Stochastische Prozesse, Sommersemester 2020, Corona-Edition

Irene Klein[1]

June 21, 2020

[1]Vielen Dank an Manuel Hahn! Seine Mitschrift in LaTeXvom Sommersemester 2019 war für dieses Skriptum unverzichtbar.

Contents

Chapter 1

Basics of stochastic processes

1.1 Basics and some examples

A stochastic process is a mathematical model of phenomena that depend on a state of the world ω and occur at times t in some time set $\mathbb{T}$. The time set can be discrete or continuous, for example, $\mathbb{T}$ could be

- $\mathbb{T} = \{0, 1, ..., T\}$
- $\mathbb{T} = \mathbb{N} \cup \{0\}$
- $\mathbb{T} = [0, T]$
- $\mathbb{T} = [0, \infty)$

A stochastic process is defined as follows: we fix a sample space Ω with a σ-algebra $\mathcal{F}$ on it. The random phenomenon is modeled by a family of random variables $X_t(.)$, indexed by $t \in \mathbb{T}$, i.e., such that each $X_t : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is measurable.

Examples:

- $(X_t)_{t\in\mathbb{T}}$ could be a Markov chain in discrete or continuous time
- $(X_t)_{t\in\mathbb{T}}$ could model a financial asset price process
- $(X_t)_{t\in\mathbb{T}}$ could describe the movement of particles in a liquid (Brownian Motion)
- $(X_t)_{t\in\mathbb{T}}$ could model the disease-free survival after some treatment
- $(X_t)_{t\in\mathbb{T}}$ could model the number of people in a population infected by COVID-19

Reminder: A function $X : \Omega \to \mathbb{R}$ is $(\mathcal{F}, \mathcal{B})$–measurable if for each $\alpha \in \mathbb{R}$ we have that

$$X^{-1}((-\infty, \alpha]) = \{\omega \in \Omega : X(\omega) \in (-\infty, \alpha]\} \in \mathcal{F}.$$

Definition 1.1.1. *Let $(X_t)_{t \in \mathbb{T}}$ be a stochastic process on $(\Omega, \mathcal{F})$. For a fixed $\omega \in \Omega$ we call the function $t \to X_t(\omega)$ a* **path** *or* **realization** *of the stochastic process $(X_t)_{t \in \mathbb{T}}$.*

Example 1.1.2 (Random Walk). Let $(\epsilon_k)_{k \geq 1}$ be a sequence of iid random variables and define $(X_n)_{n=0,1,2,...}$ as follows:

$$\begin{aligned} X_0 &= 0 \\ X_n &= \sum_{k=1}^{n} \epsilon_k, \text{ for each } n \geq 1. \end{aligned}$$

A particular choice of the iid sequence could for example be

$$\epsilon_k = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

A random walk defined as above has the discrete time set $\mathbb{T} = \mathbb{N} \cup \{0\}$.

Remark 1.1.3. Whenever we consider a stochastic process with a discrete time set $\mathbb{T}$ (as, for example, in the case of a random walk where $\mathbb{T} = \mathbb{N} \cup \{0\}$) we denote the time index by n rather than t (a notation we use for processes in continuous time).

Example 1.1.4 (Survival process). Let Z be a non-negative random variable on some probability space $(\Omega, \mathcal{F}, P)$. Define a stochastic process as follows:

$$X_t = \mathbb{1}_{\{Z > t\}}.$$

The random variable Z describes the random time a patient survives after some treatment. The stochastic process (X_t) takes only the values 1 and 0. $X_t(\omega) = 1$ as long as the patient is alive, i.e., $t < Z(\omega)$. Then the process jumps to 0. Note that this is a process in continuous time and that the paths $t \mapsto X_t(\omega)$ are right-continuous a.s.

Example 1.1.5 (Interpolated random walk). We would like to introduce an example that mimics the behavior of the random walk as in Example 1.1.2 but is a stochastic process in *continuous time*. Remember, the random walk of Example 1.1.2 is defined in *discrete* time, i.e., $X_0, X_1, X_2, \ldots, X_n, \ldots$. We will start with exactly the process of Example 1.1.2 called $(X_n)_{n=0}^{\infty}$. We will define a new stochastic process from that. Fix any $t \in [0, \infty)$. Then, of course, we can find natural numbers n and $n+1$ such that $n \leq t < n+1$. (We can find those natural numbers for every such t, but of course they depend on the particular t, so precisely it means: for each t there is $n(t) \in \mathbb{N}$ such that $n(t) \leq t < n(t)+1$.) Then we define $(Y_t)_{t\in[0,+\infty)}$ by $Y_0 = 0$ and

$$Y_t = X_n + (t-n)\epsilon_{n+1} \tag{1.1}$$

for each $t > 0$ such that $n \leq t < n+1$.

1.2 Filtrations

We would like to have a way to describe the development of information as time increases. Intuitively it is clear that for every phenomenon that evolves in time we have more and more information as time passes. How do we model this mathematically? We will not only consider one σ-algebra but an increasing family of σ-algebras that are indexed by the time set $\mathbb{T}$. Precisely, this means:

Definition 1.2.1. *Let $(\mathcal{F}_t)_{t\in\mathbb{T}}$ be an increasing family of σ-algebras, that is: (i) for every t we have that $\mathcal{F}_t$ is a σ-algebra and (ii) if $s < t$ then $\mathcal{F}_s \subseteq \mathcal{F}_t$. The σ-algebra $\mathcal{F}_t$ contains all events that we can "distinguish" at time t, it is the cumulated information over the time period from 0 up to t. Moreover we assume that the σ-algebra $\mathcal{F}$ from before is the largest σ-algebra, it contains all information, i.e., $\mathcal{F}_t \subseteq \mathcal{F}$ for all $t \in \mathbb{T}$.*

Example 1.2.2. Look back to the example of the random walk, see Example 1.1.2 with the particular ε_k taking the values $+1, -1$. Which information do we have if we know the random walk up to time $n = 0, 1, 2, \ldots$?

It is easy to see that at time 0, as $X_0 = 0$ (constantly 0) we don't have any information more than the trivial one:

$$\mathcal{F}_0 = \{\emptyset, \Omega\}$$

For each $n \geq 1$ it is easy to see that $\mathcal{F}_n = \sigma\{X_1, X_2, \ldots, X_n\} = \sigma\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$, and hence we get

$$\mathcal{F}_1 = \{\emptyset, \{\epsilon_1 = 1\}, \{\epsilon_1 = -1\}, \Omega\}$$

$$\mathcal{F}_2 = \sigma(\{\{\epsilon_1 = 1, \epsilon_2 = 1\}, \{\epsilon_1 = 1, \epsilon_2 = -1\}, \{\epsilon_1 = -1, \epsilon_2 = 1\}, \{\epsilon_1 = -1, \epsilon_2 = -1\}\})$$

...

$$\mathcal{F}_n = \sigma(\text{all possible sets of } \{\varepsilon_k = \pm 1, k = 1, \ldots, n\})$$

What we found in Example 1.2.2 is the so called "natural fitration" of the process $(X_n)_{n=0}^{\infty}$ or the "filtration generated by the random walk".

Definition 1.2.3. *The generated filtration of a stochastic process* $(X_t)_{t\in\mathbb{T}}$ *is given by*

$$\mathcal{F}_t^X = \sigma(\{X_s, s \le t\})$$

Remark 1.2.4. (i) For the concept of filtration there is no substantial difference between discrete and continuous time. However, let us formulate the definition in discrete time for the time set $\mathbb{T} = \mathbb{N} \cup \{0\}$. Then the condition $s < t$ can be reduced to $n < n+1$ as we only consider integers. So we get: a filtration is a family $(\mathcal{F}_n)_{n=0}^{\infty}$ of σ-algebras such that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n = 0, 1, 2, \ldots$.

(ii) The generated filtration in discrete time is given as we did in Example 1.2.2: $\mathcal{F}_n = \sigma\{X_k, k = 1, \ldots, n\}$.

Example 1.2.5. Let us now find the generated filtration of the survival process $(X_t)_{t\ge 0}$ as given in Example 1.1.4. We have that

$$\mathcal{F}_t = \sigma(\{X_s : s \le t\}) = \sigma(\{Z \le s\} : \text{for all } s \le t\}) \cup \{Z > t\}$$

This means that in $\mathcal{F}_t$, $Z\mathbb{1}_{\{Z\le t\}}$ does have full measurability, but $\{Z > t\}$ is an *atom*. (This means there are no non-trivial subsets A of $\{Z > t\}$ with $A \in \mathcal{F}_t$. Precisely, it holds that if $A \subseteq \{Z > t\}$ and $A \in \mathcal{F}_t$ then $A = \emptyset$ or $A = \{Z > t\}$). So, **after** t we only know if Z has already happened or not, eg. if it's bigger than t or not. In contrast to that every measurable subset of $\{Z \le t\}$ is contained in $\mathcal{F}_t$ (i.e. full measurability of $Z\mathbb{1}_{\{Z\le t\}}$, we know everything about Z **before** t).

Definition 1.2.6. *Suppose* $(\mathcal{F}_t)_{t\in\mathbb{T}}$ *is a filtration on* $(\Omega, \mathcal{F})$*. A stochastic process* $(X_t)_{t\in\mathbb{T}}$ *on* $(\Omega, \mathcal{F})$ *is* **adapted** *to* $(\mathcal{F}_t)_{t\ge 0}$ *if, for all* $t \in \mathbb{T}$*:*

$$X_t : (\Omega, \mathcal{F}_t) \to (\mathbb{R}, \mathcal{B}) \ \textit{measurable} \tag{1.2}$$

Remark 1.2.7 (Notation). In all the following we will also use the short notation: X_t is $\mathcal{F}_t$-measurable if the measurability as above, see (1.2), holds.

What is the interpretation of an adapted stochastic process? It means that when we have the information $\mathcal{F}_t$ we know how the process behaved until time t, i.e., we have measurability of the whole stochastic process $(X_s)_{0\le s\le t}$. How is the property of being adapted connected to the filtration that is generated by a stochastic process $(X_t)_{t\in\mathbb{T}}$? We summarize the properties in the following remark:

Remark 1.2.8. (i) As, by definition, the generated filtration is given by $\mathcal{F}_t^X = \sigma(\{X_s, s \leq t\})$ it is obvious that each X_s, for $s \leq t$, is measurable with respect to $\mathcal{F}_t^X$. Hence $(X_t)_{t\in\mathbb{T}}$ is always adapted to its generated filtration $(\mathcal{F}_t^X)_{t\in\mathbb{T}}$.

(ii) On the other hand, if we consider any filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ such that the stochastic process $(X_t)_{t\in\mathbb{T}}$ is adapted to this filtration it is obvious that the minimal information that has to be contained in $\mathcal{F}_t$ is $\sigma\{X_s, s \leq t\}$. And so, for all t, we have $\mathcal{F}_t^X \subseteq \mathcal{F}_t$. This means that $(\mathcal{F}_t^X)_{t\in\mathbb{T}}$ is the smallest filtration, such that the stochastic process $(X_t)_{t\in\mathbb{T}}$ is adapted.

Chapter 2

Conditional expectation

2.1 Definition and properties

Definition 2.1.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ *be measurable with* $E[|X|] < \infty$*. Let* $\mathcal{G} \subset \mathcal{F}$ *be a smaller* σ*-algebra. Then there exists a measurable function* $Y : (\Omega, \mathcal{G}) \to (\mathbb{R}, \mathcal{B})$ *with* $\mathbb{E}[|Y|] < \infty$ *such that the following condition holds:*

$$\textit{for all } A \in \mathcal{G}: \quad \mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A Y] \tag{2.1}$$

This $\mathcal{G}$*-measurable random variable is called the conditional expected value (conditional expectation) of* X *given* $\mathcal{G}$*. Note that* Y *is unique with respect to equality a.s., i.e., if* Z *is another integrable* $\mathcal{G}$*-measurable random variable that satisfies (2.1) then* $Y = Z$ *a.s. We will use the notation* $E[X|\mathcal{G}]$ *for the conditional expected value of* X *given* $\mathcal{G}$*, that means for the* Y *given above it holds that*

$$E[X|\mathcal{G}] = Y \ \ a.s.$$

Remark 2.1.2. What is the purpose of the conditional expectation? If X is a function that is $\mathcal{F} - \mathcal{B}$ measurable and $\mathcal{G} \subset \mathcal{F}$ is a strictly smaller σ-algebra then, in general, X is not $\mathcal{G} - \mathcal{B}$-measurable, see Example 2.1.3 below. If we calculate the usual expected value $E[X]$ we find a constant in $\mathbb{R}$ that is a "good" estimator for the random variable, in the sense that, trivially, this constant hast the same mean. Indeed, if $E[X] = \alpha$, then $E[\alpha] = E[X]$. Hence, in the mean, this is the best possible approximation of the $\mathcal{F}$-measurable random variable X by a constant α (which means that we only have trivial information, see Remark 2.1.4 below for a precise statement). Now suppose we know more, we have the information $\mathcal{G}$ and our aim is to find a $\mathcal{G}$-measurable estimator of X. Then the conditional expected value is the best possible approximation in the mean, because it satisfies condition (2.1) above which means that X and the estimator Y are equal in the mean on $\mathcal{G}$. We will see later on that it corresponds to the "least squares" approximation (under appropriate integrability conditions), see Theorem 2.2.1.

Example 2.1.3. Given is the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ where $\mathcal{B}([0,1])$ are the Borel sets on $[0,1]$ and λ is the Lebesgue measure restricted to $[0,1]$ (which is a probability measure). The function $X(\omega) = \omega^2$ is measurable with respect to $\mathcal{F} = \mathcal{B}([0,1])$. Define the smaller σ-algebra $\mathcal{G}$ which is given by

$$\mathcal{G} = \left\{ \emptyset, [0,1], \left[0, \frac{1}{2}\right], \left(\frac{1}{2}, 1\right] \right\}.$$

Then X is not $\mathcal{G}$-measurable. Indeed, for $\alpha = \frac{1}{16}$ we have that

$$\{\omega : X(\omega) \leq \alpha\} = \left\{ \omega : X(\omega) \leq \frac{1}{16} \right\} = \left[0, \frac{1}{4}\right] \notin \mathcal{G}.$$

Remark 2.1.4. Suppose that the σ-algebra $\mathcal{G}$ is trivial, i.e.,

$$\mathcal{G} = \{\emptyset, \Omega\}.$$

Then conditional expected value with respect to the trivial information $\mathcal{G}$ is equal to the usual expected value (that is a constant), i.e.,

$$E[X|\mathcal{G}] = E[X] \text{ a.s.}$$

Indeed, we have to show that $Y = E[X]$ is $\mathcal{G}$-measurable, this means for each $\alpha \in \mathbb{R}$ it has to hold that $\{\omega : Y(\omega) \leq \alpha\} \in \mathcal{G}$. As $\mathcal{G}$ is trivial this means

$$\{Y \leq \alpha\} = \emptyset \text{ or } \Omega.$$

Hence $Y = c$ a.s. for some constant $c \in \mathbb{R}$. By Condition (2.1) applied to $A = \Omega \in \mathcal{G}$ we see that

$$c = E[c \mathbb{1}_\Omega] = E[X \mathbb{1}_\Omega] = E[X],$$

and so $Y = E[X|\mathcal{G}] = c = E[X]$ a.s.

Proof of the uniqueness almost surely. Let us now prove the uniqueness a.s. of the conditional expected value (as was claimed in Definition 2.1.1). Suppose Y and Z are two integrable $\mathcal{G}$-measurable random variables that satisfy condition 2.1. By linearity of the usual expected value we get, for all $A \in \mathcal{G}$,

$$E[(Y - Z)\mathbb{1}_A] = E[Y \mathbb{1}_A] - E[Z \mathbb{1}_A] = E[X \mathbb{1}_A] - E[X \mathbb{1}_A] = 0, \tag{2.2}$$

where the second equality holds because of condition 2.1 applied to Y and Z. Assume now that $P(Y \neq Z) > 0$. Without loss of generality we can assume that $P(Y > Z) > 0$. It holds that

$$\{Y > Z\} = \bigcup_{n=1}^{\infty} \left\{ Y > Z + \frac{1}{n} \right\},$$

and the sets $\{Y > Z + \frac{1}{n}\}$ are increasing, i.e., $\{Y > Z + \frac{1}{n}\} \uparrow \{Y > Z\}$. Hence it holds that

$$\lim_{n\to\infty} P\left(Y > Z + \frac{1}{n}\right) = P(Y > Z) > 0.$$

It follows that there exists n_0 and $\epsilon > 0$ such that $P(Y > Z + \frac{1}{n_0}) = \epsilon$. Now, as Y and Z are $\mathcal{G}$-measurable random variables, the set $A = \{Y - Z > \frac{1}{n_0}\}$ is contained in $\mathcal{G}$ and $P(A) = \epsilon > 0$. Now we apply equation 2.2 above to this A to get a contradiction:

$$0 = E[(Y-Z)\mathbb{1}_A] = E[(Y-Z)\mathbb{1}_{\{Y-Z>\frac{1}{n_0}\}}] > \frac{1}{n_0} P\left(Y > Z + \frac{1}{n_0}\right) = \frac{1}{n_0}\epsilon > 0.$$

Hence it has to hold that $P(Y = Z) = 1$. □

Let us study some easy examples first.

