $h(W_T)$ as a stochastic integral of W. Indeed, the fact that the $\mathrm{d}t$-term vanishes means that the function $v_t(t,x)+\frac{1}{2}v_{xx}(t,x)\sigma^2x^2$ must vanish along the trajectories of the space-time process $(t, S_t^1)_{0<t<T}$. But by the explicit expression in (1.11), each S_t^1 is lognormally distributed and hence has all of $(0,\infty)$ in its support. So the support of the space-time process contains $(0,T)\times(0,\infty)$, and so $v(t,x)$ must satisfy the (linear, second-order) *partial differential equation (PDE)*

$$0 = \frac{\partial v}{\partial t} + \frac{1}{2}\sigma^2x^2\frac{\partial^2 v}{\partial x^2} \qquad \text{on } (0,T)\times(0,\infty). \tag{2.5}$$

Moreover, the definition of v via (2.1) gives the *boundary condition*

$$v(T, \cdot\,) = h(\cdot\,) \qquad \text{on } (0, \infty), \tag{2.6}$$

because $v(T, S_T^1) = V_T^* = H = h(S_T^1)$ and the support of the distribution of S_T^1 contains $(0, \infty)$. So even if we cannot compute the integral in (2.2) explicitly, we can at least obtain $v(t, x)$ *numerically* by solving the PDE (2.5), (2.6).

Remarks. 1) Instead of using the above probabilistic argument, one can also derive the PDE (2.5) *analytically*. Using in (2.2) the substitution $u = x \exp(\sigma\sqrt{T-t}\, y - \frac{1}{2}\sigma^2(T-t))$ gives $y = (\log \frac{u}{x} + \frac{1}{2}\sigma^2(T-t))/\sigma\sqrt{T-t}$, hence $\mathrm{d}y = \frac{1}{u\sigma\sqrt{T-t}}\,\mathrm{d}u$, and then

$$v(t,x) = \int_0^\infty h(u) \frac{1}{\sqrt{2\pi\sigma^2(T-t)}} \frac{1}{u} \exp\left(- \frac{(\log \frac{u}{x} + \frac{1}{2}\sigma^2(T-t))^2}{2\sigma^2(T-t)} \right) \mathrm{d}u.$$

One can now first check, by using that $h(S_T^1)$ is in $L^1(Q^*)$, that v may be differentiated by differentiating under the integral sign, and by *brute force computations*, one can then check in this way that v indeed satisfies the PDE (2.5). The deeper reason behind this is the fact that the density function $\varphi(t,z) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{z^2}{2t}}$ of an $\mathcal{N}(0,t)$-distribution satisfies the heat equation $\varphi_t = \frac{1}{2}\varphi_{zz}$.

2) The above approach works not only for the Black–Scholes model, but more generally in a *Markovian setting*, because conditional expectations given $\mathcal{F}_t$ can there typically be written as functions of the state variables at time t. The martingale property then essentially translates into saying that the generator of the driving Markov process applied to the above functions must vanish. For diffusion state variables, the generator is a second-order differential operator and so this leads to PDEs; for Lévy state variables, one has additional integral terms coming from the jumps of the driving Lévy process, and so one obtains PIDEs (partial integro-differential equations). However, there are a number of substantial technical issues; for instance, regularity or existence of smooth solutions to the resulting equations is often not clear, and one must also be careful whether or not one has uniqueness of solutions. Not all the literature is equally rigorous and precise about these issues. ◇

When comparing the PDE (2.5), (2.6) to some of those found in the literature, one might be puzzled by the simple form of (2.5). This is because we have expressed everything in discounted units. If the *undiscounted* payoff is $\widetilde{H} = \tilde{h}(\widetilde{S}^1_T)$ and the undiscounted value at time t is written as $\tilde{v}(t, \widetilde{S}^1_t)$, we have the relations

$$\tilde{h}(\widetilde{S}^1_T) = \tilde{h}(e^{rT} S^1_T) = \widetilde{H} = e^{rT} H = e^{rT} h(S^1_T)$$

and

$$\tilde{v}(t, \widetilde{S}^1_t) = e^{rt} v(t, S^1_t)$$

so that

$$\begin{aligned} v(t,x) &= e^{-rt} \tilde{v}(t, xe^{rt}), \\ \tilde{v}(t,\tilde{x}) &= e^{rt} v(t, \tilde{x}e^{-rt}). \end{aligned}$$

