It is not a priori quite clear to the author why exactly the standard framework of measure-theoretic probability (and financial calculus, in particular) is defined as such. This article aims to derive it from common sense and not pull any insights from higher intelligence.

1. Available information

A fundamental concept underlying financial mathematics is the mathematical model of *available information* — the way we encode what we know and what we don't know about a real-world event or process.

1.1. Setup

Imagine a probability experiment that can be "run" and on which "observations" can be made.

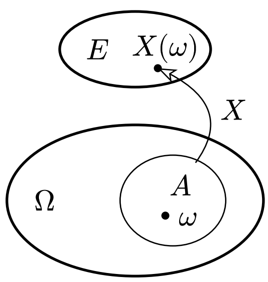
1. A single run fixes (determines) *everything* that can be observed about the experiment.
2. A *run* is commonly denoted by ω, and the set of *all* runs is commonly denoted by Ω.
3. Probabilities are assigned to sensible subsets of Ω, forming a σ-algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$ and a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$.
4. In practice, we do not know exactly which run $\omega \in \Omega$ occurred. Instead,
   a. we can only know if the run is *among whole sets* A of runs;
   b. we can usually only "look" at the experiment, and *observe* values of projections $X : \Omega \to E$ onto other sets E.

1.1.1. Events

Such a set of runs A, that we can know if ω lies in, is called *an event* and is only allowed to be in $\mathcal{F}$. Thus it is in the domain of $\mathbb{P}$ and has a well-defined *probability* $\mathbb{P}(A)$. The model does not really talk about whether a run is in a set not in $\mathcal{F}$.

1.1.2. Random variables

If we observe a particular value $e = X(\omega)$ through a projection $X : \Omega \to E$, then we know that the run that happened lies in $X^{-1}(\{e\}) \subset \Omega$. But we're only allowed to know if ω is in $\mathcal{F}$-sets, so the only (sets of) values $A \subset E$ (including the case $A = \{e\}$) we are allowed to observe from X must have measurable preimages $X^{-1}(A) \in \mathcal{F}$.



When we observe $e = X(\omega) \in E$, we know that $A = X^{-1}(e)$ happened, but not which ω exactly.

Thus to model observables we use *measurable* functions $X : \Omega \to E$ into measure spaces E, and we call X a *random variable.*

In this sense, $\mathcal{F}$ is a restriction on the nature of all random variables, containing the information *that is allowed to be known* and revealed by any observables.

2. Experiments that happen in time

Imagine we're observing an experiment that is laid out as a process over time $\mathbb{N}$.

2.1. Atomicity of history

A single run ω of the experiment contains the whole history of the process. That is, from a mathematical point of view, the future is predetermined.

As time evolves in the real world, though, we can only observe what is *currently* happening, and not the whole history, even though in theory it already "happened". Many runs of the experiment may have the same observable properties up to a particular time $n \in \mathbb{N}$, thus we only know that the run ω we're observing lies in a (probably large) measurable subset of Ω containing runs of the same observed history up to now but different futures.

For any n, let $\mathcal{F}_n \subset \mathcal{F}$ contain all maximal sets of events that share a common observable property up to time n or earlier. That is, $\mathcal{F}_n$ contains only, and all, sets of runs that a particular run can be determined to lie in, by observing it up to time n. Then trivially $\mathcal{F}_n \subset \mathcal{F}_m$ for all $n \leq m$.

If two runs ω and ω' differ in no observable way up to time n, they are not $\mathcal{F}_n$-separable[1].

Thus in order to model slowly-appearing history over Ω (recall, Ω is a set of atoms ω containing entire individual histories) we introduce a nested sequence of σ-algebras in $\mathcal{F}$

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_n \subset \ldots$$

called a *filtration* over Ω. It is a way to *lay out in time* the otherwise indivisible experiment runs ω without actually chopping them in time pieces. So at each step n we're allowed to know only which event in $\mathcal{F}_n$ has occurred and nothing more specific (like which run ω has happened).