Example 2.1.5. $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be measurable and integrable. Suppose that there exists a partition of Ω into countably many measurable sets, i.e., $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$ where $\Omega_k \in \mathcal{F}$ and $\Omega_i \cap \Omega_j = \emptyset$, for all $i \neq j$. Suppose the σ-algebra $\mathcal{G}$ is generated by those sets, that means $\mathcal{G} = \sigma\{\Omega_k, k = 1, 2, \ldots, \infty\}$. We will calculate $E[X|\mathcal{G}]$. By definition, $E[X|\mathcal{G}]$ has to be $\mathcal{G}$-measurable, therefore it has to be of the following form:

$$E[X|\mathcal{G}] = \sum_{k=1}^{\infty} \alpha_k \mathbb{1}_{\Omega_k} \text{ a.s.},$$

for some $\alpha_k \in \mathbb{R}$, $k = 1, \ldots, \infty$. Moreover, for all i and $A = \Omega_i \in \mathcal{G}$ we get by (2.1) that

$$\begin{aligned} E[X\mathbb{1}_A] = E[X\mathbb{1}_{\Omega_i}] &= E\left[\left(\sum_{k=1}^{\infty} \alpha_k \mathbb{1}_{\Omega_k}\right) \mathbb{1}_{\Omega_i}\right] \\ &= E\left[\sum_{k=1}^{\infty} \alpha_k \mathbb{1}_{\Omega_k \cap \Omega_i}\right] \\ &= \alpha_i E[\mathbb{1}_{\Omega_i}] = \alpha_i P(\Omega_i). \end{aligned}$$

This holds because $\Omega_k \cap \Omega_i = \emptyset$ if $k \neq i$. Hence we get

$$\alpha_i = \frac{E[X \mathbb{1}_{\Omega_i}]}{P(\Omega_i)},$$

and we found the following formula for the conditional expected value in this example:

$$E[X|\mathcal{G}] = \sum_{k=1}^{\infty} \frac{E[X \mathbb{1}_{\Omega_k}]}{P(\Omega_k)} \mathbb{1}_{\Omega_k} \text{ a.s.}$$

Example 2.1.6. Let's apply the formula of Example 2.1.5 to the concrete example 2.1.3. We have that $\Omega = \Omega_1 \cup \Omega_2 = [0, \frac{1}{2}) \cup [\frac{1}{2}, 1]$ and so

$$E[X|\mathcal{G}] = \alpha_1 \mathbb{1}_{\left[0,\frac{1}{2}\right)} + \alpha_2 \mathbb{1}_{\left[\frac{1}{2},1\right]} \text{ a.s.},$$

where

$$\begin{aligned}
\alpha_1 &= \tfrac{1}{P\left(\left[0,\frac{1}{2}\right)\right)} E[X \mathbb{1}_{\left[0,\frac{1}{2}\right)}] = 2\int_0^{\frac{1}{2}} \omega^2 d\lambda(\omega) = 2\tfrac{\omega^3}{3}\Big|_0^{\frac{1}{2}} = \frac{1}{12} \\
\alpha_2 &= \tfrac{1}{P\left(\left[\frac{1}{2},1\right]\right)} E[X \mathbb{1}_{\left[\frac{1}{2},1\right]}] = 2\int_0^{\frac{1}{2}} \omega^2 d\lambda(\omega) = 2\tfrac{\omega^3}{3}\Big|_{\frac{1}{2}}^{1} = \frac{7}{12}.
\end{aligned}$$

Example 2.1.7 (Connection to the conditional density). Suppose the random vector (X, Y) has the bivariate density $f(x, y)$. Let

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{+\infty} f(x, y)dy \\
f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y)dx,
\end{aligned}$$

then $f_X(x)$ is the density of X and $f_Y(y)$ is the density of Y. The conditional density of X given $Y = y$ is given by

$$f(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ 0 & \text{else} \end{cases}$$

Define now the function $g(y)$, $y \in \mathbb{R}$, as follows

$$g(y) = \int_{-\infty}^{+\infty} x f(x|y)dx. \tag{2.3}$$

If we now plug the random variable Y into the function $g(\cdot)$ of (2.3) then it holds that

$$g(Y) = E[X|Y] \text{ a.s.},$$

where $E[X|Y] = E[X|\mathcal{G}]$ with $\mathcal{G} := \sigma(Y)$. Proof, see Exercise 1.8.

Theorem 2.1.8 (A list of properties of conditional expectation)**.** *In the following let X, Z, X_n, $n \geq 1$, be integrable random variables on $(\Omega, \mathcal{F}, P)$. The following holds*

(a) *If X is $\mathcal{G}$-measurable, then $E[X|\mathcal{G}] = X$ a.s.*

(b) *Linearity: $E[aX + bZ|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Z|\mathcal{G}]$ a.s.*

(c) *Positivity: If $X \geq 0$ a.s. then $E[X|\mathcal{G}] \geq 0$ a.s.*

(d) *Monotone convergence (MON): If $0 \leq X_n$ and $X_n \uparrow X$ a.s. then*

$$E[X_n|\mathcal{G}] \uparrow E[X|\mathcal{G}] \quad a.s..$$

(e) *Fatou: If $0 \leq X_n$, for all n, then*

$$E[\liminf_n X_n|\mathcal{G}] \leq \liminf_n E[X_n|\mathcal{G}].$$

(f) *Dominated convergence (DOM): If $|X_n| \leq Z$, for all $n \geq 1$ and $Z \in L^1$ (i.e. integrable) and $\lim X_n = X$ a.s. then*

$$\lim_n E[X_n|\mathcal{G}] = E[X|\mathcal{G}] \quad a.s.$$

(g) *Jensen's inequality: Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\varphi(X)$ is integrable. Then*

$$\varphi\left(E[X|\mathcal{G}]\right) \leq E[\varphi(X)|\mathcal{G}]$$

(h) *Tower Property: If $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$ then*

$$E\left[E[X|\mathcal{G}_2]|\mathcal{G}_1\right] = E[X|\mathcal{G}_1] \quad a.s.$$

(i) *Taking out what is known: Let Y be $\mathcal{G}$-measurable and let $E[|XY|] < \infty$. Then*

$$E[XY|\mathcal{G}] = YE[X|\mathcal{G}] \text{ a.s.}$$

(j) *Role of independence: If X is independent of $\mathcal{G}$, then*

$$E[X|\mathcal{G}] = E[X] \text{ a.s.}$$

Proof. (a) is trivial, check it as an exercise.

(b) Let $\tilde{X} = E[X|\mathcal{G}]$ a.s. and $\tilde{Z} = E[Z|\mathcal{G}]$ a.s. $\mathcal{G}$-measurability and integrability of $a\tilde{X} + b\tilde{Z}$ is clear. Then, for $A \in \mathcal{G}$, we have that

$$\begin{aligned} E[\mathbb{1}_A(a\tilde{X} + b\tilde{Z})] &= aE[\mathbb{1}_A\tilde{X}] + bE[\mathbb{1}_A\tilde{Z}] \\ &= aE[\mathbb{1}_A X] + bE[\mathbb{1}_A Z] \\ &= E[\mathbb{1}_A(aX + bZ)], \end{aligned}$$

where the first and the third equality follow because of linearity of the usual expected value and the second equality follows by property (2.1) applied to $\tilde{X}$ and $\tilde{Y}$. Hence, again by property (2.1) and by uniqueness a.s. of the conditional expected value it follows that

$$a\tilde{X} + b\tilde{Z} = E[aX + bZ|\mathcal{G}] \text{ a.s.}$$

(c) Let $X \geq 0$ a.s. Let $Y = E[X|\mathcal{G}]$ a.s. Suppose (c) would not hold, i.e., $P(Y < 0) > 0$. Let $A_n = \{Y < -\frac{1}{n}\}$, then, because by definition, Y is $\mathcal{G}$-measurable we have that $A_n \in \mathcal{G}$, for all $n \geq 1$. Moreover $A_n \uparrow A = \{Y < 0\}$ which means that $A_n \subset A_{n+1}$, for all n, and $A = \cup_{n=1}^{\infty}\{Y < -\frac{1}{n}\}$. It follows that

$$\lim_{n\to\infty} P(A_n) = P(A) = P(Y < 0) > 0.$$

Hence there exists $\epsilon > 0$ and n_0 such that

$$P(A_{n_0}) = \epsilon > 0.$$

Hence, because $X \geq 0$ a.s. and $A_{n_0} \in \mathcal{G}$, it follows by (2.1) that

$$0 \leq E[X\mathbb{1}_{A_{n_0}}] = E[Y\mathbb{1}_{A_{n_0}}] = E[Y\mathbb{1}_{\{Y<-\frac{1}{n_0}\}}] < -\frac{1}{n_0}P(A_{n_0}) = -\epsilon\frac{1}{n_0} < 0,$$

which is a contradiction. Hence $Y \geq 0$ a.s. and (c) holds.

(d), (e), (f), (g), (h), (j) are left as exercises, see Exercises 1.10, 1.11, 1.12, 1.13, 1.4, 1.5, respectively.

(i) is Exercise 1.6. Hint: you can assume first that $Y \geq 0$ a.s., and, in particular, that $Y = \mathbb{1}_A$ for an $A \in \mathcal{G}$. For the indicator the statement follows by definition. Now use standard machinery of probability theory to approximate $Y \geq 0$ by sums of indicators in a monotone way and use monotone convergence. Then use linearity to get the result for general $\mathcal{G}$-measurable $Y = Y^+ - Y^-$. □

2.2 Existence of conditional expected value

Theorem 2.2.1. *Let $X \in L^2(\Omega, \mathcal{F}, P)$ that means X is $\mathcal{F}$-measurable and $E[|X|^2] < \infty$. Let $\mathcal{G} \subset \mathcal{F}$ be a smaller σ-algebra. Then the conditional expected value Y ($= E[X|\mathcal{G}]$ a.s.) exists and Y is the orthogonal projection of X onto the subspace $L^2(\Omega, \mathcal{G}, P)$. That means*

$$\|X - Y\|_{L^2} = \min_{W \in L^2(\Omega,\mathcal{G},P)} \|X - W\|_{L^2}$$

and

$$E[(X - Y)Z] = 0 \text{ for all } Z \in L^2(\Omega, \mathcal{G}, P).$$

(We use the notation $\| \cdot \|_{L^2} = E[| \cdot |^2]^{\frac{1}{2}}$ in the above.)

From the above theorem (which has the additional condition of square integrability of X) we get as a corollary the existence of conditional expected value as given in Definition 2.1.1.

Corollary 2.2.2. *Let $X \in L^1(\Omega, \mathcal{F}, P)$ (that is as in Definition 2.1.1). Let $\mathcal{G} \subset \mathcal{F}$ be a smaller σ-algebra. Then the conditional expected value $E[X|\mathcal{G}]$ exists.*

Let us first give the proof of Corollary 2.2.2.

Proof of Corollary 2.2.2. Let X be $\mathcal{F}$-measurable and $E[|X|] < \infty$, i.e., X as in Definition 2.1.1. We can decompose X into its positive and its negative part $X = X^+ - X^-$ where $X^+ = X\mathbb{1}_{\{X \geq 0\}}$ and $X^- = X\mathbb{1}_{\{X \leq 0\}}$. This shows, by linearity of conditional expectation, that it is enough to prove the result for $X \geq 0$ a.s. Such an X we can approximate by a sequence of bounded random variables $X_n \uparrow X$. As each X_n is bounded it is in $L^2(\Omega, \mathcal{F}, P)$. So, for each $n \geq 1$, we can apply Theorem 2.2.1 to see that, for each $n \geq 1$, there exists the conditional expected value $Y_n = E[X_n|\mathcal{G}]$ a.s. By assumption it holds that $0 \leq X_n \leq X_{n+1}$ a.s. for all n and so, by property c) of Theorem 2.1.8 (properties of conditional expected value) it holds that $0 \leq Y_n \leq Y_{n+1}$ a.s., that means $Y_n \uparrow$. Define

$$Y(\omega) = \limsup_n Y_n(\omega),$$

then, as all Y_n are $\mathcal{G}$-measurable, we have that Y is $\mathcal{G}$-measurable as well. Take now any $A \in \mathcal{G}$. Then the following holds:

$$E[\mathbb{1}_A Y] = \lim_{n \to \infty} E[\mathbb{1}_A Y_n] \tag{2.4}$$

$$= \lim_{n \to \infty} E[\mathbb{1}_A X_n] \tag{2.5}$$

$$= E[\mathbb{1}_A X], \tag{2.6}$$

where (2.4) holds because of monotone convergence for the usual expected value, (2.5) holds because of the conditional expected value property (2.1) applied to each Y_n and (2.6) holds again because of monotone convergence for the usual expected value. In particular we can choose $A = \Omega \in \mathcal{G}$, then the equality $E[X] = E[Y]$ (for $X \geq 0$) gives that $E[|Y|] < \infty$ (because $Y \geq 0$ a.s. as well and hence $Y = |Y|$ a.s.). So we proved existence of the conditional expected value $Y = E[X|\mathcal{G}]$ a.s. $\square$

Now we will proceed with the proof of Theorem 2.2.1. We need the following lemma about completeness of L^2-spaces which you might know from the lecture Probability Theory (if not, then at least you have enough tools there to deduce the proof from completeness with respect to the convergence in probability plus some additional ideas), see Exercise 1.14. And, the result can be found in many books.

Lemma 2.2.3 (Completeness of L^2)**.** *Let $(Y_n)_{n\geq 1}$ be a Cauchy sequence in $L^2(\Omega, \mathcal{G}, P)$, that means, for each $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $l, m \geq N$:*

$$\|Y_l - Y_m\|_{L^2} = E[(Y_l - Y_m)^2]^{\frac{1}{2}} < \epsilon.$$

Then the sequence $(Y_n)_{n\geq 1}$ has a limit in $L^2(\Omega, \mathcal{G}, P)$, i.e., there exists $Y \in L^2(\Omega, \mathcal{G}, P)$ such that

$$\lim_{n\to\infty} \|Y_n - Y\|_{L^2} = \lim_{n\to\infty} E[(Y_n - Y)^2]^{\frac{1}{2}} = 0.$$

Now we will prove a more general version of Theorem 2.2.1 for general complete linear subspaces $\mathcal{K}$ which we then will apply to $\mathcal{K} = L^2(\Omega, \mathcal{G}, P)$ which is complete by the above lemma.

Theorem 2.2.4. *Let $\mathcal{K}$ be a linear subspace of $L^2(\Omega, \mathcal{F}, P)$ that is complete (which means that every L^2-Cauchy sequence in $\mathcal{K}$ has a limit which is an element of $\mathcal{K}$). Then for every $X \in L^2(\Omega, \mathcal{F}, P)$ there exists $Y \in \mathcal{K}$ such that*

(i) $\|X - Y\|_{L^2} = \min_{W\in\mathcal{K}} \|X - W\|_{L^2}$ *and*

(ii) $E[(X - Y)Z] = 0$ *for all $Z \in \mathcal{K}$.*

Moreover, Y is unique with respect to equality a.s. and is called the orthogonal projection of X onto the subspace $\mathcal{K}$.