For the function $\tilde{v}$, we can then compute the partial derivatives

$$\begin{aligned} \frac{\partial \tilde{v}}{\partial t}(t,\tilde{x}) &= r\tilde{v}(t,\tilde{x}) + e^{rt}\frac{\partial v}{\partial t}(t,\tilde{x}e^{-rt}) - e^{rt}\frac{\partial v}{\partial x}(t,\tilde{x}e^{-rt})\tilde{x}re^{-rt}, \\ \frac{\partial \tilde{v}}{\partial \tilde{x}}(t,\tilde{x}) &= e^{rt}\frac{\partial v}{\partial x}(t,\tilde{x}e^{-rt})e^{-rt} = \frac{\partial v}{\partial x}(t,\tilde{x}e^{-rt}), \\ \frac{\partial^2 \tilde{v}}{\partial \tilde{x}^2}(t,\tilde{x}) &= \frac{\partial^2 v}{\partial x^2}(t,\tilde{x}e^{-rt})e^{-rt}, \end{aligned}$$

and by plugging in, we obtain from (2.5) the PDE

$$0 = \frac{\partial \tilde{v}}{\partial t} + r\tilde{x}\frac{\partial \tilde{v}}{\partial \tilde{x}} + \frac{1}{2}\sigma^2\tilde{x}^2\frac{\partial^2 \tilde{v}}{\partial \tilde{x}^2} - r\tilde{v} \qquad \text{on } (0,T)\times(0,\infty)$$

with the boundary condition

$$\tilde{v}(T,\,\cdot\,) = \tilde{h}(\,\cdot\,).$$

[It is a nice [→ *exercise*] to convince oneself that this is correct. Possible ways include straightforward but tedious calculus, or alternatively again a martingale argument.]

7.3 The Black–Scholes formula

In the special case of a *European call option*, the value process and the corresponding strategy can be computed explicitly, and this has found widespread use in the financial industry. Suppose the undiscounted strike price is $\widetilde{K}$ so that the undiscounted payoff is

$$\widetilde{H} = (\widetilde{S}_T^1 - \widetilde{K})^+.$$

Then $H = \widetilde{H}/\widetilde{S}_T^0 = (S_T^1 - \widetilde{K}e^{-rT})^+ =: (S_T^1 - K)^+$, and we obtain from (2.2) that the discounted value of H at time t is

$$V_t^H = V_t^* = E_{Q^*}[H \,|\, \mathcal{F}_t] = E_{Q^*}[(S_T^1 - K)^+ \,|\, \mathcal{F}_t] = E_{Q^*}\Big[\Big(xe^{\sigma\sqrt{T-t}\,Y - \frac{1}{2}\sigma^2(T-t)} - K\Big)^+\Big]\Big|_{x=S_t^1},$$

with $Y \sim \mathcal{N}(0,1)$ under Q^*. An elementary computation with normal distributions yields for $x > 0$, $a > 0$ and $b \geq 0$ that

$$E_{Q^*}\big[\big(xe^{aY - \frac{1}{2}a^2} - b\big)^+\big] = x\Phi\bigg(\frac{\log\frac{x}{b} + \frac{1}{2}a^2}{a}\bigg) - b\Phi\bigg(\frac{\log\frac{x}{b} - \frac{1}{2}a^2}{a}\bigg),$$

where

$$\Phi(y) = Q^*[Y \leq y] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,\mathrm{d}z$$

is the cumulative distribution function of the standard normal distribution $\mathcal{N}(0,1)$. Plugging in $x = S_t^1$, $a = \sigma\sqrt{T-t}$, $b = K$ and then passing to undiscounted quantities via $S_t^1 = \widetilde{S}_t^1 e^{-rt}$, $K = \widetilde{K}e^{-rT}$ therefore yields the famous *Black–Scholes formula* in the form

$$\tilde{V}_t^{\widetilde{H}} = \tilde{v}(t, \widetilde{S}_t^1) = \widetilde{S}_t^1 \Phi(d_1) - \widetilde{K}e^{-r(T-t)}\Phi(d_2) \tag{3.1}$$

with

$$d_{1,2} = \frac{\log(\widetilde{S}_t^1/\widetilde{K}) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}. \tag{3.2}$$

Note that the drift μ of the stock does not appear here; this is analogous to the result that the probability p of an up move in the CRR binomial model does not appear in

the binomial option pricing formula (3.2), (3.3) in Section 3.3. What does appear is the volatility σ, in analogy to the difference $\log(1+u) - \log(1+d)$ which gives an indication of the spread between future stock prices from one time point to the next.