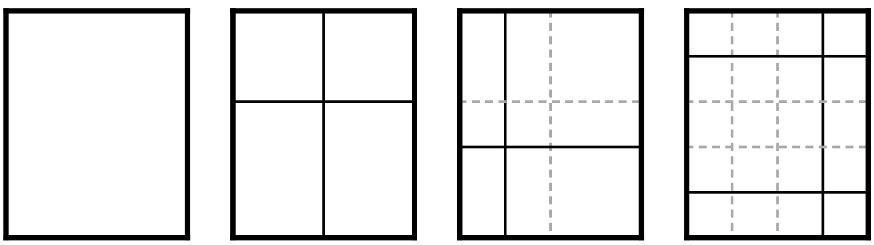
2.2. Random processes

A sequence of real-world observations is then modelled by
- projections $X_n : \Omega \to E$ that distinguish only properties observable up to time n,
- i.e. $X_n(\omega) \neq X_n(\omega')$ (with both values in disjoint measurable sets in E) is only allowed if ω and ω' differ *in an observable way up to time* n, so that X_n does not accidentally reveal information about ω that is not supposed to be known at time n.
- That is, the preimages of observable (sets of) values in E must be maximal w.r.t. observables up to time n,
- i.e. $X_n^{-1}(A)$ must be in $\mathcal{F}_n$ for all measurable $A \subset E$.

Or, X_n must be $\mathcal{F}_n$-measurable for all n. Such a sequence $X_\bullet$ is called a *stochastic process*.

2.3. Filtration example diagram

Schematically we can imagine a filtration on a rectangular Ω like this:



$\mathcal{F}_0 = \{\emptyset, \Omega\}$ $\mathcal{F}_1$ $\mathcal{F}_2$ $\mathcal{F}_3$

A run ω is a point in the rectangle. Each of the four rectangles represents the same Ω but with a different σ-algebra on it. The σ-algebras $\mathcal{F}_n$ contain all rectangles outlined in each case (both dashed and non-dashed[2]):
1. In the beginning, we know nothing (except that the run *happened*), so no subrectangles.
2. In $\mathcal{F}_1$, the model allows to observe which of the four regions ω lies in, but nothing more.

[1] I borrow that term from topology. "Two points x, y are *separated* by $\mathcal{F}_n$" means that there are sets $A, B \in \mathcal{F}_n$ that contain one but not the other: $x \in A, y \notin B$ and $x \in B, y \notin A$.

[2] non-dashed borders indicate the newly introduced measurable sets that weren't measurable in the previous step

3. In $\mathcal{F}_2$, we're allowed to observe which of the 9 (smaller!) rectangles it is in, and so on.

2.4. Example

For an experiment with 3 consecutive coin flips we can put

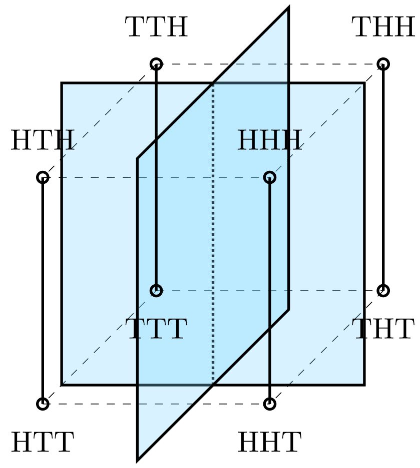
$$\Omega = \{\text{Heads}, \text{Tails}\}^3,$$

a discrete cube in 3D.

$\mathcal{F}_1$ contains only the front and back halves (apart from the whole cube and the empty set), $\mathcal{F}_2$ is generated by the four vertical edges, and $\mathcal{F}_3$ — by all 8 corners of the cube.

If $X_i : \Omega \to \{\text{Heads}, \text{Tails}\}$ is the outcome of the i th coin flip, then X_i is allowed to (though does not necessarily have to) differ only on different sets from $\mathcal{F}_i$, which it does.

Again, keep in mind that the entire runs of the experiment are fully encoded in the actual corners. The random process X_1, X_2, X_3 is only a sequence revealing smaller and smaller subsets in Ω around the true ω.



