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 $k = 0, 1, \dots, T$ with $T \in I\!N$.

The flow of *information* over time is described by a *filtration* $I\!F = (\mathcal{F}_k)_{k=0,1,\dots,T}$; 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 ($I\!R^d$-valued) *stochastic process* in this discrete-time setting is simply a family $X = (X_k)_{k=0,1,\dots,T}$ of ($I\!R^d$-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 $I\!F$) if each X_k is $\mathcal{F}_k$-measurable, i.e. observable at time k; it is called *predictable* (with respect to $I\!F$) if each X_k is even $\mathcal{F}_{k-1}$-measurable, for $k = 1, \dots, T$. (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 $r_1, \ldots, r_T$ and $Y_1, \ldots, Y_T$. 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 $k = 0, 1, \ldots, T$. 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 $k = 1, \ldots, T$. 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 $k = 0, 1, \ldots, T$. 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 $I\!\!F$ is the filtration generated by Y, i.e.,

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

is the smallest σ-field that makes all stock prices up to time k observable. Then $\widetilde{S}^1$ is obviously adapted to $I\!\!F$. 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

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

This explains a posteriori why we have started above directly with $I\!\!F$ 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 $I\!\!F = (\mathcal{F}_k)_{k=0,1,\dots,T}$ for some $T \in I\!\!N$, without repeating this explicitly. We shall only be more specific when we want to exploit special properties of a particular model $(\Omega, \mathcal{F}, I\!\!F, P)$. 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 $I\!\!R^d$-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. $\diamond$

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

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

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. ◇

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. $\diamond$

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. $\diamond$

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

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

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 $I\!R^d$-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 book-keeping* 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. [$\rightarrow$ *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 $\tau : \Omega \to \{0, 1, \dots, T\}$ 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

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

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

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

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 $S_k^\tau = S_{k\wedge\tau}$ 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 $I\!\!F$, 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

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

(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

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

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 $I\!\!F = (\mathcal{F}_k)_{k=0,1,\dots,T}$. A (real-valued) stochastic process $X = (X_k)_{k=0,1,\dots,T}$ is called a *martingale* (with respect to Q and $I\!\!F$) if it is adapted to $I\!\!F$, 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 $I\!\!R^d$-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 $(\Omega, \mathcal{F}, I\!\!F, P)$ 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 $X = (X_k)_{k=0,1,\dots,T}$ null at 0 (i.e. with $X_0 = 0$) is called a *local martingale* (with respect to Q and $I\!\!F$) if there exists a sequence of stopping times $(\tau_n)_{n\in I\!\!N}$ increasing to T such that for each $n \in I\!\!N$, the stopped process $X^{\tau_n} = (X_{k\wedge\tau_n})_{k=0,1,\dots,T}$ is a $(Q, I\!\!F)$-martingale. We then call $(\tau_n)_{n\in I\!\!N}$ 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. $\diamond$

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

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

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

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 $|\vartheta_k| \leq n$ for $k \leq \sigma_n$ 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 $\vartheta_k := I_{\{k \leq \tau\}}$ 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*] $\diamond$

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 $L^T = (L_k)_{k=0,1,\dots,T}$ 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 $k = 0, 1, \dots, T$ 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 $Y_1, \dots, Y_T$ are i.i.d. and take the finitely many values $1 + y_1, \dots, 1 + y_m$ with respective probabilities $p_1, \dots, p_m$. 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 $1+y_j$, $j = 1, \dots, m$. 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

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

be the set of all strings of length T formed by elements of $\{1, \dots, m\}$. 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 $Y_1, \dots, Y_T$ 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

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

Intuitively, this means that up to time k, we can observe the values of $Y_1, \dots, Y_k$ 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 $(\bar{x}_1, \dots, \bar{x}_k)$ of length k with elements $\bar{x}_i \in \{1, \dots, m\}$ such that A consists of all those $\omega \in \Omega$ that start with the substring $(\bar{x}_1, \dots, \bar{x}_k)$, i.e.

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

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 $A = A_{\bar{x}_1, \dots, \bar{x}_k}$ from $\mathcal{F}_k$ splits into precisely m subsets $A_1 = A_{\bar{x}_1, \dots, \bar{x}_k, 1}, \dots, A_m = A_{\bar{x}_1, \dots, \bar{x}_k, m}$ 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

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

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 $(\mathcal{G}_k)_{k=0,1,\dots,T}$ is in general not a filtration; we do not have $\mathcal{G}_k \subseteq \mathcal{G}_\ell$ for $k \le \ell$. $\diamond$

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 $(\bar{x}_1,\dots,\bar{x}_k)$ that parametrises the atom. For any atom $A = A_{\bar{x}_1,\dots,\bar{x}_k}$ of $\mathcal{F}_k$, we then look at its m successor atoms $A_1 = A_{\bar{x}_1,\dots,\bar{x}_k,1},\dots,A_m = A_{\bar{x}_1,\dots,\bar{x}_k,m}$ 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 $A_j \subseteq A$)

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

Because A is the disjoint union of $A_1, \dots, A_m$, we have $0 \le Q[A_j \,|\, A] \le 1$ for $j = 1, \dots, m$ 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 $\bar{\omega} = (\bar{x}_1, \dots, \bar{x}_T)$,

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

and iterate from here to obtain

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

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 $Q[A_j \,|\, A] = Q[A_{\bar{x}_1,\dots,\bar{x}_k,j} \,|\, A_{\bar{x}_1,\dots,\bar{x}_k}] = q_j(\bar{x}_1, \dots, \bar{x}_k)$. 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 $Y_1, \dots, Y_T$ 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, $Y_1, \dots, Y_T$ 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. $\diamond$