To compute the *replicating strategy*, we recall from (2.4) that the stock price holdings at time t are given by

$$\vartheta_t^H = \frac{\partial v}{\partial x}(t, S_t^1).$$

Moreover, $v(t,x) = e^{-rt}\tilde{v}(t, xe^{rt})$ so that

$$\frac{\partial v}{\partial x}(t,x) = e^{-rt}\frac{\partial \tilde{v}}{\partial x}(t, xe^{rt}) = e^{-rt}\frac{\partial \tilde{v}}{\partial \tilde{x}}(t, xe^{rt})e^{rt} = \frac{\partial \tilde{v}}{\partial \tilde{x}}(t, \tilde{x}).$$

Computing the above derivative explicitly [$\rightarrow$ *exercise*] gives

$$\vartheta_t^H = \frac{\partial \tilde{v}}{\partial \tilde{x}}(t, \widetilde{S}_t^1) = \Phi(d_1) = \Phi\left(\frac{\log(\widetilde{S}_t^1/\widetilde{K}) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}\right), \tag{3.3}$$

which always lies between 0 and 1.

One very useful feature of the above results is that the explicit formula (3.1), (3.2) allows to compute all partial derivatives of the option price with respect to the various parameters. These sensitivities are usually called *Greeks* and denoted by (genuine or invented) Greek letters. Examples are

- *Delta*: the partial derivative with respect to the asset price $\widetilde{S}_t^1$, computed in (3.3), also called *hedge ratio*.
- *Gamma*: the second partial derivative with respect to $\widetilde{S}_t^1$; it measures the reaction of Delta to a stock price change.
- *Rho*: the partial derivative with respect to the interest rate r.
- *Vega*: the partial derivative with respect to the volatility σ.
- *Theta*: the partial derivative with respect to $T-t$, the time to maturity.

- *Vanna*: the partial derivative of Delta with respect to σ, or the second partial derivative of the option price, once with respect to $\widetilde{S}^1_t$ and once with respect to σ.
- *Vomma*: the second partial derivative of the option price with respect to σ.
- *Charm*: the partial derivative of Delta with respect to $T - t$, the time to maturity.
- *Volga*: another term for Vomma.

Of course, the above definitions per se make sense for any model; but in the Black–Scholes model, one has even explicit expressions for them.

Remark. One can find in the literature many different derivations for the Black–Scholes *formula*. One especially popular approach is to first derive the binomial call pricing formula in the CRR model via arbitrage arguments, as we have done in Section 3.3, and to then pass to the limit by appropriately rescaling the parameters. More precisely, one considers for each $n \in \mathbb{N}$ a binomial model with time step T/n so that letting n increase corresponds to more and more frequent trading. It is intuitively plausible that the CRR models should then converge in some sense to the BS model, and one can make this mathematically precise via Donsker's theorem. Obtaining the Black–Scholes formula as a limit is similar but simpler; it is essentially an application of the central limit theorem.

The above limiting "derivation" of the Black–Scholes formula is mathematically much simpler; but it is also far less satisfactory, especially at the conceptual level. Most importantly, it does not give the key insight of the *methodology behind the formula*, namely that the price is the initial capital for a self-financing replication strategy in the continuous-time model. We do have the corresponding insight for each binomial model; but the elementary analysis usually done in the literature does not study whether that important structural property is preserved when passing to the limit. To obtain that insight (and to develop it further in other applications or maybe generalisations), stochastic calculus in continuous time is indispensable.

It is interesting to note that the above view was also shared by the Nobel Prize Committee; when it awarded the 1997 Nobel Prize in Economics to Robert C. Merton and Myron Scholes (Fischer Black had died in 1995), the award was given "for a new

method to determine the value of derivatives". The emphasis here is clearly on "method", as opposed to "formula". ◇

8 Appendix: Some basic concepts and results

This short chapter recalls some basic notations, concepts and results from probability theory. It is not exhaustive and not meant to serve as a replacement for a serious text in probability theory.