$\mathcal{F}_2$ is generated by the four vertical edges, i.e. the outcomes of the first two flips. Both X_1 and X_2 are $\mathcal{F}_2$-measurable, though neither of the two on its own separates the entire $\mathcal{F}_2$, only the vector (X_1, X_2) does.

3. Stopping times

Here we consider a property that is present up to some time point (potentially infinite) in the observed process (or equivalently, a property present only from some time point onward in the observed process).

Again, we model the time layout of the process by a filtration $\mathcal{F}_\bullet$. Then the presence of a property like above is modelled by a monotonic $\mathcal{F}_\bullet$-adapted boolean sequence $P_\bullet$ (with values in $2 = \{0, 1\}$ with the $\mathcal{P}(2)$ σ-algebra).

It is natural to ask now, what is the turning point? It is given by

$$T_X = \inf_{P_n = 1} n = \inf \{n \in \mathbb{N} : P_n = 1\}$$

At any time $n \in \mathbb{N}$, P_n tells us whether P has happened up to now. In terms of T_X, this is the event $T_X \leq n$, so it is $\mathcal{F}_n$ measurable.

Conversely, for every $\mathbb{N}$-valued random variable T with $\mathcal{F}_n$-measurable $T \leq n$ for every n, we can consider a monotone boolean sequence

$$X_n = \mathbb{1}_{\{T \leq n\}}$$

and then T is given from $X_\bullet$ as the infimum above.

A *stopping time* for a filtration $\mathcal{F}_\bullet$ is any $\mathbb{N}$-valued random variable for which

$$\forall n : \qquad \{T \leq n\} \ \text{ is } \ \mathcal{F}_n\text{-measurable},$$

i.e. it is the time at which an observed property (in the sense of a monotonic boolean sequence) "stops" holding.

4. Conditional expectation

If $\mathcal{G} \subset \mathcal{F}$ are σ-algebras on Ω, and $X : \Omega \to E$ is a function,

then	$\mathcal{G}$-measurability	is a stronger	restriction on X than $\mathcal{F}$-measurability,
because	$X^{-1}(A) \in \mathcal{G}$		condition than $X^{-1}(A) \in \mathcal{F}$.

That is, an event $A \in E$ is allowed to reveal less information on the run ω in the $\mathcal{G}$ case than in the $\mathcal{F}$ case, just because $\mathcal{G}$ contains less information than $\mathcal{F}$.

Thus an $\mathcal{F}$-measurable X need not be $\mathcal{G}$-measurable.

TODO

5. Martingales

Recall that an $\mathcal{F}_\bullet$**-martingale** is an $\mathcal{F}_\bullet$-adapted process $X_\bullet$ with zero-$\mathcal{F}_n$-expectation increments $\Delta X_{n+1} = X_{n+1} - X_n$:

Definition 5.1 (Martingale): An $\mathcal{F}_\bullet$-adapted process $X_\bullet$ is a called a **martingale** if

$$\forall n : \mathbb{E}[\Delta X_{n+1} | \mathcal{F}_n] = 0.$$

I prefer that phrasing of the definition because I find it more natural than the standard one. By writing

$$0 = \mathbb{E}[\Delta X_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_{n+1} | \mathcal{F}_n] - \mathbb{E}[X_n | \mathcal{F}_n]$$

as $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_n | \mathcal{F}_n]$ and applying $\mathcal{F}_n$-measurability to X_n, one gets equivalently

Definition 5.2 (Martingale, standard): An $\mathcal{F}_\bullet$-adapted process $X_\bullet$ is a called a **martingale** if

$$\forall n : \mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n.$$

Note that we can always decompose a process into the sum of its increments up a given time, i.e.

$$\forall n : X_n = X_0 + \sum_{i=1}^{n} \Delta X_i.$$

Here again ΔX_i denotes the "backwards" increment, $\Delta X_{i+1} = X_{i+1} - X_i$.

Now it is natural to ask, what happens if we scale the increments of a martingale, i.e. multiply each term in the above sum by some factor.

