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 $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$ 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 (NA') $\Longrightarrow$ (NA) $\Longrightarrow$ (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. ◇

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 $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with

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

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 [$\rightarrow$ *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

$(G(\vartheta^{(n)}))_{n\in I\!N} = (\vartheta^{(n)} \bullet S)_{n\in I\!N}$ 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,

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

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 ∞. ◇

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

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

which implies that

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

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. [$\rightarrow$ *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 $\mathbb{P}_{\mathrm{e}}(S)$ or simply $\mathbb{P}_{\mathrm{e}}$ the set of all EMMs for S and by $\mathbb{P}_{\mathrm{e,loc}}$ the set of all ELMMs for S. Clearly, $\mathbb{P}_{\mathrm{e}} \subseteq \mathbb{P}_{\mathrm{e,loc}}$.

Saying that $\mathbb{P}_{\mathrm{e(,loc)}}(S)$ 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:*

$$\textit{(NA)} \quad \Longleftrightarrow \quad \mathbb{P}_{\mathrm{e}}(S) \neq \emptyset \quad \Longleftrightarrow \quad \mathbb{P}_{\mathrm{e,loc}}(S) \neq \emptyset.$$

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.

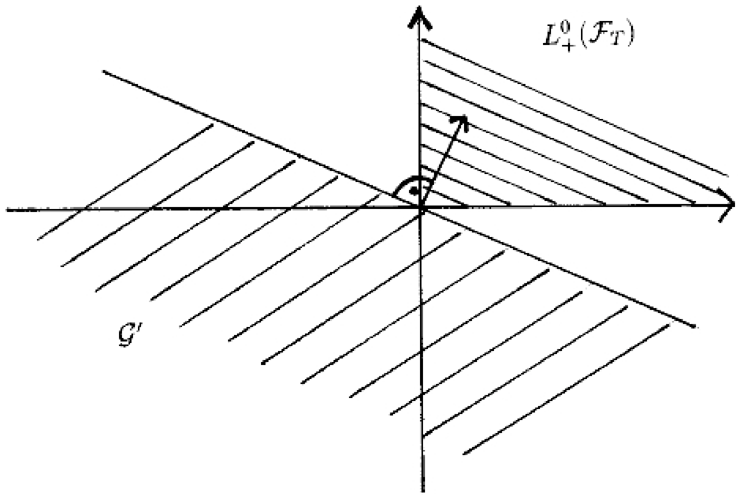
3) Note that we make no integrability assumptions about S (under P); so it is also noteworthy that S, being a Q-martingale, is automatically integrable under (some) Q. (To put this into perspective, one should add that it is a minor point; one can always easily construct [$\to$ *exercise*] a probability 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 [$\to$ *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 $Z = \sum_{A \,\mathrm{atom\,in}\, \mathcal{F}_T} Z I_A = \sum_{A \,\mathrm{atom\,in}\, \mathcal{F}_T} z_A I_A$ with $z_A \in I\!R$. We consider the set of all $\mathcal{F}_T$-measurable $Z \geq 0$ with $\sum_{A \,\mathrm{atom\,in}\, \mathcal{F}_T} z_A = 1$ and identify this with the subset

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

of $I\!R^n_+$, where n denotes the (finite, by assumption) number of atoms in $\mathcal{F}_T$. Then $\mathcal{K} \subseteq I\!R^n_+$ 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 $I\!R^n$ (see e.g. Lamberton/Lapeyre [12, Theorem A.3.2] implies that there exists a vector $\lambda \in I\!R^n$ with $\lambda \neq 0$ such that

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

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

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

(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 $I\!R^n$, 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 $\lambda^{\mathrm{tr}} h = 0$ for all $h \in \mathcal{H}$ translates via the identification of $\mathcal{H}$ and $\mathcal{G}'$ and the definition of $\mathcal{G}'$ into

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

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 $I\!\!F = (\mathcal{F}_k)_{k=0,1,\dots,T}$ 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 $I\!\!F$ 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

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

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 $D = (D_k)_{k=1,\dots,T}$ 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 $Y_1, \dots, Y_T$ 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 $y_1, \dots, y_m$ and r, the above recipe boils down to finding $q_1, \dots, q_m > 0$ 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 $U_1, \dots, U_T$ 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

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