8.1 Very basic things

Let $\Omega \neq \emptyset$ be a nonempty set. We denote by 2^Ω the power set of Ω; this is the family of all subsets of Ω. A *σ-field* or *σ-algebra* on Ω is a family $\mathcal{F}$ of subsets of Ω which contains Ω and which is closed under taking complements and countable unions, i.e. if $A \in \mathcal{F}$, then also $A^c \in \mathcal{F}$, and if A_i, $i \in I\!N$, are in $\mathcal{F}$, then also $\bigcup_{i \in I\!N} A_i$ is in $\mathcal{F}$. Of course, $\mathcal{F}$ is then also closed under countable intersections. A σ-field is called finite if it contains only finitely many sets.

A pair $(\Omega, \mathcal{F})$ with $\Omega \neq \emptyset$ and $\mathcal{F}$ a σ-algebra on Ω is called a *measurable space*. One concrete example is $(I\!R, \mathcal{B}(I\!R))$, where $\mathcal{B}(I\!R)$ denotes the Borel-σ-field on $I\!R$. For any mapping $X : \Omega \to I\!R$ and any subset $B \subseteq I\!R$, we use the shorthand notation

$$X^{-1}(B) := \{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\}.$$

This is sometimes called the pre-image of the set B under the mapping X. We say that X is *measurable* (or more precisely *Borel-measurable*) if for every $B \in \mathcal{B}(I\!R)$, we have $\{X \in B\} \in \mathcal{F}$. One can show that this is equivalent to having $\{X \leq c\} \in \mathcal{F}$ for every $c \in I\!R$. More precisely, we could also say that $X : \Omega \to I\!R$ is measurable with respect to $\mathcal{F}$ and $\mathcal{B}(I\!R)$. If we replace the measurable space $(I\!R, \mathcal{B}(I\!R))$ by another measurable space $(\Omega', \mathcal{F}')$, say, we have an analogous definition of a measurable function from Ω to Ω', with respect to $\mathcal{F}$ and $\mathcal{F}'$.

For any subset A of Ω, the indicator function I_A is the function defined by

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

The function I_A is measurable if and only if $A \in \mathcal{F}$.

Sometimes, we start with $\Omega \neq \emptyset$ and a function $X : \Omega \to \mathbb{R}$ (or more generally to Ω'). Then $\sigma(X)$ is by definition the smallest σ-field $\mathcal{G}$, say, on Ω such that X is measurable with respect to $\mathcal{G}$ and $\mathcal{B}(\mathbb{R})$ (or $\mathcal{G}$ and $\mathcal{F}'$, respectively). We call $\sigma(X)$ the *σ-field generated by X*. Sometimes, we also consider a σ-field generated by a whole family of mappings; this is then analogously the smallest σ-field that makes all the mappings in that family measurable.

If $(\Omega, \mathcal{F})$ is a measurable space, a *probability measure* on $\mathcal{F}$ is a mapping $P : \mathcal{F} \to [0, 1]$ such that $P[\Omega] = 1$ and P is *σ-additive*, i.e.

$$P\left[\bigcup_{i \in \mathbb{N}} A_i\right] = \sum_{i \in \mathbb{N}} P[A_i] \quad \text{whenever } A_i,\ i \in \mathbb{N}, \text{ are sets in } \mathcal{F} \text{ that are pairwise disjoint.}$$

The triple $(\Omega, \mathcal{F}, P)$ is then called a *probability space*.

We say that a statement holds *P-almost surely* or *P-a.s.* if the set

$$A := \{\omega : \text{the statement does not hold}\}$$

is a P-nullset, i.e. has $P[A] = 0$. We sometimes also use instead the formulation that a statement holds for P-almost all ω. For example, $X \geq Y$ P-a.s. means that $P[X < Y] = 0$ or, equivalently, $P[X \geq Y] = 1$. Note that we also use here the shorthand notation

$$P[X \geq Y] := P[\{X \geq Y\}] := P[\{\omega \in \Omega : X(\omega) \geq Y(\omega)\}].$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ a measurable function. We also say that X is a (real-valued) *random variable*. If Y is another random variable, we call X and Y equivalent if $X = Y$ P-a.s. We then denote by L^0 or $L^0(\mathcal{F})$ the family of all equivalence classes of random variables on $(\Omega, \mathcal{F}, P)$. For $0 < p < \infty$, we denote by $L^p(P)$ the family of all equivalence classes of random variables X which are *p-integrable* in the sense that $E[|X|^p] < \infty$, and we write then $X \in L^p(P)$ or $X \in L^p$ for short. Finally, L^∞

is the family of all equivalence classes of random variables that are bounded by a constant c, say (where the constant can depend on the random variable).