Definition 5.3 (Martingale transform): For an $\mathcal{F}_\bullet$-martingale $X_\bullet$ and a predictable $H_\bullet$, the martingale transform of $X_\bullet$ by $H_\bullet$ is denoted by $(H \odot X)_\bullet$ and defined by

$$(H \odot X)_n := \sum_{i=1}^{n} H_i \Delta X_i.$$

Using this notion, a martingale can be characterized not only by having its *individual* increments-means vanish, but by having (only) the final expectation of of all of its increments-transforms vanish:

Proposition 5.1: Let $X_\bullet$ be $\mathcal{F}_\bullet$-adapted.

$$X_\bullet \text{ is a martingale} \qquad \text{if an only if} \qquad \text{for any predictable } H_\bullet,\ \mathbb{E}[(H \odot X)_N] = 0.$$

6. Discrete-time market model

This section follows the first chapter of the book "Introduction to Stochastic Calculus Applied to Finance" (Lamberton & Lapeyre), whose explanations I found missing or unsatisfactory so I enhance the presentation with as much motivation as possible.

The market price vector is a stochastic process

$$S_i, \ i = 0, \ldots$$

with values in $\mathbb{R}^{1+d}$ adapted to a filtration

$$\mathcal{F}_i, \ i = 0, \ldots$$

of the market state known at time i. We consider $d \in \mathbb{N}$, $\mathcal{F}_\bullet$ and $S_\bullet$ fixed from now on.

The number $1 + d$ reflects that we track d risky assets and one riskless asset (the time-value of money) in the first component of $\mathbb{R}^{1+d}$.

Definition 6.1: A *trading strategy* is any $\mathbb{R}^{1+d}$-valued *predictable* sequence

$$\phi_i, \ i = 0, \ldots$$

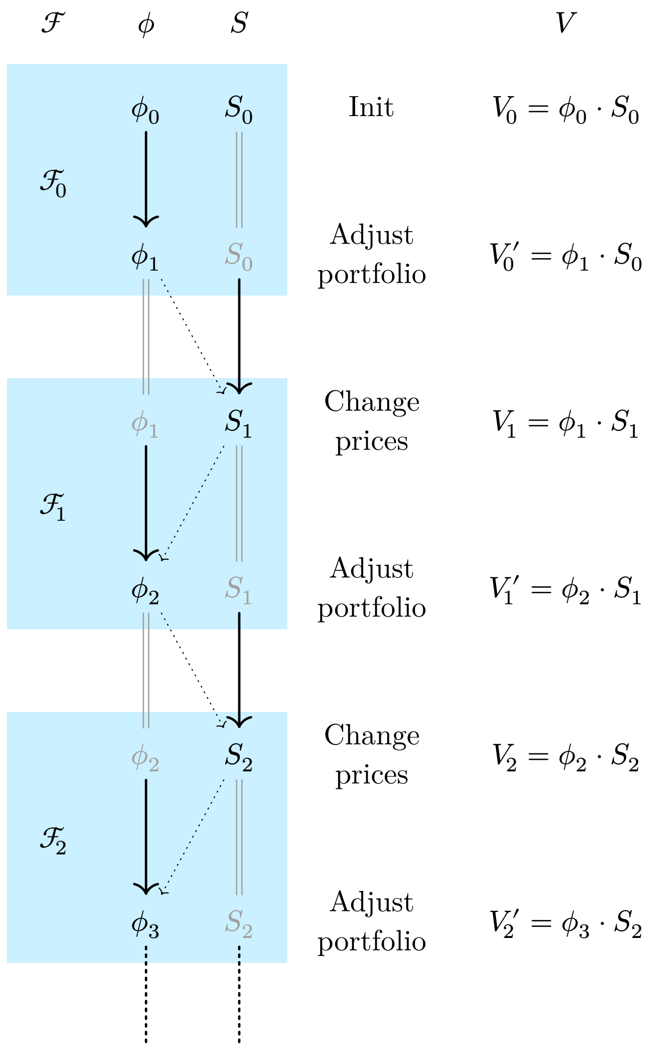
with respect to that filtration.