Proof. For the proof of (i) define

$$\alpha = \inf\{\|X - W\|_{L^2} : W \in \mathcal{K}\}. \tag{2.7}$$

Choose a sequence $(Y_n)_{n\geq 1} \in \mathcal{K}$ with $\|X - Y_n\|_{L^2} \to \alpha$ for $n \to \infty$. We will show that $(Y_n)_{n\geq 1}$ is an L^2-Cauchy sequence. Indeed, fix l, m large and define $U = X - \frac{1}{2}(Y_l + Y_m)$ and $V = \frac{1}{2}(Y_l - Y_m)$ then:

$$\begin{aligned}\|X - Y_m\|_{L^2}^2 + \|X - Y_l\|_{L^2}^2 &= \|U + V\|_{L^2}^2 + \|U - V\|_{L^2}^2 \\ &= 2\|U\|_{L^2}^2 + 2\|V\|_{L^2}^2 \\ &= 2\|X - \tfrac{1}{2}(Y_l + Y_m)\|_{L^2}^2 + 2\|\tfrac{1}{2}(Y_l - Y_m\|_{L^2}^2 \\ &\geq 2\alpha^2 + \tfrac{2}{4}\|Y_l - Y_m\|_{L^2}^2,\end{aligned}$$

where the first equality holds by definition of U and V, the second equality is a trivial calculation, the third equality again the definition of U, V. The inequality holds because $\mathcal{K}$ is a linear subspace, therefore, as Y_l, $Y_m \in \mathcal{K}$ we have that $\frac{1}{2}(Y_l + Y_m) \in \mathcal{K}$ as well and hence $\|X - \frac{1}{2}(Y_l + Y_m)\|_{L^2} \geq \alpha$. Now, let's write the relevant inequality down again:

$$\|X - Y_m\|_{L^2}^2 + \|X - Y_l\|_{L^2}^2 \geq 2\alpha^2 + \tfrac{1}{2}\|(Y_l - Y_m\|_{L^2}^2. \tag{2.8}$$

If we let $m \to \infty$ and $l \to \infty$ it is clear by the definition of the sequence $(Y_n)_{n\geq 1}$ that the left hand side of inequality (2.8) converges to $2\alpha^2$. As we have $2\alpha^2 + \frac{1}{2}\|Y_l - Y_m\|_{L^2}^2$ on the right hand side of the inequality this is only possible if

$$\lim_{l,m\to\infty} \|Y_l - Y_m\|_{L^2} \to 0,$$

and we proved that $(Y_n)_{n\geq 1}$ is a Cauchy sequence in $\mathcal{K}$.

By assumption $\mathcal{K}$ is complete therefore there exists a limit $Y \in \mathcal{K}$ such that $\lim_{n\to\infty} \|Y_n - Y\|_{L^2} = 0$. As $Y \in \mathcal{K}$ we have that $\|X - Y\|_{L^2} \geq \alpha$ and so

$$\alpha \leq \|X - Y\|_{L^2} = \|X - Y_n + Y_n - Y\|_{L^2} \leq \|X - Y_n\|_{L^2} + \|Y_n - Y\|_{L^2},$$

where the right hand side converges to α as $\|X - Y_n\|_{L^2} \to \alpha$ and $\|Y_n - Y\|_{L^2} \to 0$. Therefore we see that

$$\|X - Y\|_{L^2} = \alpha,$$

and the inf in equation (2.7) is a minimum which is attained in Y. So we proved (i).

For the proof of (ii) let $Z \in \mathcal{K}$, $Z \neq 0$. Then, for every $t \in \mathbb{R}$, we have that $Y + tZ \in \mathcal{K}$, because $\mathcal{K}$ is a linear space (Y is the one from (i)). By definition of Y it holds that

$$\|X - (Y + tZ)\|_{L^2}^2 \geq \|X - Y\|_{L^2}^2.$$

Calculating the left and the right hand side of the above inequality we see that

$$E[(X-Y)^2] - 2tE[(X-Y)Z] + t^2E[Z^2] \geq E[(X-Y)^2].$$

Hence, if we define $a = E[(X-Y)Z]$ and $b = E[Z^2]$ we see that, for all $t \in \mathbb{R}$,

$$-2at + bt^2 \geq 0, \tag{2.9}$$

where we know that $b = E[Z^2] > 0$. Let $f(t) = -2at + bt^2$, with $b > 0$. Then this quadratic function satisfies $f(t) = 0$ for $t_1 = 0$ and $t_2 = \frac{2a}{b}$. Suppose $a \neq 0$ then $t_1 \neq t_2$ and the function becomes strictly negative for any t which is strictly between t_1 and t_2. Hence the condition (2.9) is not satisfied, which is a contradiction. Therefore

$$a = E[(X-Y)Z] = 0.$$

As this can be done for every $Z \in \mathcal{K}$, we proved (ii). □

Proof of Theorem 2.2.1. Apply Theorem 2.2.4 for the linear subspace $\mathcal{K} = L^2(\Omega, \mathcal{G}, P)$ to find the orthogonal projection Y of X onto $L^2(\Omega, \mathcal{G}, P)$. Let $A \in \mathcal{G}$, then clearly $\mathbb{1}_A \in L^2(\Omega, \mathcal{G}, P)$. Hence by (ii) of Theorem 2.2.4 it holds that $E[(X-Y)\mathbb{1}_A] = 0$ and therefore

$$E[X\mathbb{1}_A] = E[Y\mathbb{1}_A],$$

and therefore Y satisfies condition 2.1 of the definition of conditional expected value (Definition 2.1.1). Hence $Y = E[X|\mathcal{G}]$ a.s. □

Chapter 3

Stopping times, martingales and the stochastic integral in discrete time

3.1 Random times and stopping times

Given is a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}}, P)$.

Definition 3.1.1. *A random time is an $\mathcal{F}$-measurable random variable $T : \Omega \to [0, +\infty]$. (See Remark 1.2 for what is meant by $\mathcal{F}$-measurable.)*

You already know the concept of random times from the lecture about Markov chains. You also know that not all random times are "nice". "Nice" means that it can be decided at time t if they have happened before a t or not, when we observe a certain stochastic process (or we have the knowledge of a certain filtration) up to time t. Those "nice" random times are defined below and are called stopping times. Why "stopping" times? Because they can be used to define a new stochastic process (the stopped process) that is only observed until the random time T, after that it stays constant. This is useful if we are interested in a stochastic process until a certain event occurs (which gives rise to a stopping time) and after the event has occurred the further development of the process is not observed any more.

Definition 3.1.2. *A* **stopping time** *T for the filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ is a random time T that has the following property: for all $t \in \mathbb{T}$, we have that $\{T \leq t\} \in \mathcal{F}_t$.*

If we think of the generated filtration of a stochastic process $(X_t)_{t\in\mathbb{T}}$ this describes in a mathematically precise way the following: if we observe the stochastic process up to time t, we know if $T \leq t$ or $T > t$, so we know if the event describing T happened before time t or not. Later we can also use stopping times as kind of an alarm clock that tell us, that, directly after T happened, to do a certain transaction (for example when trading with financial assets).

Remark 3.1.3. Let us specify the situation in discrete time. If we have the time set $\mathbb{T} = \mathbb{N} \cup \{0\}$, then also the stopping time should only take discrete values, i.e., $T : \Omega \to \mathbb{N} \cup \{0\} \cup \{+\infty\}$. Hence the condition in Definition 3.1.2 amounts to the following: a discrete random time T is a stopping time if and only if, for all $n \geq 0$,

$$\{T = n\} \in \mathcal{F}_n.$$

Let us prove this fact:
($\Rightarrow$) T is a stopping time if and only if $\{T \leq n\} \in \mathcal{F}_n$, for all $n \in \mathbb{T}$ (by Definition 3.1.2. But $\{T = n\} = \{T \leq n\} \setminus \{T < n\}$ and, as we consider only integers we have that $\{T < n\} = \{T \leq n-1\}$ which is $\in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$. Therefore $\{T = n\} \in \mathcal{F}_n$.
($\Leftarrow$) Suppose now that $\{T = n\} \in \mathcal{F}_n$ for all n. We have to show that T is a stopping time as in Definition 3.1.2. I.e., we have to show that $\{T \leq n\} \in \mathcal{F}_n$, for all n. Indeed: $\{T \leq n\} = \bigcup_{k=0}^{n} \{T = k\} \in \mathcal{F}_n$, because $\{T = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$, for all $k \leq n$. And the proof is finished.

In the lecture of Markov chains we already saw some examples of stopping times, for example the first hitting time H^A of a subset A of the state space I and the first passage time T_i of a state i (i.e.the first moment when the process is in the state i).

In continuous time, i.e., if $\mathbb{T} = [0, \infty)$ there occur some subtle troubles concerning filtrations and stopping times. We need some extra condition on filtrations (in order to be able to prove helpful results around important stochastic processes in continuous time, such as Brownian motion). These conditions are so widely used and important that they got the name "the usual conditions". Here they are:

Definition 3.1.4. *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration with respect to the time set $\mathbb{T} = [0, \infty)$. We say that the filtration satisfies the* **usual conditions** *if the following holds:*

(1) $(\mathcal{F}_t)_{t\geq 0}$ is complete, that is all null-sets (sets $A \in \mathcal{F}$ with $P(A) = 0$) are already in $\mathcal{F}_0$.

(2) $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous, that is $\mathcal{F}_t^+ = F_t$, where

$$\mathcal{F}_t^+ = \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}.$$

The right continuity means that the information **immediately after** t already belongs to $\mathcal{F}_t$.

Remark 3.1.5 (Notation). Note that whenever the time set is $\mathbb{T} = [0, +\infty)$, instead of writing $t \in \mathbb{T}$ we can alternatively write $t \in [0, \infty)$ or just $t \geq 0$.

Now here is a property that gives us more ways to prove that a random time is a stopping time (at least under the usual conditions):

Lemma 3.1.6. (i) *If T is a stopping time on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)})$ then, for all $t \geq 0$, $\{T < t\} \in \mathcal{F}_t$.*

(ii) *Conversely, if T is a random time on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)})$ and $\{T < t\} \in \mathcal{F}_t$, for all $t \geq 0$* **and, additionally** *the filtration $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous, then T is a stopping time.*

Proof. Exercises 4.3 and 4.4. □

Further properties of stopping times:

Lemma 3.1.7. *Let S and T be stopping times for a given filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$. Then $S \wedge T = min(S, T)$, $S \vee T = max(S, T)$ and $S + T$ are stopping times for the same filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$.*

Proof. Exercise 2.1. Hint: if $\mathbb{T} = \{0, 1, 2, \dots\}$ is discrete the proof of $S + T$ being a stopping time is very easy. If $\mathbb{T} = [0, +\infty)$ the proof is a bit more tricky. □

Let us give the definition of a Brownian motion (short: BM) here.

Definition 3.1.8 (BM, standard BM)**.** *An adapted stochastic process $(W_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ is a Brownian motion if*

- *For $s < t$ the increment $W_t - W_s$ has a normal distribution with mean 0 and variance $t - s$, i.e., $\mathcal{N}(0, t - s)$.*
- *For $s < t$ the increment $W_t - W_s$ is independent of the past (i.e., independent of $\mathcal{F}_s$).*
- *BM has continuous paths, that means, $t \to W_t(\omega)$ is continuous for almost all ω.*

If the BM satisfies additionally that $W_0 = 0$ then it is called a **standard** *BM.*

Here is a concrete example of a stopping time with time set $\mathbb{T} = [0, \infty)$ that is very important in the study of BM.

Example 3.1.9. Given is a standard BM. Let $a > 0$ and define

$$T_a(\omega) = \inf\{t > 0 : W_t(\omega) \geq a\} = \inf\{t > 0 : W_t(\omega) = a\}.$$

The time T_a is the first passage time of the level a. As Brownian paths are continuous functions we have that $W_{T_a} = a$ a.s., because the BM becomes bigger than the level a by crossing it: it cannot jump over it and end up higher than a before actually hitting it.

Lemma 3.1.10. *T_a is a stopping time,i.e., for all $t \geq 0$: $\{T_a \leq t\} \in \mathcal{F}_t$.*

Proof. Exercise 4.5 □

3.2 Martingales

We will now introduce a stochastic process with certain properties that reflect that it behaves as a "fair game". A fair game means that in expectation your gains and losses are always zero. How does the mathematical description of that look like?

Definition 3.2.1. *A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}}, P)$ is a* **martingale** *if X is an adapted stochastic process which is*

(i) **integrable**, *i.e., $E[|X_t|] < \infty$ for all $t \in \mathbb{T}$ and*

(ii) *$E[X_t|\mathcal{F}_s] = X_s$ a.s. for all $s < t$, i.e. X satisfies the* **martingale property**.

We will see in Remark 3.2.2, (ii), that the martingale property ensures that the gain/loss from time s up to time t, which is just the increment $X_t - X_s$, is 0 in expectation, given that we use all the information that we have at time s. In the game interpretation this would mean that at time s we put a stake of 1 into a game and our gain at time t is then $1(X_t - X_s)$. In our decision we can only include the information that we have at time s because we do our bet at time s. And, for a martingale, we know that with this information we expect to neither lose nor gain (i.e. =0), hence it is a fair game, the gains and losses cancel each other out in the mean. See Example 3.3.1 below for a more detailed game interpretation that leads to the definition of the stochastic integral in discrete time.

Remark 3.2.2. (i) If X is a martingale, then $E[X_t] = E[X_0]$ for all $t \in \mathbb{T}$. This immediately follows from the tower property as $E[X_t] = E[E[X_t|\mathcal{F}_0]] = E[X_0]$. If, moreover $\mathcal{F}_0 = \{\emptyset, \Omega\}$ then $E[X_t] = X_0$.

(ii) The martingale property is equivalent to $E[X_t - X_s|\mathcal{F}_s] = 0$ a.s.

Indeed, clearly $E[X_t|\mathcal{F}_s] - X_s = 0$ a.s. Now as X_s is measurable with respect to $\mathcal{F}_s$ (because a martingale is by definition adapted) we can take X_s inside the conditional expectation and have: $E[X_t|\mathcal{F}_s] - E[X_s|\mathcal{F}_s] = 0$ a.s. By linearity we get $E[X_t - X_s|\mathcal{F}_s] = 0$ a.s.

(iii) If the time set is discrete, i.e., $\mathbb{T} = \{0, 1, 2, ...\}$, then the martingale property says $E[X_n|\mathcal{F}_{n-1}] = X_{n-1}$ a.s. for all $n \geq 1$.

Example 3.2.3. The random walk, see Example 1.1.2, is a martingale for its generated filtration if we assume that the iid random variables are integrable and have mean 0, i.e., $E[|\epsilon_k|] < \infty$ and $E[\epsilon_k] = 0$. (This is, for example, the case if we use the particular choice where ϵ_k takes the values ± 1 each with probability $\frac{1}{2}$.)

Let's prove that the random walk as above is a martingale.
1) Integrability: $X_n = \sum_{k=1}^n \epsilon_k$, hence $E[|X_n|] \leq \sum_{k=1}^n E[|\epsilon_k|] < \infty$ because each summand is finite and it's a sum of finitely many (n) summands.
2) Martingale property: we use Remark 3.2.2, (2), to observe that it is enough to show: $E[X_n - X_{n-1}|\mathcal{F}_{n-1}] = 0$ a.s., for all n. We have that

$$E[X_n - X_{n-1}|\mathcal{F}_{n-1}] = E[\epsilon_n|\mathcal{F}_{n-1}] = E[\epsilon_n] = 0 \text{ a.s.},$$

where the first equality is clear by the definition of X_n and X_{n-1} as sums, the second equality holds because $\mathcal{F}_{n-1} = \sigma\{\epsilon_1, \ldots, \epsilon_{n-1}\}$ and ϵ_n is independent of all ϵ_k, $k = 1, \ldots, n-1$. The last equality holds by assumption (zero moment for the iid sequence). So the martingale property holds.

Example 3.2.4. Brownian motion is a martingale.
Let's do the proof for standard BM, so assume $W_0 = 0$.
1) Integrability: first observe that $W_t = W_t - 0 = W_t - W_0$ and hence $W_t \sim \mathcal{N}(0, t)$. By the properties of normal distribution we know that $E[|W_t|] < \infty$, for all t.
2) Martingale property: again we use Remark 3.2.2, (2), and see that

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

where the first equality holds because $W_t - W_s$ is independent of $\mathcal{F}_s$. The second equality holds as $W_t - W_s \sim \mathcal{N}(0, t-s)$ and hence it has mean 0.