If $(\Omega, \mathcal{F}, P)$ is a probability space, then an *atom* of $\mathcal{F}$ is a set $A \in \mathcal{F}$ with the properties that $P[A] > 0$ and that if $B \subseteq A$ is also in $\mathcal{F}$, then either $P[B] = 0$ or $P[B] = P[A]$. Intuitively, atoms are the "smallest P-indivisible sets" in a σ-field. Atoms are pairwise disjoint up to P-nullsets. The space $(\Omega, \mathcal{F}, P)$ is called *atomless* if $\mathcal{F}$ contains no atoms; this can only happen if $\mathcal{F}$ is infinite. On the other hand, a finite σ-field $\mathcal{F}$ can be very conveniently described via its atoms because every set in $\mathcal{F}$ is then a union of atoms.

8.2 Conditional expectations: A survival kit

This section gives a short summary of some basic notions about conditional expectations. We provide the definition and the most important properties, but hardly any proofs.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and U a real-valued random variable, i.e. an $\mathcal{F}$-measurable mapping $U : \Omega \to \mathbb{R}$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a fixed sub-σ-field of $\mathcal{F}$; the intuitive interpretation is that $\mathcal{G}$ gives us some *partial information*. The goal is then to find a *prediction* for U on the basis of the information conveyed by $\mathcal{G}$, or, put differently, a best estimate for U that uses only information from $\mathcal{G}$.

Definition. A *conditional expectation of U given $\mathcal{G}$* is a real-valued random variable Y with the following two properties:

(2.1) Y is $\mathcal{G}$-measurable.

(2.2) $E[U I_A] = E[Y I_A]$ for all $A \in \mathcal{G}$.

Y is then called a *version of the conditional expectation* and is denoted by $E[U \,|\, \mathcal{G}]$.

Theorem 2.1. *Let U be an integrable random variable, i.e. $U \in L^1(P)$. Then:*

1) *There exists a conditional expectation $E[U \,|\, \mathcal{G}]$, and $E[U \,|\, \mathcal{G}]$ is again integrable.*

2) *$E[U \,|\, \mathcal{G}]$ is unique up to P-nullsets: If Y, Y' are random variables satisfying* (2.1) *and* (2.2)*, then $Y' = Y$ P-a.s.*

Proof. 1) is nontrivial and not proved here; possible proofs use the Radon–Nikodým theorem or a projection argument in $L^2(P)$ combined with an extension argument.

2) Due to (2.1), the set $A := \{Y > Y'\}$ is in $\mathcal{G}$ so that (2.2) implies

$$0 = E[(Y - Y')I_A].$$

But by the definition of A, we have $(Y - Y')I_A \geq 0$ P-a.s., and so we get $(Y - Y')I_A = 0$ P-a.s., hence $P[A] = 0$ by the definition of A, i.e. $Y \leq Y'$ P-a.s. The converse inequality is proved in the same way. **q.e.d.**

We next recall without proofs some properties of and computation rules for conditional expectations. Let U, U' be integrable random variables so that $E[U \,|\, \mathcal{G}]$ and $E[U' \,|\, \mathcal{G}]$ exist. We denote by $b\mathcal{G}$ the set of all bounded $\mathcal{G}$-measurable random variables. Then we have:

(2.3) $E[UZ] = E\big[E[U \,|\, \mathcal{G}]Z\big]$ for all $Z \in b\mathcal{G}$.

Linearity: $E[aU + bU' \,|\, \mathcal{G}] = aE[U \,|\, \mathcal{G}] + bE[U' \,|\, \mathcal{G}]$ P-a.s., for all $a, b \in \mathbb{R}$.

Monotonicity: If $U \geq U'$ P-a.s., then $E[U \,|\, \mathcal{G}] \geq E[U' \,|\, \mathcal{G}]$ P-a.s.