Predictable means that ϕ_{i+1} is $\mathcal{F}_i$-measurable, i.e. the information used (revealed) by the *next* portfolio ϕ_{i+1} is restricted to the information provided by the current market state $\mathcal{F}_i$, which is only guaranteed to contain the current stock prices (because S_i is $\mathcal{F}_i$-measurable) and past stock prices (because $\mathcal{F}_i \subset \mathcal{F}_{i+1}$).

Definition 6.2: The *value of the portfolio* defined by a strategy $\phi_\bullet$ is

$$V_i(\phi) := \phi_i \cdot S_i.$$

Now a slight confusion may arise as to what is the time order of the quantities S_i and ϕ_i. They both use the same index i so we might be under the impression that the time evolution is described by a sequence of pairs (ϕ_i, S_i). While not false, this may be deceiving when we define *self-financing* strategies. Consider the diagram on the right.



Here, in each timebox $\mathcal{F}_i$, first prices change and then the portfolio is adjusted, so we have an alternating sequence of price and portfolio changes.
The strategy ϕ is then self-financing iff the portfolio value doesn't change within a timebox ($V_i = V_i'$), but only responds to stock price changes $S_i \to S_{i+1}$.

Diagonal dotted arrows indicate the causal relationship between the time-evolution of the strategy and the stock prices in the process $(\phi_0, S_0), (\phi_1, S_1), \ldots$.

Notation 6.3: Denote the *portfolio adjustment* (again, non-standard terminology) performed at time i by

$$\Delta\phi_i \coloneqq \phi_{i+1} - \phi_i.$$

(consider the diagram above to convince yourself that $\Delta\phi_i$ is $\mathcal{F}_i$-measurable)

It is a vector of the sells/buys of each stock performed at time i. That is,

$\boxed{\Delta\phi_i^j \text{ is the amount of stock } j \text{ that is sold (if } \Delta\phi_i < 0\text{) or bought (if } \Delta\phi_i > 0\text{) at time } i.}$

Notation 6.4: Denote the *price changes* that happened at time $n+1$ by

$$\Delta S_{n+1} \coloneqq S_{n+1} - S_n.$$

Remark 6.5: $\Delta\phi_n$ was defined as the *forward* difference (between times $n+1$ and n), while the ΔS_n was defined as the *past* difference (between times n and $n-1$), so that both $\Delta\phi_n$ and ΔS_n are $\mathcal{F}_n$-measurable.

We also get a Leibniz rule for $V_\bullet$ ($\mathrm{d}V = \mathrm{d}\phi \cdot S + \mathrm{d}S \cdot \phi$).

Remark 6.6: The two increments let us express the portfolio value with a recursive formula:

$$\underbrace{V_{n+1}(\phi)}_{\substack{\text{next}\\\text{value}}} = \underbrace{V_n(\phi)}_{\substack{\text{prev}\\\text{value}}} + \underbrace{\Delta\phi_n \cdot S_n}_{\substack{\text{changes in}\\\text{strategy}}} + \underbrace{\phi_{n+1} \cdot \Delta S_n}_{\substack{\text{changes in}\\\text{prices}}}.$$

That is, the next value is the previous value plus the changes from the strategy and the changes in the prices.

6.1. Self-financing strategies

When the second term in the sum on the right in the remark above, $\Delta\phi_n \cdot S_n$, vanishes for all n, we get a very important type of strategies. Denote the first part of recursive formula by

$$V_n'(\phi) = V_n(\phi) + \Delta\phi_n \cdot S_n,$$

so that

$$V_{n+1}(\phi) = V_n'(\phi) + \phi_{n+1} \cdot \Delta S_n.$$

Those two mutually-recursive formulas reflect the structure of the alternating sequence $V_0, V_0', V_1, V_1', \ldots$ in the diagram above, so

Remark 6.1.1: $V'_n(\phi)$ is the *adjusted value* of the portfolio **after** the next strategy choices ϕ_{n+1} are applied but **before** the prices have been updated.