One can also study stochastic processes that have the property of being systematically unfair in one or the other direction. Either you always loose in the mean or either you always win in the mean. The respective processes are called supermartingales and submartingales.

Definition 3.2.5. *Let $X = (X_t)_{t\in\mathbb{T}}$ be an* **adapted** *and* **integrable** *stochastic process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}}, P)$. X is called*

(i) *supermartingale if $E[X_t|\mathcal{F}_s] \leq X_s$ a.s. for all $t \geq 0$,*

(ii) *submaringale if $E[X_t|\mathcal{F}_s] \geq X_s$ a.s. for all $t \geq 0$.*

Observe that a process X that is a submartingale and a supermartingale is a martingale.

Example 3.2.6. Find an example for a supermartingale and an example for a submartingale based on a random walk with respect to its generated filtration. What conditions on the iid sequence do you need?

3.3 Stochastic integral in discrete time

Example 3.3.1 (Stochastic integral in discrete time as accumulated gains/losses in a game). Let $\mathbb{T} = \mathbb{N} \cup \{0\}$ and $(X_n)_{n\in\mathbb{T}}$ be a stochastic process on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{T}}, P)$. As in the interpretation of a martingale as a fair game in the last section we again imagine we play the following game. At each time $n-1 = 0, 1, 2, \ldots$ we make a stake in a game (for example our stake is the amount of 1). Then the process evolves from X_{n-1} (at time $n-1$) to X_n at time n. We get the amount $X_n - X_{n-1}$ (if our stake was 1). So depending on the sign of the increment $X_n - X_{n-1}$ we make a gain or loss at time n.

Now let's suppose our stake is not 1. Suppose we decide at time $n-1$ to put a stake of H_n into the game, i.e., instead of making a gain/loss of $X_n - X_{n-1}$ at time n we get H_n times this amount: $H_n(X_n - X_{n-1})$. H_n depends on the state of the world $\omega \in \Omega$ so H_n will be a random variable. The question is on which information can we base our decision, that means, which σ-algebra is the one that we have "access" to? This is a game where we have to make our decision at time $n-1$. The outcome of the game is seen at time n. Of course we cannot use the information what the result of the game will be (otherwise this would no longer be a risky game). This means we cannot use the information at time n when the outcome is already known. Imagine a casino: you cannot make a bet on red if red is already the outcome, but you have to make your bet when the ball is still moving (and that is **strictly before** the outcome is known). In our case this means:

for a game that is decided at time n we have to make our decision strictly before that, that is, at time $n-1$. In mathematical terms this means that the size of the stake H_n for the game where the outcome is known at time n has to be a random variable with respect to the information **strictly before** n, which means, H_n has to be $\mathcal{F}_{n-1}$-measurable. Because $\mathcal{F}_{n-1}$ is exactly the information up to time $n-1$ and in our discrete time setting here $n-1$ is the biggest time **before** the time n.

Suppose we are playing this game from time $n = 1$ until time $n = N$, where N is some natural number. This means there is a game at each time $n = 1, 2, \ldots, N$. Our decisions H_n are made at time $n-1$, i.e., H_1 is $\mathcal{F}_0$-measurable, H_2 is $\mathcal{F}_1$-measurable, H_3 is $\mathcal{F}_2$-measurable,..., H_N is $\mathcal{F}_{N-1}$-measurable. This gives a new stochastic process $(H_n)_{n=1}^N$ with the property that H_n is $\mathcal{F}_{n-1}$-measurable for each $n = 1, \ldots, N$. A process like this will be called predictable because H_n with the time index n is already measurable with the information at time $n-1$. Now our gain/loss in the time n game is $H_n(X_n - X_{n-1})$. So when we play from time $n = 1$ until time N the sum of all our gains/losses is equal to

$$\sum_{n=1}^{N} H_n(X_n - X_{n-1}) \tag{3.1}$$

This sum describes our **accumulated gains/losses** if we play the game over the time period from 1 up to N. And the sum in (3.1) is exactly the stochastic integral in discrete time of the predictable stochastic process $(H_n)_{n=1}^N$ with respect to the adapted stochastic process $(X_n)_{n=0}^N$.

Now we will formally introduce the stochastic integral. But in fact all the ingredients can already be found in Example 3.3.1.

Definition 3.3.2. *Let the time set $\mathbb{T} = \mathbb{N} \cup \{0\}$. A discrete time process $(H_n)_{n\in\mathbb{N}}$ (i.e. $n \geq 1$) on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{T}}, P)$ (i.e. $n \geq 0$) is called predictable (or previsible) if H_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$.*

Definition 3.3.3 (Stochastic integral in discrete time)**.** *Let $\mathbb{T} = \mathbb{N} \cup \{0\}$. Let $(X_n)_{n\in\mathbb{T}}$ be an adapted stochastic process and $(H_n)_{n\in\mathbb{N}}$ be a predictable stochastic process on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{T}}, P)$. Then the discrete time stochastic integral at time N is given by*

$$\sum_{n=1}^{N} H_n(X_n - X_{n-1}) =: (H \cdot X)_N = \int_0^N H_n dX_n$$

Let's use the notation with the lower case n again, i.e. $(H \cdot X)_n = \sum_{k=1}^{n} H_n(X_k - X_{k-1})$ for each $n \geq 1$ and define $(H \cdot X)_0 = 0$. Then $((H \cdot X)_n)_{n\in\mathbb{T}}$ defines a new stochastic process, the stochastic integral process. Note that

$((H \cdot X)_n)_{n\in\mathbb{T}}$ *is an adapted stochastic process on the same filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{T}}, P)$

Exercise: Prove that $((H \cdot X)_n)_{n\in\mathbb{T}}$ really is adapted.

We will now study the case when X is a martingale. We will see that, under some technical conditions that ensure integrability, a stochastic integral over a martingale is again a martingale. Thinking of the game interpretation of the stochastic integral this is often referred to by the phrase "You can't win betting on a martingale." This means if the underlying process is a martingale, whatever your stakes H are, in the mean, your gains and losses will always be 0. Here is the corresponding theorem:

Theorem 3.3.4. *Let* $\mathbb{T} = \mathbb{N} \cup \{0\}$ *and let* $(H_n)_{n\in\mathbb{N}}$ *be a bounded and predictable stochastic process and* $(X_n)_{n\in\mathbb{T}}$ *be a martingale on a filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{T}}, P)$*. Then* $((H \cdot X)_n)_{n\in\mathbb{T}}$ *is a martingale.*

Proof. First note that the assumption on H of the theorem means the following: there exists a constant K such that $P(|H_n| \leq K) = 1$ for all $n = 1, 2, \dots,$ i.e., $|H_n| \leq K$ a.s. for all $n \in \mathbb{N}$. In order to prove that the stochastic integral process is a martingale we have to show first integrability for each n. Indeed,

$$\begin{aligned} E[|\sum_{k=1}^{n} H_k(X_k - X_{k-1})|] &\leq E[\sum_{k=1}^{n} |H_k(X_k - X_{k-1})|] \\ &\leq K \sum_{k=1}^{n} E[|X_k - X_{k-1}|] \\ &\leq K \sum_{k=1}^{n} (E[|X_k|] + E[|X_{k-1}|]) \\ &< \infty, \end{aligned}$$

where we use first triangle inequality then the fact that $|H_k| \leq K$ for each k, then triangle inequality for $|X_k - X_{k-1}| \leq |X_k| + |X_{k-1}|$ and then the fact that each $E[|X_k|] < \infty$ for all k (because X is a martingale and therefore integrable) and we use that the sum has a finite number of summands.

Now we prove the martingale property. Indeed,

$$\begin{aligned} E[(H \cdot X)_n - (H \cdot X)_{n-1} | \mathcal{F}_{n-1}] &= E[H_n(X_n - X_{n-1}) | \mathcal{F}_{n-1}] \\ &= H_n E[X_n - X_{n-1} | \mathcal{F}_{n-1}] = 0 \text{ a.s.}, \end{aligned}$$

where the first equality in line 2 is due to the fact that H_n is $\mathcal{F}_{n-1}$-measurable because H is predictable. Therefore we can use the "taking out what is known"-property of conditional expected value. The second equality in line 2 follows from the martingale property of X. $\square$

Exercise: Which technical conditions could you pose to get the martingale property of the stochastic integral instead of boundedness of H and only martingale property for X?

We have already seen that if (X_n) is a martingale then $E[X_n] = X_0$ (if $\mathcal{F}_0$ is the trivial σ-algebra). Now let's pose the question if this also holds when we replace the deterministic time n by a stopping time T taking values in $\mathbb{T} \cup \{+\infty\}$, where again $\mathbb{T} = \mathbb{N} \cup \{0\}$. The answer is: yes and no. Yes, if T is a bounded stopping time (that means T takes values in a set $\{0, 1, 2, \ldots, N\}$ for some N, i.e. $P(|T| \leq N) = 1$.) And no if T is an unbounded stopping time. We will see these two facts in a theorem and in a counterexample.

Theorem 3.3.5 (Doob's optional stopping Theorem)**.** *Suppose* $(X_n)_{n\in\mathbb{T}}$*, where* $\mathbb{T} = \mathbb{N} \cup \{0\}$ *(discrete time), is a martingale on* $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\in\mathbb{T}}, P)$ *and* T *is a bounded stopping time, i.e., there exists a constant* $N \in \mathbb{N}$ *such that*

$$P(\{\omega : T(\omega) \leq N\}) = 1.$$

Then $E[X_T] = E[X_0]$ *and if* $\mathcal{F}_0$ *is the trivial* σ*-algebra then* $E[X_T] = X_0$.

Assume that S *is a further stopping time with* $S \leq T$ *a.s. Then*

$$E[X_T|\mathcal{F}_S] = X_S \ \textit{a.s.}$$

Remark 3.3.6. (i) The random variable X_T is given as follows: $X_T(\omega) = X_n(\omega)$ if $T(\omega) = n$.

(ii) The stopping time σ-algebra is defined in Exercise 2.2 in the collection of Exercises. There you can also see some of its properties (and you should prove these properties).

(iii) Note that the equation $E[X_T|\mathcal{F}_S] = X_S$ a.s. means that, for bounded stopping times $S < T$ the martingale property holds also for these **bounded (!)** random times $S \leq T$ a.s.

Proof of Doob's Theorem. We will prove the first part of the theorem, the part about $S \leq T$ a.s.is left as an exercise (Exercise 2.6 in the collection of all exercises). By assumption T is a bounded stopping time in discrete time. This means T takes values in $\{0, 1, ..., N\}$ where N is a natural number, $N < \infty$.

Define, for each $n \geq 1$ the following random variable $H_n(\omega) = \mathbb{1}_{\{T(\omega) \geq n\}}$. It holds that

$$\{T \geq n\} = \Omega \setminus \{T < n\} = \Omega \setminus \{T \leq n-1\},$$

and $\Omega \setminus \{T \leq n-1\} \in \mathcal{F}_{n-1}$ and hence $\{T \geq n\} \in \mathcal{F}_{n-1}$. This implies that, for each $n \geq 1$, we have that H_n is $\mathcal{F}_{n-1}$-measurable. Therefore the stochastic process $(H_n)_{n\geq 1}$ is predictable. Moreover the process $(H_n)_{n\geq 1}$ is bounded (by 1), because each H_n is an indicator and indicators are of course bounded by 1. By Theorem 3.3.4 we know that the stochastic integral process over this integrand is a martingale, i.e., $((H \cdot X)_n)_{n\geq 0}$ is a martingale. This immediately implies that

$$E[(H \cdot X)_n] = (H \cdot X)_0 = 0, \tag{3.2}$$

for all $n \geq 1$, see Remark 3.2.2, (i). Fix now an arbitrary $n > N$. Let's look at $(H \cdot X)_n(\omega)$ for a fixed $\omega \in \Omega$. Assume that for this ω we have $T(\omega) = m$. By the boundedness assumption on T this has to satisfy $m \leq N$ (at least this holds a.s. therefore we take an ω in this probability-one set). Then

$$\begin{aligned}
(H \cdot X)_n(\omega) &= \sum_{k=1}^{n} H_k(\omega)(X_k - X_{k-1})(\omega) \\
&= \sum_{k=1}^{n} \mathbb{1}_{\{T(\omega)\geq k\}}(X_k(\omega) - X_{k-1}(\omega)) \\
&= \sum_{k=1}^{n} \mathbb{1}_{\{m\geq k\}}(X_k(\omega) - X_{k-1}(\omega)) \\
&= \sum_{k=1}^{m} (X_k(\omega) - X_{k-1}(\omega)) \\
&= X_m(\omega) - X_0(\omega) = X_T(\omega) - X_0(\omega),
\end{aligned}$$

where the last step needs the observation that $X_T(\omega) = X_m(\omega)$ because $T(\omega) = m$. And we also use that we chose n such that that $n > N \geq m$ in the equality before the last one.

If we now combine (3.2) with the last calculation we get

$$0 = E[(H \cdot X)_n] = E[X_T - X_0],$$

and the result follows, i.e., $E[X_T] = E[X_0]$ and for trivial $\mathcal{F}_0$ therefore $E[X_T] = X_0$. $\square$

Example 3.3.7 (Counterexample with an unbounded stopping time). Let $(X_n)_{n\geq 0}$ be a random walk as in Example 1.1.2, where $\epsilon_k \in \{1, -1\}$ with probability $\frac{1}{2}$, each. Recall the lecture about Markov processes. Let

$$T = \inf\{n \geq 1 : X_n = 1\},$$

i.e. T is the first passage time of the state 1. We saw in the lecture about Markov processes that

$$P(T < \infty) = 1,$$

hence it is a **finite** stopping time. But be careful: finite is not the same as bounded! In fact, we saw in the Markov processes lecture that $E[T] = +\infty$ and therefore T cannot be bounded. (Ask yourselves now as an exercise: why?)

Let's now look at $E[X_T]$. We have that on the set $\{T < \infty\}$ at time T the stochastic process is in the state 1 (because T is the first passage time of 1) and hence

$$E[X_T] = E[X_T \mathbb{1}_{\{T<\infty\}}] = E[1\mathbb{1}_{\{T<\infty\}}] = P(T < \infty) = 1.$$

Here the first equality holds because $P(T < \infty) = 1$ and the rest was explained above.

On the other hand the random walk starts in 0 with probability 1, hence $X_0 = 0$, and so

$$E[X_T] = 1 \neq 0 = X_0,$$

and the conclusion of Doob's Theorem does NOT hold. Why? Because T is only finite, but NOT bounded!

Chapter 4

Application: financial market with a finite Ω

For all the results and examples in this Chapter we refer, for example, to [5] and [6].

4.1 A first very easy example

We look at the easy situation of a sample space Ω consisting of only two possible states of the world, i.e. $\Omega = \{g, b\}$. Let's call the states $g =$ "good" and $b =$ "bad". Let the time set consist of times $t = 0$ and $t = 1$, i.e., $\mathbb{T} = \{0, 1\}$. As σ-algebra we take the power set $\mathcal{P}(\Omega) = \{\emptyset, \{g\}, \{b\}, \Omega\}$. Moreover we have a probability measure on $(\Omega, \mathcal{P}(\Omega))$ given by $P(g) = P(b) = \frac{1}{2}$.

We will now introduce a toy model of a financial market. There are two assets:

- a riskless asset (the bond) which is always equal to 1, that means $B_0 = B_1 = 1$,
- a risky asset (the stock) where $S_0 = 1$ and

$$S_1(\omega) = \begin{cases} 2 & \text{if } \omega = g \\ \frac{1}{2} & \text{if } \omega = b \end{cases}$$

The bond can be considered as a bank account (where the interest rate $r = 0$, which can be interpreted that we are looking at assets which are already discounted). The stock is a risky asset traded on a stock exchange. At time $t = 0$ it is worth 1 and at time $t = 1$ it either increases to 2 if the

the state "good" occurs and it decreases to $\frac{1}{2}$ if the state "bad" occurs. The stock induces a filtration $\mathcal{F}_0 = \sigma\{S_0\} = \{\emptyset, \Omega\}$ and $\mathcal{F}_1 = \sigma\{S_0, S_1\} = \mathcal{P}(\Omega)$.

There are more complicated assets that are traded and that depend on the price of the stock S at time 1 (which is unknown at time 0 as $\mathcal{F}_0$ is the trivial σ-algebra). These assets are called "derivative assets". As an example we introduce the "European call option".