Projectivity: $E[U \,|\, \mathcal{G}] = E\big[E[U \,|\, \mathcal{G}] \,\big|\, \mathcal{H}\big]$ P-a.s., for every σ-field $\mathcal{H} \subseteq \mathcal{G}$.

Further elementary properties are:

(2.4) $E[U \,|\, \mathcal{G}] = U$ P-a.s. if U is $\mathcal{G}$-measurable.

(2.5) $E\big[E[U \,|\, \mathcal{G}]\big] = E[U]$.

(2.6) $E[ZU \,|\, \mathcal{G}] = ZE[U \,|\, \mathcal{G}]$ P-a.s., for all $Z \in b\mathcal{G}$.

(2.7) $E[U \,|\, \mathcal{G}] = E[U]$ P-a.s. for U independent of $\mathcal{G}$.

In fact, (2.4) is clear from the definition, (2.5) follows immediately from (2.2) with $A = \Omega$, and (2.6) follows from (2.3) with the help of the definition. The right-hand side of (2.7) is clearly $\mathcal{G}$-measurable, and U and I_A are by assumption independent for every $A \in \mathcal{G}$; hence we obtain

$$E[UI_A] = E[U]E[I_A] = E\big[E[U]I_A\big]$$

and therefore (2.7), as (2.2) holds as well.

Remarks. **1)** Instead of integrability of U, one could also assume that $U \geq 0$; then analogous statements are true. One point of caution applies: if $U \geq 0$, then U as well as $E[U \,|\, \mathcal{G}]$ could take the value $+\infty$, and so one must be careful to avoid expressions involving $\infty - \infty$ as these are not well defined.

2) More generally, (2.3) and (2.6) hold as soon as U and ZU are both integrable or both nonnegative; this is often useful.

3) If U is $\mathbb{R}^d$-valued, one simply does everything component by component to obtain analogous results. ◇

For concrete computations of conditional expectations, the following result is often very useful.

Lemma 2.2. *Let U, V be random variables such that U is $\mathcal{G}$-measurable and V is independent of $\mathcal{G}$. For every measurable function $F \geq 0$ on $\mathbb{R}^2$, we then have*

$$E[F(U,V)\,|\,\mathcal{G}] = E[F(u,V)]\big|_{u=U} =: f(U). \tag{2.8}$$

Proof. For F of the form $F(u,v) = g(u)h(v)$ with $g, h \geq 0$ and measurable, we have on the one hand

$$f(u) = E[F(u,V)] = g(u)E[h(V)]$$

and on the other hand by (2.6) and (2.7) that

$$E[F(U,V)\,|\,\mathcal{G}] = E[g(U)h(V)\,|\,\mathcal{G}] = g(U)E[h(V)\,|\,\mathcal{G}] = g(U)E[h(V)] = f(U),$$

because $g(U)$ is $\mathcal{G}$-measurable and $h(V)$ is like V independent of $\mathcal{G}$. For general F, one then uses an argument via the so-called monotone class theorem. **q.e.d.**

Intuitively, (2.8) says that under the assumptions of Lemma 2.2, one can compute the conditional expectation $E[F(U,V)\,|\,\mathcal{G}]$ by "fixing the known value U and taking the expectation over the independent quantity V".

In analogy to Fatou's lemma and the dominated convergence theorem, one has the following convergence results for conditional expectations.

Theorem 2.3. *Suppose $(U_n)_{n\in\mathbb{N}}$ is a sequence of random variables.*

1) *If $U_n \geq X$ P-a.s. for all n and some integrable random variable X, then*

$$E\Big[\liminf_{n\to\infty} U_n \,\Big|\, \mathcal{G}\Big] \leq \liminf_{n\to\infty} E[U_n\,|\,\mathcal{G}] \qquad P\text{-a.s.}$$

2) *If (U_n) converges to some random variable U P-a.s. and if $|U_n| \leq X$ P-a.s. for all n and some integrable random variable X, then*

$$E\Big[\lim_{n\to\infty} U_n \,\Big|\, \mathcal{G}\Big] = E[U\,|\,\mathcal{G}] = \lim_{n\to\infty} E[U_n\,|\,\mathcal{G}] \qquad P\text{-a.s.} \tag{2.9}$$

Remark. In analogy to what happens for usual expectations, one might be tempted to think that (2.9) is still true if one replaces the assumption that all the U_n are dominated by an integrable random variable by the weaker requirement that the sequence (U_n) is uniformly integrable. But while this is still enough to conclude that $E[U] = \lim_{n\to\infty} E[U_n]$ (in fact, one even has convergence of (U_n) to U in $L^1(P)$), it does not imply that the conditional expectations converge P-a.s. (although they then do converge in L^1).