Now the condition that the second term in the earlier recursive formula for $V(\phi)$ vanishes for all n can be phrased simply as the equality $V(\phi) = V'(\phi)$ (as processes). It is instructive to derive a direct expression for V':

$$\begin{aligned} V'_n(\phi) &\equiv V_n(\phi) && + \Delta\phi_n \cdot S_n, \\ &= \cancel{\phi_n \cdot S_n} && + (\phi_{n+1} \cdot S_n - \cancel{\phi_n \cdot S_n}) \\ &= \phi_{n+1} \cdot S_n. \end{aligned}$$

Definition 6.1.2: A strategy ϕ_i is *self-financing* if the value of the portfolio stays the same after the strategy adjustment:

$$V'_i(\phi) = V_i(\phi) \qquad \text{for all } i,$$

i.e.

$$\phi_{i+1} \cdot S_i = \phi_i \cdot S_i \qquad \text{for all } i.$$

Now we can get back to the condition on the vanishing second term in the recursive formula for V:

Remark 6.1.3: The self-financing condition $V_i = V'_i$ is equivalent to

$$\Delta\phi_i \cdot S_i = 0,$$

i.e. all buys and sells cancel each other in value, i.e. no money is lost or needs to be brought in for the adjustment.

Then a self-financing strategy is one where the recursive formula for V has a vanishing second term, i.e.

$$V_{n+1}(\phi) = V_n(\phi) + \phi_{n+1} \cdot \Delta S_n$$

and that allows us to write

$$V_n = V_0 + \sum_{i=1}^{n} \phi_n \cdot \Delta S_n$$

It should be close to mind that even for a strategy that is not self-financing (i.e. some of the $\Delta\phi_n \cdot S_n$ does not vanish), if we were to put the quantity $\Delta\phi_n \cdot S_n$ into ϕ^0_n, i.e. if we invest (borrow) the surplus (shortage) of sells-minus-buys into (from) the riskless asset, we'd get a self-financing strategy.

Proposition 6.1.1: Any $\mathbb{R}^d$-valued predictable sequence (a "strategy" only on the risky assets)

$$\left(\phi_n^1, ..., \phi_n^d\right) \qquad \text{for} \quad n = 0, 1, ...$$

is a restriction of a unique self-financing strategy ($\mathbb{R}^{1+d}$-valued)

$$\left(\boldsymbol{\phi_n^0}, \phi_n^1, ..., \phi_n^d\right) \qquad \text{for} \quad n = 0, 1, ...$$

for any choice of an initial value $V_0(\phi) \in \mathbb{R}$ (or a choice of any of ϕ_n^0 for $n \geq 0$).

In other words, the self-financing strategies restricting to a given $\left(\phi_n^1, ..., \phi_n^d\right)_n$ are a one-parameter family indexed by V_0.

Proof: Fix $\underline{\phi}_n = \left(\phi_n^1, ..., \phi_n^d\right)$ for all $n \geq 0$. Any strategy $\phi_\bullet$ that (on the risky assets) restricts to this process is determined by the choices of bank account amounts $\varphi_\bullet^0$.

The self-financing condition $\Delta\phi_n \cdot S_n = 0$ imposes a restriction on $\varphi_\bullet^0$, though, and reads

$$\underbrace{\left(\phi_{n+1}^0 - \phi_n^0\right)} \cdot S_n^0 + \sum_{i=1}^d \left(\phi_1^i - \phi_n^i\right) \cdot S_n^i = 0 \qquad \text{for all } n = 0, 1, ...$$

Here a choice of any element of the sequence $\phi_\bullet^0$ determines the rest, so the self-financing strategies restricting to $\underline{\phi}_n$ are a one-parametric family indexed by (e.g.) ϕ_0^0.

Since $V_0 \equiv \phi_0^0 \cdot S_0 + \sum_{i=1}^d \phi_0^i \cdot S_0^i$ establishes a 1:1 relation between V_0 and ϕ_0^0 (given $\underline{\phi}_0$), we can also consider this parameter to be V_0. □

6.2. Arbitrage

Definition 6.2.1 (Admissible strategy): A *self-financing* strategy is called **admissible** if $V_n(\phi) \geq 0$ for all n (and all $\omega \in \Omega$).