Definition 4.1.1. *A European call option C with strike price K is a contract that gives you the right to buy the stock at time 1 for the price K (which is fixed at $t = 0$). If we assume that at time 1 the option is exercised we see that it is described by the following random variable (which is the value of C at time 1):*

$$C_1 = (S_1 - K)_+$$

In the above definition we use, for any random variable X, the notation X_+ for the positive part of the random variable, i.e., $X_+(\omega) = X(\omega)\mathbb{1}_{\{\omega: X(\omega) \geq 0\}}$. If you have the right (but not the obligation!) to buy S at time 1 for the price K then you would do the following: if $K > S_1(\omega)$ you would definitely not buy S for this price. If you would like to have the stock you obviously would prefer to buy it at its (cheaper!) market price $S_1(\omega)$. So, in this case, the option is of no value for you, that means it's worth 0. But if $K \leq S_1(\omega)$ you would buy S for the price K. And then, to see what the value of the call option C at time 1 is, you would immediately sell the stock for the price $S_1(\omega)$ and hence your profit would be $S_1(\omega) - K$ (because you invested K and you received $S_1(\omega)$). This gives the following value of C_1:

$$C_1(\omega) = \begin{cases} 0 & \text{if } S_1(\omega) < K \\ S_1(\omega) - K & \text{if } S_1(\omega) \geq K \end{cases},$$

and this is exactly the positive part of the random variable $S_1 - K$, i.e., $C_1 = (S_1 - K)_+$.

Now suppose on our toy market model we would be able to buy the contract C, the call option. We would have to buy it at time 0 for a certain price C_0. The problem is that, at time 0, we do not know what the value of S_1 will be (this we will only see at time 1). So the contract C amounts to a "bet" on the outcome of S at time 1. And the question is: what would we be willing to pay at time 0 for such a bet? Or let's put it rather as follows, for the contract C that, at time 1, satisfies $C_1 = (S_1 - K)_+$:

What would be a fair price $\mathbf{C_0}$ at time 0?

In order to be able to do some concrete calculations let us now fix the strike price $K = 1$ in the toy market. We can now easily determine the possible values of $C_1 = (S_1 - K)_+ = (S_1 - 1)_+$. Indeed: $S_1(g) = 2 > 1 = K$, so $S_1(g) - 1 = 2 - 1 = 1$. And $S_1(b) = \frac{1}{2} < 1 = K$ and $S_1(b) - 1 = -\frac{1}{2}$ and hence

$$C_1(\omega) = (S_1(\omega) - 1)_+ = \begin{cases} 1 & \text{if } \omega = g \\ 0 & \text{if } \omega = b \end{cases}$$

Now what could be an idea for the price at time 0?

4.1.1 First idea: equivalence principle as in insurance

There is the classical approach of insurance mathematics to calculate the premium for an insurance contract as the expected value of the discounted cash flows. In our case that would mean that we get as candidate $\tilde{C}_0$ for the price at time 0 the following:

$$\tilde{C}_0 = E[e^{-r} C_1] = E[C_1],$$

because the interest rate $r = 0$ and the time point $t = 1$ hence the discount factor $e^{-rt} = e^{-r} = e^0 = 1$. So we could just try $\tilde{C}_0 = E[C_1]$ as price at time 0 for the contract C.

Is this a reasonable price in the sense that a "rational" investor would buy the contract C for the price $\tilde{C}_0$?

By a rational investor we would mean that she would buy C for the price suggested above if there is no better choice (i.e. a rational investor would try to maximize expected gains under as small risk as possible). Ok, the expected profit of C would be $E[C_1 - \tilde{C}_0] = E[C_1] - \tilde{C}_0 = 0$ because we would have to pay $\tilde{C}_0$ at time 0 and make the expected profit $E[C_1]$ which in the end gives us $E[C_1] - \tilde{C}_0 = 0$. So by buying and exercising the contract our expected profit would be 0. But we also have a risk: if the situation "bad" occurs our loss would be $-\tilde{C}_0 = -E[C_1]$. To see some numbers let's calculate the expectation

$$E[C_1] = 1 \cdot P(g) + 0 \cdot P(b) = \frac{1}{2}.$$

So, if scenario "bad=b" occurs we would loose our investment, i.e., we would loose $\frac{1}{2}$. But scenario b has a positive probability of occurring, namely $P(b) = \frac{1}{2} > 0$. So we would buy a contract that has an expected profit of 0 but still, if we are unlucky, we could loose $\frac{1}{2}$. The question is: is there maybe a better choice to get an expected profit of 0 without the risk? The answer is: yes. Instead of buying the contract C for the price $\tilde{C}_0 = \frac{1}{2}$ we could invest the amount $\frac{1}{2}$ in the bond and buy $\frac{1}{2}$ of a bond (that means put $\frac{1}{2}$ into a bank account). The interest rate is 0 so at time 1 we would still have $\frac{1}{2}$ in the bank account. Our profit would be $\frac{1}{2}B_1 - \frac{1}{2}B_0 = \frac{1}{2} - \frac{1}{2} = 0$, hence again a profit of 0. But, of course, here is no risk, because $B_0 = B_1 = 1$ is deterministic. Hence a rational investor would not buy C for $\tilde{C}_0 = \frac{1}{2}$ but instead put the amount $\frac{1}{2}$ into the riskless (!) bank account. So, the price $\tilde{C}_0$ is **too expensive**. The fact that in buying the contract C there is **a risk** involved should be reflected in the price at time 0.

4.1.2 A better idea: "replicate" C_1 by a portfolio containing B and S. Pricing by "No Arbitrage"

Here we have to make some assumptions on our market model. We assume that the market is frictionless (this means we have no costs for transactions and we have no trading restrictions, that means we can buy or sell any fraction of the assets B and S). So we are able to trade any linear combination $\alpha B + \beta S$ with $\alpha, \beta \in \mathbb{R}$ on the market. As we have only time points $t = 0, 1$ we can only do the following: decide at time $t = 0$ about the amount α that we would like to put into the bond and the amount β that we would like to put into the stock. We cannot base our decision on the values of S_1 because we cannot say: ok, if g occurs then I would like to buy S at time 0: at time 0 we only have the information $\mathcal{F}_0$ which is the trivial σ-algebra. Therefore α and β are just real numbers (but cannot depend on g, b).

Now we try to **replicate** the value of the option at time 1 by a portfolio Π_t, $t = 0, 1$. Replicate means that at time $t = 1$ we have that $\Pi_1 = C_1$. And a portfolio is just a linear combination of B and S, i.e., $\Pi = \alpha B + \beta S$, where $\alpha, \beta \in \mathbb{R}$. If we look for a replicating portfolio we have to find real numbers α and β such that $\Pi_1(\omega) = \alpha B_1(\omega) + \beta S_1(\omega) = C_1(\omega)$, for $\omega = g, b$. This gives the following 2 linear equations

(i) $\alpha + 2\beta = 1$ (for $\omega = g$)

(ii) $\alpha + \frac{1}{2}\beta = 0$ (for $\omega = b$)

If we solve the two linear equations we get $\alpha = -\frac{1}{3}$ and $\beta = \frac{2}{3}$. Let's check again if the portfolio Π satisfies what we want:

$$\Pi_1(\omega) = -\frac{1}{3}B_1(\omega) + \frac{2}{3}S_1(\omega) = \begin{cases} -\frac{1}{3} + \frac{4}{3} = 1 & \text{for } \omega = g \\ -\frac{1}{3} + \frac{2}{3}\frac{1}{2} = 0 & \text{for } \omega = b, \end{cases}$$

so, indeed, $\Pi_1(\omega) = C_1(\omega)$ for all $\omega \in \Omega$, i.e., $\omega = g, b$.

So we found a new financial asset, namely the portfolio Π which has exactly the same value as C at time $t = 1$. As Π is just a linear combination of B and S, we can immediately find the price of the portfolio at time $t = 0$ namely

$$\Pi_0 = -\frac{1}{3}B_0 + \frac{2}{3}S_0 = \frac{1}{3}.$$

It is now intuitively clear that if we have two assets with the same price at time $t = 1$ that their prices should also be equal at time 0. Indeed, suppose this would not be the case and assume $C_0 < \Pi_0$. Then we could do the following: at time 0 we buy the asset C for the price C_0 and sell ("go short in") the portfolio Π for the price Π_0. Hence we would make the gain $\Pi_0 - C_0 =: \epsilon > 0$ because we have to pay C_0 but we receive Π_0. Now, at time $t = 1$ we own C_1 which is $= \Pi_1$ so we can fulfill our obligations without further costs. Hence we made a strictly positive gain (at time 0) that was completely riskless. Now suppose we are doing this 100 times or 1000000 times or n times with n very large then our riskless profit would be n and could be made arbitrarily large. A riskless gain, that is nonnegative and strictly positive with positive probability is called an **arbitrage**. And in mathematical finance we make the assumption of **No Arbitrage**, this means, that there should be no arbitrage opportunities in the market. This assumption is justified by the fact that if there were an arbitrage opportunity (i.e. as said before an opportunity of making a strictly positive profit without any risk) then there are many organized traders that see it, make use of it and then the opportunity is immediately gone. So it is a reasonable assumption that for the average trader there just does not exist arbitrage in the market.

Hence by No Arbitrage and by the fact that $\mathbf{\Pi_1 = C_1 \Rightarrow \Pi_0 = C_0}$.

And hence we found a fair price at time 0 for the European call option C, namely $C_0 = \Pi_0 = \frac{1}{3}$. Observe that this price is strictly smaller than our first tentative price $\tilde{C}_0 = \frac{1}{2}$ that we found before by the equivalence principle. Here the fact that there is risk is reflected in the price $C_0 = \frac{1}{3}$ (in contrast to the too expensive price $\tilde{C}_0 = \frac{1}{2}$ from before).

4.1.3 Another idea: change the probability measure; equivalent martingale measures

The probability measure P reflects a certain view of the world. It gives the weight $\frac{1}{2}$ to both possible states of the world. Suppose we would change our view of the world and shift the weights between g and b in such a way that we still have a probability measure that gives strictly positive weight to both states. This means in our toy model that we agree with the measure P on the sets that occur with probability 0 or 1 (loosely speaking we agree on the "impossible" and the "certain" events but we give different weights to the events that are not certain/impossible). This means we choose an **equivalent** probability measure. Let's give a general definition:

Definition 4.1.2 (Absolutely continuous and equivalent probability measures)**.** *Suppose P and Q are probability measures on a measurable space $(\Omega, \mathcal{F})$. We say that*

(i) $Q \ll P$ *(Q is* **absolutely continuous** *with respect to P) if for every $A \in \mathcal{F}$ with $P(A) = 0$ we have that $Q(A) = 0$,*

(ii) $Q \sim P$ *(Q and P are equivalent) if $P \ll Q$ and $Q \ll P$.*

Note that $Q \sim P$ means, for $A \in \mathcal{F}$, $Q(A) = 0 \iff P(A) = 0$. In the given toy model with $\Omega = \{g, b\}$ this means the following. As $P(g) = P(b) = \frac{1}{2}$ the only P-nullset is $\emptyset$. Hence an equivalent probability measure $Q \sim P$ can have as nullset only the $\emptyset$, too. Hence it has to hold that $Q(g) > 0$ and $Q(b) > 0$ if $Q \sim P$. As Q should be a probability measure the total mass has to be 1, i.e., $Q(g) + Q(b) = 1$.

Suppose we could find an equivalent probability measure Q such that the risky asset S behaves in the mean like a riskfree asset, which means that, in the mean, we do not lose or gain anything. (This should remind you of the definition of a martingale, a fair game, see Definition3.2.1.) Mathematically this means that there exists a probability measure $Q \sim P$ such that with respect to the measure Q the stochastic process $(S_t)_{t=0,1}$ is a martingale, i.e., a fair game, a game that is riskfree in the mean. In the very easy setting of two time points the martingale property reduces to

$$E_Q[S_1] = S_0$$

because $\mathcal{F}_0$ is trivial and hence $E_Q[S_1|\mathcal{F}_0] = E_Q[S_1] = E_Q[S_0] = S_0$. By the notation $E_Q[.]$ we mean the expected value with respect to the probability measure Q.

Let's see if we can find such a Q in our case. Let's fix $Q(g) = x$ and $Q(b) = y$ and determine a system of equations which will let us calculate the unknowns x and y. We would like to have $Q \sim P$ therefore we need $0 < x < 1$ and $0 < y < 1$. Q should be a probability measure which leads to

Equation 1: $x + y = 1$

The martingale property $E_Q[S_1] = S_1(g)Q(g) + S_1(b)Q(b) = S_0$ leads to

$$\textbf{Equation 2: } 2x + \frac{1}{2}y = 1.$$

The 2 equations give a unique solution, namely $x = Q(g) = \frac{1}{3}$ and $y = Q(b) = \frac{2}{3}$.

Now let's come back to the European call option C with strike price $K = 1$. As the stochastic process S now, in the mean, behaves like a riskless asset it might make sense to go back to the idea of the equivalence principle **but with the measure Q** instead of P. So, let's calculate the mean of C_1:

$$E_Q[C_1] = 1Q(g) + 0Q(b) = 1 \cdot x = 1 \cdot \frac{1}{3} = \frac{1}{3}.$$

Observe that we found the price $C_0 = \frac{1}{3}$ from Subsection 4.1.2 again! This is not a coincidence - there is one of the most important theorems of mathematical finance behind this. And it is useful to have the possibility of calculating the same price as in Subsection 4.1.2 because in more complicated examples it can be easier to find the equivalent martingale measures than to find the concrete replicating portfolio.

4.1.4 Fundamental Theorem of Asset Pricing: the connection between Pricing by NA and equivalent martingale measures

Let's formulate the Fundamental Theorem of Asset Pricing here in the generality that we will study later on for a finite discrete time set $\mathbb{T} = \{0, 1, 2, \dots, T\}$.

Theorem 4.1.3 (Fundamental Theorem of Asset Pricing)**.** *A market model satisfies NA (no arbitrage) if and only if there exists a probability measure $Q \sim P$ such that the stochastic process $(S_t)_{t=0,\dots,T}$ is a martingale.*

Definition 4.1.4. *A probability measure Q that satisfies the properties of Theorem 4.1.3 is called an* **equivalent martingale measure***. Let us call the set of all such equivalent martingale measures $\mathcal{M}^e(S)$, i.e.,*

$$\mathcal{M}^e(S) = \{Q \sim P : \textit{such that } (S_t)_{t=0}^T \textit{ is a martingale with respect to } Q\}$$

Theorem 4.1.3 holds in full generality even in continuous time (however in continuous time there are some subtle issues: one has to modify the NA condition to a slightly stronger condition called no free lunch with vanishing risk and the measures are not equivalent martingale measures but something that is called equivalent σ-martingale measures, see Delbaen and Schachermayer [3], [4]). In discrete finite time the above result holds in the given form, see Harrison and Pliska [7] in the case of a finite Ω, and Dalang,Morton and Willinger [2] in the case of a general sample space Ω. Later on in these lecture notes we will see the proof for finite Ω, i.e., the Harrison-Pliska result.

Remark 4.1.5 (Complete and incomplete markets). Note that in the above theorem we say there exists "an" equivalent martingale measure Q. We do not say there exists "the" equivalent martingale measure. This is a hint that there exist markets where martingale measures are not unique. It is in fact the case that the set of all equivalent martingale measures $\mathcal{M}^e(S)$ consists

(i) **either** of exactly one point: $\mathcal{M}^e(S) = \{Q\}$

(ii) **or** it is a set containing infinitely many probability measures. $\mathcal{M}^e(S)$ is a convex set.

If the case (i) holds then the market model is called **complete** and if (ii) holds the market model is called **incomplete**.

Of course in our toy model with two states of the world the following situation occurs: **if** there exists an equivalent martingale measure, then it is unique. So, a state space with two states of the world (if it is free of arbitrage) always leads to a complete market. And we already saw in Subsection 4.1.3 that the system of two equations that gave us the values $x = Q(g)$ and $y = Q(b)$ had a **unique** solution. This is not surprising as we found a system of 2 linear equations in 2 unknowns. There still is the possibility that there does not exist any martingale measure in a market with two possible states of the world. By Theorem 4.1.3 we already know that then there has to be arbitrage. Let's look at this situation in an example.