8.3 Stochastic processes and functions

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{T}$ an index set. Usually, we use $\mathcal{T} = \{0, 1, \ldots, T\}$ with some $T \in \mathbb{N}$, or $\mathcal{T} = [0, T]$ with some $T \in (0, \infty)$, or $\mathcal{T} = [0, \infty)$. A (real-valued) *stochastic process* with index set $\mathcal{T}$ is then a family of random variables X_t, $t \in \mathcal{T}$, which are all defined on the same probability space $(\Omega, \mathcal{F}, P)$. We often write $X = (X_t)_{t \in \mathcal{T}}$.

Mathematically, a stochastic process can be viewed as a function depending on two parameters, namely $\omega \in \Omega$ and $t \in \mathcal{T}$. If we fix $t \in \mathcal{T}$, then $\omega \mapsto X_t(\omega)$ is simply a random variable. If we fix instead $\omega \in \Omega$, then $t \mapsto X_t(\omega)$ can be viewed as a function $\mathcal{T} \to \mathbb{R}$, and we often call this the *path* or the *trajectory* of the process corresponding to ω. But also viewing a stochastic process as a mapping $X : \Omega \times \mathcal{T} \to \mathbb{R}$ is useful in some circumstances.

We say that a stochastic process is *continuous* if all or P-almost all its trajectories are continuous functions. We call a stochastic process *RCLL* if all or P-almost all its trajectories are right-continuous (RC) functions admitting left limits (LL). We say that a stochastic process is of *finite variation* if all or P-almost all its trajectories are functions of finite variation. Recall that a function is of finite variation if and only if it can be written as the difference of two increasing functions.

Finally, we say that a stochastic process has a property *locally* if there exists a sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ increasing to ∞ P-a.s. such that when restricted to the stochastic interval $[\![0, \tau_n]\!] = \{(\omega, t) \in \Omega \times \mathcal{T} : 0 \leq t \leq \tau_n(\omega)\}$, the process has the property under consideration. (Actually, this is a bit tricky. In some cases, for example when considering integrators, one can simply keep the process constant after τ_n at its time-τ_n level; in other cases, for example when considering integrands, one must set the process to 0 after time τ_n.)

9 References

[1] N. H. Bingham, R. Kiesel: *Risk-Neutral Valuation. Pricing and Hedging of Financial Derivatives*, second edition, Springer, 2004

[2] J. Cvitanić, F. Zapatero: *Economics and Mathematics of Financial Markets*, MIT Press, 2004

[3] R.-A. Dana, M. Jeanblanc: *Financial Markets in Continuous Time*, Springer, 2003

[4] F. Delbaen, W. Schachermayer: *The Mathematics of Arbitrage*, Springer, 2006

[5] C. Dellacherie, P.-A. Meyer: *Probabilities and Potential B. Theory of Martingales*, North-Holland, 1982

[6] R. Durrett: *Probability. Theory and Examples*, fourth edition, Cambridge University Press, 2010

[7] R. J. Elliott, P. E. Kopp: *Mathematics of Financial Markets*, second edition, Springer, 2005

[8] H. Föllmer: *Calcul d'Itô sans probabilités*, Séminaire de Probabilités XV, Lecture Notes in Mathematics 850, Springer, 1981, pp. 143–150

[9] H. Föllmer, A. Schied: *Stochastic Finance. An Introduction in Discrete Time*, third edition, de Gruyter, 2011

[10] J. Jacod, P. Protter: *Probability Essentials*, second edition, Springer, 2003

[11] J. Jacod, A. N. Shiryaev: *Limit Theorems for Stochastic Processes*, second edition, Springer, 2003

[12] D. Lamberton, B. Lapeyre: *Introduction to Stochastic Calculus Applied to Finance*, second edition, Chapman & Hall/CRC, 2008

[13] P. E. Protter: *Stochastic Integration and Differential Equations*, second edition, version 2.1, Springer, 2005

10 Index