We fix an $N \in \mathbb{N}$, called the *horizon*. All indices n, i, j, k below vary between 0 and N.

The market is said to have an **arbitrage opportunity** if some admissible strategy with zero initial value delivers a strictly positive value on a non-null set.
That is, an arbitrage opportunity is a

zero-investment	**risk-free**	**profit**.
($V_0 = 0$)	($V_n \geq 0$ for *every* ω and n)	($V_N \gneqq 0$ on a non-null set)

A market without arbitrage opportunities is called viable.

Definition 6.2.2 (Viable market): The market is called **viable** if every admissible strategy with $V_0(\phi) = 0$ satisfies $V_N(\phi) = 0$.

As a silly (counter)example, if

$$\text{at} \quad n = 0, \quad S^1_0 = S^0_0$$
$$\text{at} \quad n > 0, \quad S^1_n = 2 \cdot S^0_n,$$

i.e. the price of risky asset 1 equals the riskless at first but then becomes twice as much as the riskless one, then the constant strategy $\phi_n = (-1, 1)$ has zero initial value and $V_n = -S^0_n + S^1_n = S^0_n > 0$. Thus the market defined by those $S^\bullet_\bullet$ is not viable.

Proposition 6.2.1 (Fundamental theorem of asset pricing):

A market is viable	if an only if	under some equivalent measure, the discounted price processes $S^k_\bullet$, $k = 0, 1, \ldots$ are martingales.

Writing out the definition of a viable market, and applying the characterization of martingales in Proposition 5.1, the proposition reads

For any self-financing $\phi_\bullet$, if $\begin{matrix} \forall n : V_n(\phi) \geq 0 \\ \text{and } V_0(\phi) = 0 \end{matrix}$, then $V_N(\phi) = 0$	if an only if	there exists $\mathbb{P}^*$ equivalent to $\mathbb{P}$, such that for any predictable $H_\bullet$, $\mathbb{E}\left[\sum_{i=1}^{N} H_i \Delta S_i\right] = 0$

6.3. Perfect hedging (attainable claims)

A **contingent claim** is a promise to pay some amount h at a *maturity time* T, the amount depending on the market state. We model it by a non-negative real random variable h that is only $\mathcal{F}_T$-measurable, so the amount might not be known in advance.

Remark 6.3.1: Consider a contingent claim h at a maturity time T.

If all admissible strategies ϕ induce a portfolio value at T that is
- *less* than h, the contingent claim is essentially "free money" for the receiver;
- *greater* than h,

Definition 6.3.2: An $\mathbb{R}$-valued random variable h is called **attainable** if there exists an admissible strategy ϕ that results in the end in a portfolio of value exactly h:

$$V_N(\phi) = h.$$

Remark 6.3.3: If h is a constant $h(\omega) = M \in \mathbb{R}_+$, it is trivially attainable by the strategy with $\phi_N^0 = M$ that is the constant zero on the risky assets (exists uniquely by the proposition above, and is obviously admissible).

Even if h is only *bounded* by a constant, $h \leq M$, we could attain the (higher) value M by the same approach. Thus attainability is *not* essentially about reaching *at least* some value — we are allowed to start with as much cash in advance as we need.

In the case of a finite Ω (which is assumed in the book), h is bounded anyway. Attaining h, and not just M, is more difficult. If h is not $\mathcal{F}_0$-measurable, we cannot just prepare the cash in advance, because we are not allowed (by $\mathcal{F}_0$) to know exactly how much we will need (even if know it won't exceed M).

For example, if $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the initial cash ϕ_0^0 is only allowed to be constant (by $\mathcal{F}_0$-measurability). Thus even in the bounded case we have to make use of the risky assets to achieve the precise target h, and we have to abide by $\mathcal{F}_n$-measurability all the way $n = 0, ..., N$.

Definition 6.3.4: The market is called *complete* if every non-negative h is attainable.