Example 4.1.6. We consider exactly the same probability space as in the whole Section, namely $\Omega = \{g, b\}$. Times $t = 0$ and $t = 1$. Same probability measure P on $(\Omega, \mathcal{P}(\Omega))$ as before, i.e., $P(g) = P(b) = \frac{1}{2}$. We still have the riskfree bond B with $B_0 = B_1 = 1$. We will modify the toy model now by taking a different risky asset S. Let again $S_0 = 1$ but we change S_1 to

$$S_1(\omega) = \begin{cases} 2 & \text{if } \omega = g \\ 1 & \text{if } \omega = b \end{cases}$$

So, in this example, the stock either strictly increases to 2 (if the "good" scenario occurs) or it stays at the same level namely 1 (if the "bad" scenario occurs, which is actually not too bad here). Let's try to find an equivalent martingale measure as before. Set again $Q(g) = x$ and $Q(b) = y$. We get again the equation $x + y = 1$ but our second equation (coming from $E_Q[S_1] = S_0$) now looks as follows:

$$2x + y = 1.$$

Trying to solve this gives the unique solution $x = 0$ and $y = 1$. But this has the effect the the probability measure Q does not satisfy $Q \sim P$. Indeed, for the set $A = \{g\}$ we have that $Q(A) = 0$ but $P(A) = \frac{1}{2} > 0$. But still, at least Q would be a martingale measure, however, only absolutely continuous ($Q \ll P$). (Check, why this holds!) We know that there must exist an arbitrage opportunity and here it is: buy S for the price $S_0 = 1$ at time 0. Then sell S at time 1. In the case of "b" you have $S_1(b) - S_0 = 1 - 1 = 0$ so you did not loose or gain anything. But in the case "g" you have $S_1(g) - S_0 = 2 - 1 = 1$, i.e., here you make a strictly positive profit and this happens with the strictly positive probability $P(g) = \frac{1}{2} > 0$. You had no risk (no possible loss, the worst that happens is 0). So you made arbitrage.

We can of course modify the example in a way that you would make a profit in both cases, for example:

$$S_1(\omega) = \begin{cases} 2 & \text{if } \omega = g \\ \frac{3}{2} & \text{if } \omega = b \end{cases}$$

The 2 linear equations for the martingale measure are then $x + y = 1$ and $2x + \frac{3}{2}y = 1$. The unique solution is $x = -1$ and $y = 2$. This cannot give a probability measure because for a probability measure, obviously, we need $0 \leq x, y \leq 1$. So there does not exists any martingale measure. The arbitrage opportunity works as above, but the profit on "b" is $\frac{3}{2} - 1 = \frac{1}{2}$ so that here we even make strictly positive gain for both states g and b, i.e. with probability $P(\{b, g\}) = P(\Omega) = 1$.

Remark 4.1.7. Here we used for the notion of arbitrage just the intuitive idea that you make a profit without risk that is strictly positive somewhere. The exact mathematical definition of an arbitrage opportunity will be given in Definition 4.3.5 below.

4.2 A model with three states of the world: first easy model of an incomplete market

Now we enrich the toy model from before by a third state of the world. We consider a sample space consisting of three possible scenarios $\Omega = \{g, m, b\}$.

Let's call the new state "m=" medium. The time set again is $\mathbb{T} = \{0, 1\}$. As σ-algebra again take the power set $\mathcal{P}(\Omega)$, i.e. the family of all subsets of $\Omega = \{g, m, b\}$. Moreover we have a probability measure on $(\Omega, \mathcal{P}(\Omega))$ that is now given by $P(g) = P(m) = P(b) = \frac{1}{3}$. Again we have two assets. The bond B as before, i.e., $B_0 = B_1 = 1$. And for the stock S we take a new value at time $t = 0$, to have some change, i.e., $S_0 = 2$ (this is not important). But the important difference is that S_1 takes 3 possible values:

$$S_1(\omega) = \begin{cases} 3 & \text{if } \omega = g \\ 2 & \text{if } \omega = m \\ 1 & \text{if } \omega = b \end{cases}$$

Now as before let us find equivalent martingale measures Q (if they exist). Let $Q(g) = x$, $Q(m) = y$ and $Q(b) = z$. As Q should be a probability measure we get the first equation:

$$(1) \quad x + y + z = 1$$

By the martingale condition we need that $E_Q[S_1] = S_0 = 2$, i.e., $3Q(g) + 2Q(m) + Q(b) = 2$, so we get the second equation

$$(2) \quad 3x + 2y + z = 2$$

This is a system of 2 linear equations and we have 3 unknowns. From linear algebra you know that we then have 1 degree of freedom, i.e., 1 parameter. Let's choose $z = Q(b) = \alpha$. We will soon see that we have some constraints on α. Let's now solve the system of 2 linear equations. This gives $x = \alpha$, $y = 1 - 2\alpha$. Now, Q has to be a probability measure and $Q \sim P$, so we need that $0 < x, y, z < 1$. (Check why!) Therefore we get $0 < \alpha$. And by the fact that $y = 1 - 2\alpha > 0$ we get $\alpha < \frac{1}{2}$. Let's call Q^α the probability measure that satisfies $Q^\alpha(g) = \alpha = Q^\alpha(b)$ and $Q^\alpha(m) = 1 - 2\alpha$ for a fixed $\alpha \in (0, \frac{1}{2})$. Then we see that we found a convex set of equivalent martingale measures, i.e.,

$$\mathcal{M}^e(S) = \{Q^\alpha, 0 < \alpha < \frac{1}{2}\}.$$

Hence we found our first example of an **incomplete** financial market, see Remark 4.1.5. Now again, we can use the set of equivalent martingale measures for finding fair prices for a European call option. Indeed, let, C be a European call option with strike price $K = \frac{3}{2}$ then

$$C_1(\omega) = \left(S_1 - \frac{3}{2}\right)_+ (\omega) = \begin{cases} \frac{3}{2} & \text{if } \omega = g \\ \frac{1}{2} & \text{if } \omega = m \\ 0 & \text{if } \omega = b \end{cases}$$

.

Remark 4.2.1. All possible fair (= arbitrage-free) prices C_0^α form an **open** interval, namely

$$\left(\inf_{Q\in\mathcal{M}^e(S)} E_Q[C_1], \sup_{Q\in\mathcal{M}^e(S)} E_Q[C_1]\right)$$

Every real number c with $c \notin \left(\inf_{Q\in\mathcal{M}^e(S)} E_Q[C_1], \sup_{Q\in\mathcal{M}^e(S)} E_Q[C_1]\right)$ gives an arbitrage opportunity (in particular the boundary points of the interval).

Let's calculate the price interval. We know that for each $\alpha \in (0, \frac{1}{2})$ there is an **emm** (=equivalent martingale measure) Q^α. Fix α and calculate

$$\begin{aligned} E_{Q^\alpha}[C_1] &= \frac{3}{2}Q^\alpha(g) + \frac{1}{2}Q^\alpha(b) + 0Q^\alpha(b) \\ &= \frac{3}{2}\alpha + \frac{1}{2}(1-2\alpha) = \frac{1}{2}(1+\alpha). \end{aligned}$$

Now we easily see that

$$\inf_{Q\in\mathcal{M}^e(S)} E_Q[C_1] = \inf_{\alpha\in(0,\frac{1}{2})} E_{Q^\alpha}[C_1] = \inf_{\alpha\in(0,\frac{1}{2})} \frac{1}{2}(1+\alpha) = \frac{1}{2},$$

and

$$\sup_{Q\in\mathcal{M}^e(S)} E_Q[C_1] = \sup_{\alpha\in(0,\frac{1}{2})} \frac{1}{2}(1+\alpha) = \frac{1}{2}(1+\frac{1}{2}) = \frac{3}{4},$$

and so we found the open interval of fair prices C_0. This means every real number C_0 with

$$C_0 \in \left(\frac{1}{2}, \frac{3}{4}\right) \quad \text{is a fair price for } C \text{ at time } 0.$$

Of course, at this stage of the lecture notes, we did not really define "fair price". We just said that it is a price that does not lead to an arbitrage. To prove that all prices $C_0 \in (\inf_{Q\in\mathcal{M}^e(S)} E_Q[C_1], \sup_{Q\in\mathcal{M}^e(S)} E_Q[C_1])$ (for a European call option with some strike price K) are fair prices is the content of exercise 3.1c. Try really to show that you cannot find any arbitrage. Now if we would take a price outside the open interval we get an arbitrage. Let's check this for one of the boundaries of the open interval $(\frac{1}{2}, \frac{3}{4})$ in the case of our concrete example.

Claim: $C_0 = \frac{1}{2}$ as time 0 price of $C_1 = (S_1 - \frac{3}{2})_+$ leads to an arbitrage opportunity.

Proof of the claim: Let us now be precise what an **arbitrage** is in our example: it is a portfolio Π consisting of the given assets in the market which are B, S and C (but C with $C_0 = \frac{1}{2}$), i.e.,

$$\Pi_t = aC_t + bB_t + cS_t, \quad t = 0, 1,$$

with the following properties.

- $\Pi_0 \leq 0$ (which means we invest a maximum amount of 0 at time 0 (negative costs would mean that we get some extra money for free)) and
- $\Pi_1(\omega) \geq 0$ for all ω and there exists an ω such that $\Pi_1(\omega) > 0$. This means that at time 1 we don't make a loss (because $\Pi_1 \geq 0$ P-a.s.) and with strictly positive probability we make a strictly positive profit, because for one $\omega \in \{g, m, b\}$ we have that $\Pi_1(\omega) > 0$ and $P(\omega) = \frac{1}{3} > 0$.

We will now find such a portfolio. Let's look for a, b, c such that $\Pi_0 = 0$. This gives us the equation $aC_0 + bB_0 + cS_0 = 0$, i.e.,

$$(1) \quad \frac{1}{2}a + b + 2c = 0.$$

What do we have at time 1?

$$\Pi_1(\omega) = aC_1(\omega) + bB_1 + cS_1(\omega) = \begin{cases} \frac{3}{2}a + b + 3c =: x & \text{for } \omega = g \\ \frac{1}{2}a + b + 2c =: y & \text{for } \omega = m \\ b + c =: z & \text{for } \omega = b \end{cases}$$

What we now need for an arbitrage is $x, y, z \geq 0$ and one of them > 0. Now it's a bit of guessing a clever choice. By equation (1) we see that $\Pi(m) = y = \frac{1}{2}a + b + 2c = 0$. Then we can express $\Pi_1(g) = \Pi_1(m) + a + c = a + c$, because $\Pi_1(m) = 0$. Let's see what happens if we try $a + c = 0$, then $c = -a$ and $\Pi_1(g) = \Pi_1(m) = 0$. Now we can solve the equation (1) with $c = -a$ to get $\frac{1}{2}a + b - 2a = 0$ and hence $b = \frac{3}{2}a$. We still have a to choose and we do that such that $z = b + c = \frac{3}{2}a - a = \frac{1}{2}a > 0$. Hence we can choose any $a > 0$ for example $a = 1$. Then the portfolio $\Pi = C + \frac{3}{2}B - S$ satisfies $\Pi_0 = 0$ (for $C_0 = \frac{1}{2}$) and $\Pi_1(g) = \frac{3}{2} + \frac{3}{2} - 3 = 0$, $\Pi_1(m) = \Pi_0 = 0$ and $\Pi(b) = \frac{3}{2} - 1 = \frac{1}{2} > 0$. Hence we found an arbitrage. And we proved the claim. $\square$

Remark 4.2.2. (i) We saw that it was crucial to choose $a > 0$. Let us see why this is the case. By taking $C_0 = \frac{1}{2}$ we underprice the option, it is too cheap, because $\frac{1}{2}$ is the left border of the **open** interval of the fair prices. Hence we should buy the option at time 0 to make a gain from this underpricing. To buy the option at time 0 means that we own it at time 1 therefore this transaction amounts to a positive weight $a > 0$ for the option. In the portfolio that we chose with $a = 1$ it means we sell one stock S for the price of 2 and buy one C (for $\frac{1}{2}$) and $\frac{3}{2}$ of Bond (from the money $2 = \frac{1}{2} + \frac{3}{2}$ we got from selling S). So at time 0 we invest 0. At time 1 we own Π_1 which is ≥ 0 and $P(\Pi_1 = \frac{1}{2}) = \frac{1}{3} > 0$. In Π_1 the option appears with the factor $a = 1 > 0$ because we bought it at time 0 and exploited the underpricing with $C_0 = \frac{1}{2}$.

(ii) By rewriting the portfolio $\Pi_1 = C_1 + \frac{3}{2}B_1 - S_1 \geq 0$ a.s. we see that if we define $\widetilde{\Pi} = -\frac{3}{2}B + S$ then

$$\widetilde{\Pi}_1 \leq C_1 \quad \text{and } P(\widetilde{\Pi}_1 < C_1) = \frac{1}{3},$$

hence we found a portfolio $\widetilde{\Pi}$ consisting only of bond and stock with $\widetilde{\Pi}_0 = C_0$ and $\widetilde{\Pi}_1 \leq C_1$ a.s. but $\widetilde{\Pi}_1 \neq C_1$ (because there is the strict inequality for the state "b"). So we did **not** find a portfolio that **replicates** the option. This is not a coincidence but this is the case because the market is incomplete. In contrast to that in a complete market (as the two states example) there **always** is a replicating portfolio. In an incomplete market there is a replication portfolio only for those derivative assets X with $\mathbb{E}_Q[X_1] = const$ over the set of emms, i.e., there is X_0 such that $E_Q[X_1] = X_0$ **for all** $Q \in \mathcal{M}^e(S)$. Note that the option C above does not satisfy this, because $\inf_{Q\in\mathcal{M}^e(S)} E_Q[C_1] = \frac{1}{2} < \sup_{Q\in\mathcal{M}^e(S)} E_Q[C_1] = \frac{3}{4}$. Hence there is no perfect hedge for this C, there does not exists a replicating portfolio $\widehat{\Pi}_1 = C_1$ a.s. and $\widehat{\Pi}_0 = C_0$ (for no value of C_0).

Example 4.2.3. Let's see if we can find a strike price K such that the option has a perfect hedge (i.e. a portfolio $(\Pi_t)_{t=0,1}$ such that $\Pi_t = C_t$ for $t = 0, 1$. Choose $K = 1$. Then the European call option X with strike price $K = 1$ satisfies

$$X_1(\omega) = (S_1 - 1)_+(\omega) = \begin{cases} 2 & \omega = g \\ 1 & \omega = m \\ 0 & \omega = b \end{cases}.$$

Take any measure $Q^\alpha \in \mathcal{M}^e(S)$ for $0 < \alpha < \frac{1}{2}$ fixed. Then

$$E_{Q^\alpha}[X_1] = 2Q^\alpha(g) + Q^\alpha(m) = 2\alpha + (1 - 2\alpha) = 1.$$

As the α cancels out we have that $E_Q^\alpha[X_1] = 1$ **for all** $0 < \alpha < \frac{1}{2}$ hence for all $Q \in \mathcal{M}^e(S)$. For the asset X_1 with $X_0 = 1$ we can find a replicating portfolio. Indeed take $\widetilde{\Pi} = S - B$ then $\widetilde{\Pi}_0 = S_0 - B_0 = 2 - 1 = 1 = X_0$ and

$$\begin{aligned}\widetilde{\Pi}_1 = S_1 - B_1 &= \begin{cases} 2 & \omega = g \\ 1 & \omega = m \\ 0 & \omega = b \end{cases} \\ &= X_1\end{aligned}$$

This is in line with Remark 4.2.2, (ii).

4.3 Fundamental Theorem of Asset Pricing for a finite Ω

4.3.1 The model

Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ be a finite set. As σ-algebra $\mathcal{F}$ we again take the power set, i.e., $\mathcal{F} = \mathcal{P}(\Omega)$. We fix a probability measure P on $(\Omega, \mathcal{F})$ such that $P(\omega_n) = p_n > 0$ for all $n = 1, \dots, N$. As P is a probability measure it obviously holds that $\sum_{n=1}^{N} p_n = 1$. The time set is the finite set $\mathbb{T} = \{0, 1, \dots, T\}$. We will have a filtration on the model describing that information increases with time, i.e., there is an increasing family of σ-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_T \subseteq \mathcal{F} = \mathcal{P}(\Omega)$.

Definition 4.3.1 (Market model)**.** *A model of a financial market is an $\mathbb{R}^d$-valued stochastic process $(S_t)_{t\in\mathbb{T}}$ that is adapted with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$. To be precise, for each $t \in \mathbb{T}$, $S_t = (S_t^1, S_t^2, \dots, S_t^d)$ is a random vector where $S_t : (\Omega, \mathcal{F}_t) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is measurable (because of the adaptedness we get that S_t is $\mathcal{F}_t$-measurable). Moreover there exists a riskless asset $S_t^0 = B_t = 1$ (we again assume that everything is already discounted, i.e., the interest rate $r = 0$. For the subtlety that shows how the case of $r > 0$ can be reduced to the discounted case later we refer to [5]. Because we will need this soon we define the $d + 1$-dimensional stochastic process $\hat{S}_t = (S_t^0, S_t^1, \dots, S_t^d)$ that contains the riskless asset S^0 and the vector of the d risky assets $(S^1, S^2, \dots, S^d)$.*

As in the easy examples before the random vector $S_t = (S_t^1, S_t^2, \dots, S_t^d)$ describes the price of d risky stocks at time t. The riskless asset S^0 is the bond. Note that in the easy examples from the last section we had $N = 2$ or 3, $\mathbb{T} = \{0, 1\}$, i.e., $T = 1$, and $d = 1$.

Now we will find a mathematical way of describing what trading means and what a portfolio is in this more general situation.

Definition 4.3.2 (Trading strategy)**.** *A trading strategy is an $\mathbb{R}^{d+1}$-valued* **predictable** *stochastic process $(\hat{H}_t)_{t=1,\dots,T}$. This means, for each $t = 1, 2, \dots, T$, the random vector $\hat{H}_t = (H_t^0, H_t^1, \dots, H_t^d)$ is $\mathcal{F}_{t-1}$-measurable, i.e.,*

$$\hat{H}_t : (\Omega, \mathcal{F}_{t-1}) \to (\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1})).$$

The random variable H_t^0 describes the amount put into the riskless asset S^0 at time t and H_t^i the amount put in the risky asset S^i, $i = 1, \dots, d$. We have to decide on this before the price of the risky assets is shown at time t hence $\hat{H}_t$ can only use the information of time $t-1$ which means $\hat{H}_t$ has to be $\mathcal{F}_{t-1}$-measurable.

Definition 4.3.3 (Value process)**.** *The value process $(V_t)_{t=0,\dots,T}$ of a portfolio using the trading strategy $(\hat{H}_t)_{t=1,\dots,T}$ is given as follows: V_0 is an initial investment and for $t = 1, \dots, T$ we have that*

$$V_t = (\hat{H}_t, \hat{S}_t) = \sum_{i=0}^{d} H_t^i S_t^i,$$

where $(.,.)$ denotes the inner product on $\mathbb{R}^{d+1}$.

Note that $(V_t)_{t\in\mathbb{T}}$ is an $\mathbb{R}$-valued adapted stochastic process. (Convince yourselves, why this is true!) The interpretation is that the total value of the portfolio (with trading strategy $\hat{H}$) at time t is the sum of the amount you put in each asset times the price at time t. From now on we are only interested in a special type of trading strategies that allow the following: at time $t-1$ we decide which amount of which asset we would like to hold at time t. At time t we decide which amount we would like to hold at time $t+1$ (so we decide on $\hat{H}_{t+1}$). However, the change we do in the step from $\hat{H}_t$ to $\hat{H}_{t+1}$ does not allow to put additional money into the portfolio and it does also not allow to take money out of the portfolio. The only thing we can do is shift our total capital in the portfolio around among the different assets. A trading strategy like that is called **self financing**. Here is the definition:

Definition 4.3.4. *A trading strategy $(\hat{H}_t)_{t=1,\dots,T}$ is called self financing if the following holds:*

$$V_t = (\hat{H}_t, \hat{S}_t) = (\hat{H}_{t+1}, \hat{S}_t).$$

The interpretation is, that at time t we decide how to distribute the money between the different assets for the next time $t+1$, this is the strategy $\hat{H}_{t+1}$. The value cannot change hence $V_t = (\hat{H}_t, \hat{S}_t) = (\hat{H}_{t+1}, \hat{S}_t)$. Then time evolves to $t+1$ and our value moves from $(\hat{H}_{t+1}, \hat{S}_t) = V_t$ to the new value

$(\hat{H}_{t+1}, \hat{S}_{t+1}) = V_{t+1}$ (only by the movement of the price from $\hat{S}_t$ to $\hat{S}_{t+1}$). The self financing property will help us to express the value process in a convenient form. Indeed, let's look at the increment $V_{t+1} - V_t$ for a fixed t:

$$\begin{aligned}
V_{t+1} - V_t &= (\hat{H}_{t+1}, \hat{S}_{t+1}) - (\hat{H}_t, \hat{S}_t) \\
&= (\hat{H}_{t+1}, \hat{S}_{t+1}) - (\hat{H}_{t+1}, \hat{S}_t) \quad \text{because of the self fin. property} \\
&= \sum_{i=0}^{d} H^i_{t+1} S^i_{t+1} - \sum_{i=0}^{d} H^i_{t+1} S^i_t \\
&= \sum_{i=1}^{d} H^i_{t+1}(S^i_{t+1} - S^i_t) + H^0_{t+1} S^0_{t+1} - H^0_{t+1} S^0_t \\
&= \sum_{i=1}^{d} H^i_{t+1}(S^i_{t+1} - S^i_t) + H^0_{t+1}(B_{t+1} - B_t) \\
&= \sum_{i=1}^{d} H^i_{t+1}(S^i_{t+1} - S^i_t),
\end{aligned}$$

where the last two equalities follow because $S^0 = B$ and $B_{t+1} = B_t = 1$. We see above that the investment in the riskless asset $S^0 = B$ cancels out. Define $H_t := (H^1_t, \ldots, H^d_t)$ to be the trading strategy which only describes what to put in the d risky stocks then we see that we proved above that

$$V_{t+1} - V_t = (H_{t+1}, S_{t+1} - S_t),$$

because the above sum can of course be written with the help of the inner product $(.,.)$ (this time on $\mathbb{R}^d$).

Let V_0 be any starting capital and define now $V_0 = (\hat{H}_0, \hat{S}_0)$. We did not have $\hat{H}_0$ before. But we see that from the self financing condition it follows that

$$(\hat{H}_0, \hat{S}_0) = (\hat{H}_1, \hat{S}_0) = H^0_1 S^0_0 + \sum_{i=1}^{d} H^i_1 S^i_0 = H^0_1 + \sum_{i=1}^{d} H^i_1 S^i_0,$$

because $S^0_0 = B_0 = 1$.

We can now rewrite the value process as follows. Let $u \in \{1, 2, \ldots, T\}$ then

$$
\begin{aligned}
V_u &= V_0 + \sum_{t=1}^{u}(V_t - V_{t-1}) \\
&= V_0 + \sum_{t=1}^{u}\sum_{i=1}^{d} H_t^i(S_t^i - S_{t-1}^i) \\
&= V_0 + \sum_{i=1}^{d}(\sum_{t=1}^{u} H_t^i(S_t^i - S_{t-1}^i)) \\
&= V_0 + \sum_{i=1}^{d}(H^i \cdot S^i)_u,
\end{aligned}
$$

where $(H^i \cdot S^i)_u$ is the stochastic integral in discrete time at time u of the predictable process (H^i) with respect to the adapted process (S^i), see Definition 3.3.3. Hence we proved that we can describe the value process with respect to a self-financing trading strategy as a sum of discrete time stochastic integrals.

4.3.2 No arbitrage, equivalent martingale measures and the Fundamental Theorem of Asset Pricing

Now we give the exact definition of an arbitrage.

Definition 4.3.5. *A self financing trading strategy $(H_t)_{t=1,\ldots,T}$, where $H_t = (H_t^1, H_t^2, \ldots, H_t^d)$, is an* **arbitrage** *opportunity if the value process (of the trading strategy H) satisfies:*

(i) $V_0 \leq 0$

(ii) $P(V_T \geq 0) = 1$ *and* $P(V_T > 0) > 0$.

We say that the model satisfies the condition **NA** *(= No Arbitrage) if there does not exist any arbitrage opportunity.*

Note that in our setting with finitely many $\omega_1, \ldots, \omega_N$ and $P(\omega_n) > 0$ for all $n = 1, \ldots N$, the condition (ii) means that $V_T(\omega_n) \geq 0$ for all $n = 1, \ldots, N$ and there exists $n_0 \in \{1, 2, \ldots, N\}$ such that $V_T(\omega_{n_0}) > 0$.

It might be difficult to exclude that there exist arbitrage strategies in a concrete given model. The following lemma helps: we don't have to check all possible arbitrages but it is enough to check those arbitrage opportunities that are done in only one time step. We do not need the content of

Lemma 4.3.6 in the following but the result is interesting in itself. Note that this lemma does not only hold for a finite Ω but also for a general Ω (in the proof we do not use that Ω is finite).

Lemma 4.3.6. *There exists an arbitrage opportunity if and only if there exists a 1-step-arbitrage, i.e., there exists $t \in \{1, 2, \dots T\}$ and an $\mathcal{F}_{t-1}$-measurable random vector η (with values in $\mathbb{R}^d$) such that*

- $(\eta, S_t - S_{t-1}) \geq 0$ *a.s.*
- $P((\eta, S_t - S_{t-1}) > 0) > 0$.

Proof. For ($\Rightarrow$) observe that if H is an arbitrage opportunity we can define

$$t = \min\{k : \text{ such that } V_k \geq 0 \text{ a.s. and } P(V_k > 0) > 0\}$$

This t satisfies $t \leq T$ because we know that H is an arbitrage opportunity. Then

(a) either $V_{t-1} = 0$ a.s.

(b) or $P(V_{t-1} < 0) > 0$.

This follows by the definition of t being the **smallest** time point such that..., which has the above consequences for $t-1$ which is of course $< t$. Now let's look at the two cases.

Suppose (a) holds then

$$(H_t, S_t - S_{t-1}) = V_t - V_{t-1} = V_t,$$

because $V_{t-1} = 0$ a.s. Then the 1-step-arbitrage holds for $\eta = H_t$.

Suppose that (b) holds. Define $\eta = H_t \mathbb{1}_{\{V_{t-1}<0\}}$. Note that because of the predictability of H and because V is adapted we have that η is $\mathcal{F}_{t-1}$-measurable. Then we have that

$$\begin{aligned}(\eta, S_t - S_{t-1}) &= \mathbb{1}_{\{V_{t-1}<0\}}(H_t, S_t - S_{t-1}) \\ &= \mathbb{1}_{\{V_{t-1}<0\}}(V_t - V_{t-1}) \\ &\geq -V_{t-1}\mathbb{1}_{\{V_{t-1}<0\}} \geq 0,\end{aligned}$$

where the first inequality in the last line holds because, by definition of t, we have that $V_t \geq 0$. Because (b) holds it follows that $P(V_{t-1} < 0) > 0$ and hence

$$P((\eta, S_t - S_{t-1}) > 0) \geq P(-V_{t-1}\mathbb{1}_{\{V_{t-1}<0\}} > 0) = P(V_{t-1} < 0) > 0.$$

So in both cases (a) and (b) we get a 1-step arbitrage.

The other direction ($\Leftarrow$) is trivial as a 1-step arbitrage obviously is an arbitrage. (Write down the strategy H induced by η!) $\square$

When does the model satisfy the condition NA? We already know from the previous sections that this has something to do with equivalent martingale measures. In the present setting we have to give the precise definitions. For the notion of equivalent probability measures check again Definition 4.1.2.

Definition 4.3.7 (Martingale measures)**.** *A probability measure Q on a measure space $(\Omega, \mathcal{F})$ is called a martingale measure for S, if*

(i) $E_Q[|S_t^i|] < \infty$ *for all* $i = 1, \ldots, d$ *and* $t = 0, \ldots, T$ *and*

(ii) $E_Q[S_{t+1}^i | \mathcal{F}_t] = S_t^i$ *a.s. for all* $t = 0, \ldots, T-1$, $i = 1, \ldots, d$.

This means that all risky assets are martingales with respect to the probability measure Q.

Note that condition (i) **always** holds for a finite Ω because

$$E_Q[|S_t^i|] = \sum_{n=1}^{N} |S_t^i(\omega_n)| Q(\omega_n) < \infty,$$

because it is a finite sum.

Let us now recall the main result of this chapter, see Theorem 4.1.3. In this chapter we will prove the following version for finite Ω.

Theorem 4.3.8 (Fundamental Theorem of Asset Pricing for finite $\Omega = \{\omega_1, \ldots, \omega_N\}$)**.** *NA holds $\iff$ there exists $Q \sim P$ such that Q is a martingale measure.*

4.3.3 Interlude on absolutely continuous measures

In the following let $(\Omega, \mathcal{F})$ be a measure space with a probability measure P on it.

Lemma 4.3.9. *Let $Z : (\Omega, \mathcal{F}) \to ([0, \infty), \mathcal{B}([0, \infty))$ be a measurable function that is integrable with respect to P, i.e. $E_P[|Z|] < \infty$. Define, for each $A \in \mathcal{F}$,*

$$\mu(A) = \int_A Z dP.$$

Then μ is a finite measure on $(\Omega, \mathcal{F})$ with $\mu \ll P$. The following additional properties hold:

(i) *μ is a probability measure if and only if $E_P[Z] = 1$.*

(ii) *$\mu \sim P$ if and only if $P(Z > 0) = 1$.*

Proof. Exercise 3.3. □

We will state the Theorem of Radon-Nikodym here (without proof) for the special case of probability measures that we will need in the sequel. For the general version and a proof we refer to [1]. This result gives a kind of converse result to Lemma 4.3.9 above.

Theorem 4.3.10 (Radon, Nikodym: special case)**.** *Let μ be a measure on $(\Omega, \mathcal{F})$ with $\mu \ll P$. Then there exists a density function of μ with respect to P, that is, a measurable and P-integrable function $Z : (\Omega, \mathcal{F}) \to ([0, \infty), \mathcal{B}([0, \infty))$ such that for all $A \in \mathcal{F}$:*

$$\mu(A) = \int_A Z dP \quad (= E_P[Z \mathbb{1}_A]) \tag{4.1}$$

If $Y : \Omega \to [0, \infty)$ is another measurable function which satisfies (4.1) then $Y = Z$ P-a.s. The Radon-Nikodym density will be written as $\frac{d\mu}{dP}$.

Lemma 4.3.11. *Suppose $Q \ll P$ are probability measures on $(\Omega, \mathcal{F})$, let $Z = \frac{dQ}{dP}$. Suppose $f \in L^1(\Omega, \mathcal{F}, Q)$, i.e., $E_Q[|f|] < \infty$. Then $Zf \in L^1(\Omega, \mathcal{F}, P)$ and it holds that*

$$E_Q[f] = E_P[Zf].$$

Proof. By Theorem 4.3.10 the Lemma holds for $f = \mathbb{1}_A$ with $A \in \mathcal{F}$. Assume now that $Q(f \geq 0) = 1$. Then there exist f_n with $f_n \uparrow f$ Q-a.s. such that $f_n = \sum_{k=1}^{N_n} \alpha_k^n \mathbb{1}_{A_k^n}$ and $E_Q[f] = \lim_{n\to\infty} E_Q[f_n]$. By linearity and because all sums are finite we get $Zf_n \in L^1(P)$ and

$$E_Q[f_n] = E_P[Zf_n].$$

Observe that $Zf_n \geq 0$ and $Zf \geq 0$, P-a.s. (Why?) Observe further that $Zf_n \uparrow Zf$. Hence we get by monotone convergence with respect to the measure P that

$$E_P[Zf_n] \to E_P[Zf],$$

and as

$$E_P[Zf_n] = E_Q[f_n] \to E_Q[f]$$

because $f \in L^1(Q)$ and by definition of f_n we see that $Zf \in L^1(P)$ and

$$E_Q[f] = E_P[Zf].$$

For the general case of a Q integrable random variable f use the decomposition into the positive and the negative part

$$f = f^+ - f^-,$$

apply the above proved result for the nonnegative f^+ and f^- and use linearity. □

Remark 4.3.12. Suppose the probability space Ω is finite, that is, $\Omega = \{\omega_1, \dots, \omega_N\}$ as we have in the model of the financial market. Suppose again that $P(\omega_n) = p_n > 0$ for $n = 1, \dots, N$. Suppose that $Q \ll P$ is another probability measure. Define $q_n = Q(\omega_n)$. Then we get the Radon Nikodym derivative of Q with respect to P as follows:

$$\frac{dQ}{dP}(\omega_n) = \frac{Q(\omega_n)}{P(\omega_n)} = \frac{q_n}{p_n}, \quad , n = 1, \dots, N.$$

In this case $Q \sim P$ if and only if $q_n > 0$ for all $n = 1, \dots, N$. (Prove the statements of the remark as an exercise!)

Two further properties of Radon-Nikodym derivatives.

Lemma 4.3.13. *Suppose $Q \ll \tilde{P}$ and $\tilde{P} \ll P$. Then $Q \ll P$ and*

$$\frac{dQ}{dP} = \frac{dQ}{d\tilde{P}} \cdot \frac{d\tilde{P}}{dP} \quad P - a.s$$

Proof. Exercise 3.4. □

Lemma 4.3.14. *Suppose $Q \sim P$, that is, $Q \ll P$ and $P \ll Q$. Let $Z = \frac{dQ}{dP}$ which is > 0 a.s. Then*

$$\frac{dP}{dQ} = \frac{1}{Z} \quad a.s$$

Proof. Exercise 3.5. □

Observe that, for equivalent probability measures $Q \sim P$, something holds P-a.s. if and only if it holds Q-a.s. Hence, in this case we can just say "a.s." without stressing the measure. This is for example done in the formulation of the above Lemma 4.3.14.

4.3.4 The proof of the Fundamental Theorem of Asset Pricing with finite Ω

This section will give the proof of Theorem 4.3.8. Note that one implication of the proof is easy and should be done by you in Exercise 3.7. The hard part is to prove that the condition (NA) gives the existence of an equivalent martingale measure. In our situation (finite Ω) the proof can be done by relying on linear algebra. If one wants to generalize the proof for a general Ω one has to use results from functional analysis. Why can we use linear algebra in the finite Ω-case? Suppose we have a random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ where $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$. Then this random variable can be identified with an element of $\mathbb{R}^N$ because the random variable X only takes N values. Indeed, it takes the values $X(\omega_1), \ldots, X(\omega_N)$ which can be written as a vector $\vec{v} \in \mathbb{R}^N$ by

$$\vec{v} = \begin{pmatrix} X(\omega_1) \\ X(\omega_2) \\ \vdots \\ X(\omega_N) \end{pmatrix}.$$

Note that, in the finite Ω case, we get, for all $1 \leq p < \infty$ that

$$E_P[|X|^p] = \sum_{n=1}^{N} |X(\omega_n)|^p P(\omega_n) < \infty,$$

hence all L^p-norms are finite. Moreover the essential sup and with that the norm of the space L^∞:

$$\|X\|_{L^\infty} = \sup_{n=1,\ldots,N} |X(\omega_n)| < \infty.$$

All these norms are equivalent because all norms on $\mathbb{R}^N$ are equivalent. And hence the spaces $L^p(\Omega, \mathcal{F}, P)$, $1 \leq p \leq \infty$, can all be identified with $\mathbb{R}^N$. This is the case, of course, **only if** Ω is finite. In the general case one needs the full theory of these spaces, that is, one needs functional analysis. Moreover, in the general case, it is a substantial difference which p we consider whereas in the finite case everything reduces to the same finite vector space $\mathbb{R}^N$ where all norms are equivalent and where there exists an inner product (namely the usual inner product $(\vec{v}, \vec{w}) = \sum_{i=1}^N v_i w_i$. With the above observations we can proceed with the proof.

Definition 4.3.15. *Define the following linear subspace K of $\mathbb{R}^N$:*

$$\begin{aligned} K &= \{f = (H \cdot S)_T,\ H = (H^1, \ldots, H^d) \textit{ predictable}\} \qquad (4.2) \\ &= \{f = V_T,\ \textit{with } V_0 = 0,\ \hat{H} \textit{ self-financing.}\} \end{aligned}$$

V is the value process of the self-financing trading strategy $\hat{H}$ as in Definition 4.3.3. And $(H \cdot S)_T := \sum_{i=1}^d (H^i \cdot S^i)_T = V_T$ (as $V_0 = 0$).

Convince yourselves that K indeed is a linear subspace of $\mathbb{R}^N$. You again need to identify each V_T with the corresponding vector

$$\begin{pmatrix} V_T(\omega_1) \\ V_T(\omega_2) \\ \vdots \\ V_T(\omega_N) \end{pmatrix} \in \mathbb{R}^N$$

(as we did for the random variable X above) and then prove that K is a linear subspace. Recall from linear algebra how a linear subspace is defined!

The interpretation of the subspace K is: these are all so-called **attainable claims**. This means in the space K you find everything that can be reached by trading in a self-financing way in the risky assets $S^1, \dots S^d$ until the time horizon T with starting investment $= 0$ (because $V_0 = 0$). We can now express the condition of No Arbitrage (NA) with the help of the linear space K.

Lemma 4.3.16. *The condition (NA) holds if and only if*

$$K \cap \mathbb{R}^N_+ = \{\vec{0}\},$$

with $\mathbb{R}^N_+ = \left\{ \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} \in \mathbb{R}^n, v_n \geq 0, n = 1, \dots, N \right\}.$

Proof. Recall the definition of No Arbitrage. The proof is obvious. □

We will use the Separating Hyperplane Theorem of linear algebra which we recall here.

Theorem 4.3.17 (Separating hyperplane in $\mathbb{R}^N$)**.** *Suppose K and M are non-empty, convex subsets of $\mathbb{R}^N$ such that*

$$K \cap M = \emptyset.$$

If K is **closed** *and M is* **compact** *then there exists* $\vec{\varphi} = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_N \end{pmatrix} \in \mathbb{R}^N$ *such that*

$$\sup_{f \in K} (\vec{\varphi}, f) \leq \alpha < \beta \leq \inf_{g \in M} (\vec{\varphi}, g).$$

For the proof of Theorem 4.3.17 we refer to any extensive book on linear algebra.

Proof of Theorem 4.3.8. As said before the proof of ($\Longleftarrow$) is left to you in Exercise 3.7.

($\Longrightarrow$) We saw above that if (NA) holds then $K \cap \mathbb{R}^N_+ = \{\vec{0}\}$. We cannot directly apply the separating hyperplane theorem here because we need an empty intersection. So we will construct an artificial set M that will help us.

First let's check if the properties on K are satisfied. Indeed, K is a linear subspace therefore it is (trivially) also convex. Being a subspace moreover implies that K is a closed subset of $\mathbb{R}^N$ (check any literature on linear subspaces or prove this fact yourselves). So the assumptions on K of Theorem 4.3.17 are satisfied. Now define a convex set M as follows

$$M = \{g = \sum_{i=1}^{N} c_n \mathbb{I}_{\{\omega_n\}}, \text{ where } c_n \geq 0,\ n = 1, \dots, N, \text{ and } \sum_{n=1}^{N} c_n = 1\}.$$

Then $M \subseteq \mathbb{R}^N$ by identifying the random variable g with the vector

$$\begin{pmatrix} g(\omega_1) \\ g(\omega_2) \\ \vdots \\ g(\omega_N) \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}.$$

Moreover, observe that M is a closed and bounded subset of $\mathbb{R}^N$ and therefore a compact set. Note that M clearly is convex. (In fact M is the standard simplex of $\mathbb{R}^N$.) Hence the set M satisfies the assumption of the set M in Theorem 4.3.17.

Let us now check if $K \cap M = \emptyset$. Indeed, by definition, $M \subseteq \mathbb{R}^N_+$. By (NA) we know that $K \cap \mathbb{R}^N_+ = \{\vec{0}\}$. But, clearly $\vec{0} \notin M$ as the components of the all-0 vector obviously do not sum up to 1. Hence, we indeed get that,

$$K \cap M = \emptyset.$$

Now we can apply Theorem 4.3.17 to find $\vec{\varphi} \in \mathbb{R}^N$ with

$$\sup_{f \in K} (\vec{\varphi}, f) \leq \alpha < \beta \leq \inf_{g \in M} (\vec{\varphi}, g).$$

We know that K is a linear space. Hence, suppose we fix $\tilde{f} \in K$, then for every $k > 0$, we get that also $k\tilde{f} \in K$. Suppose $(\vec{\varphi}, \tilde{f}) = a > 0$ then we would have that

$$\lim_{k \to \infty} (\vec{\varphi}, k\tilde{f}) = \lim_{k \to \infty} ka = +\infty,$$

and the $\sup_{f \in K}(\vec{\varphi}, f)$ could not be bounded above by a finite number α which would be a contradiction. Therefore, for every $\tilde{f} \in K$ we get that $(\vec{\varphi}, \tilde{f}) \leq 0$ and hence we find that $\alpha = 0$, i.e.,

$$\sup_{f \in K} (\vec{\varphi}, f) \leq 0 < \beta \leq \inf_{g \in M} (\vec{\varphi}, g).$$

This means, in particular, that for every $g \in M$ we have that $(\vec{\varphi}, g) > 0$. Choose now the particular g-s in M which are the unit vectors of $\mathbb{R}^N$, i.e., choose $g = e_n$ where e_n has a 1 in the n-th coordinate and otherwise 0. This gives

$$0 < (\vec{\varphi}, g) = (\vec{\varphi}, e_n) = \varphi_n,$$

and this holds for all $n = 1, \dots N$. Hence $\vec{\varphi}$ is a vector with strictly positive entries. From this we can define a probability measure on Ω by the following. Denote by $C = \sum_{n=1}^{N} \varphi_n$ and define, for $n = 1, \dots, N$,

$$Q(\omega_n) = q_n = \frac{\varphi_n}{C}.$$

Then, clearly $q_n > 0$, for all n, and $\sum_{n=1}^{N} q_n = 1$. So Q is a probability measure and, moreover $Q \sim P$ (see Remark 4.3.12).

We claim that this is a martingale measure. Indeed, we have to prove that, for $i = 1, \dots, n$, S^i is a martingale with respect to Q. Fix i. It is enough to show that, for each $t \in \{0, \dots, T-1\}$ and $A \in \mathcal{F}_t$, we have that

$$E_Q[\mathbb{1}_A(S^i_{t+1} - S^i_t)] = 0,$$

see the martingale property and the definition of conditional expected value. Note that $f = \mathbb{1}_A(S^i_{t+1} - S^i_t) \in K$ by choosing the trading strategy H as $H^j_u = 0$ for all $j \neq i$ and $u = 1, \dots, T$. And $H^i_{t+1} = \mathbb{1}_A$, otherwise $H^i_u = 0$. Then

$$V_T = \sum_{j=1}^{d} (H^j \cdot S^j)_T = \sum_{j=1}^{d} \sum_{u=1}^{T} H^j_u (S^j_u - S^j_{u-1}) = H^i_{t+1}(S^i_{t+1} - S^i_t) = \mathbb{1}_A(S^i_{t+1} - S^i_t) = f.$$

Hence we get that

$$(\vec{\varphi}, f) \leq 0,$$

and because Q is just $\vec{\varphi}/C$ and $C > 0$ we get

$$(Q, f) = \sum_{n=1}^{N} f(\omega_n) q_n = E_Q[f] \leq 0.$$

Now, because K is a linear space, also $-f \in K$ and hence

$$E_Q[-f] \leq 0,$$

and therefore

$$0 = E_Q[f] = E_Q[\mathbb{1}_A(S^i_{t+1} - S^i_t)].$$

This holds for all assets i and all time points t therefore Q is an equivalent martingale measure. $\square$

Chapter 5

Brownian motion

5.1 Definition and outlook

Recall the definition of Brownian motion from Chapter 3, see Definition 3.1.8. An adapted stochastic process $(W_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ is a standard Brownian motion, if $W_0 = 0$, and

(1) $W_t - W_s \sim \mathcal{N}(0, t-s)\ \forall s < t$

(2) $W_t - W_s$ is independent of $\mathcal{F}_s\ \forall s < t$

(3) $t \mapsto W_t(\omega)$ is continuous for almost all ω.

Brownian motion is an important stochastic process that appears in many applications and is also important in statistics (asymptotic statistics). In the lecture Finanz- und Versicherungsmathematik (UK) we will study financial markets in a continuous time setting. In particular we will have a closer look at the Black-Scholes model where the risky financial asset is modeled by a stochastic process of the following type:

$$S_t = S_0 + \int_0^t \mu S_u du + \int_0^t \sigma S_u dW_u.$$

As we see in the above expression there will appear a stochastic integral with respect to Brownian motion (the dW_u expression). There we need the concept of the Ito integral which will be introduced in the Winter term. We will see that there is a unique solution to the above stochastic integral equation, namely

$$S_t = S_0 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right).$$

With this model for the risky asset (the stock) we will then address analogous topics as in Chapter 4. This means we will have to see if there exists an emm and how to find it. The stochastic integral with respect to Brownian motion will not only appear in the definition of the risky asset but also, as in the finite discrete time setting, when we model trading with self-financing strategies. The technical details are more involved in the continous time setting. This will lead us later on to Ito integral, Ito's formula and Girsanov Theorem.

5.2 Some words on existence: Brownian motion as a weak limit of random walks

There are various proofs of existence of a Brownian motion. One possibility is to derive this stochastic process by an approximation by properly scaled and interpolated random walks. We will give the idea here without the technical details (it involves weak convergence on the space of all continuous functions, i.e., the path space of Brownian motion - recall that all paths of BM are continuous functions). Recall the definition of a random walk $(X_n)_{n=0,1,\ldots}$, see Example 1.1.2. We will now scale a random walk in the appropriate way that will lead us to Brownian motion. Indeed, fix $N \in \mathbb{N}$ and define the random walk at times $t_n = \frac{n}{N}$, $n = 0, 1, \ldots$ (so the distance between the time points is always $\frac{1}{N}$) and with jump size $\frac{1}{\sqrt{N}}$. For $N \to \infty$, the jump as well as the step size converge to 0; however, for very large N, the jump size $\frac{1}{\sqrt{N}}$ is then substantially larger than the step size (the distance between the time points), which is $\frac{1}{N}$. That means, suppose $(X_n)_{n=0,1,\ldots}$ is the random walk of Example 1.1.2, then the **scaled random walk** is given by

$$X_{t_n}^N := \frac{1}{\sqrt{N}} X_n = \frac{1}{\sqrt{N}} \sum_{k=1}^{n} \epsilon_k, \tag{5.1}$$

where $t_n = \frac{n}{N}$, for all $n \in \{0\} \cup \mathbb{N}$. In principle, for $N \to \infty$, this is already the way how to approximate Brownian motion. However, we aim to do the approximation with continuous time stochastic processes instead of discrete time random walks (so that we can work on the path space of BM = continuous functions). Therefore, in an analogous way to the interpolated random walk of Example 1.1.5, we will do an interpolation for the discrete time stochastic process $(X_{t_n}^N)_{n=0,1,\ldots}$ here that leads to a continuous time stochastic process that mimics the behavior of X^N. For $\frac{n}{N} \le t < \frac{n+1}{N}$, i.e.,

$t_n \le t < t_{n+1}$, define

$$Y_t^N = X_{t_n}^N + (\alpha^n t + d^n)\epsilon_{n+1},$$

where we look for α^n and d^n such that (1) $Y_{t_n}^N = X_{t_n}^N$ and (2) $Y_{t_{n+1}}^N = X_{t_{n+1}}^N$, which means that we connect $X_{t_n}^N$ and $X_{t_{n+1}}^N$ by a straight line. This leads us to the following two equations

$$\begin{aligned} &(1) \quad \alpha^n t_n + d^n = 0 \\ &(2) \quad \alpha^n t_{n+1} + d^n = \frac{1}{\sqrt{N}} \end{aligned}$$

Solving the equations gives $\alpha^n = \sqrt{N}$ and $d^n = -\frac{n}{\sqrt{N}}$. Hence for $\frac{n}{N} \le t < \frac{n+1}{N}$ we get

$$\begin{aligned} Y_t^N &= X_{t_n}^N + \left(\sqrt{N}t - \frac{n}{\sqrt{N}}\right)\epsilon_{n+1} \\ &= \frac{1}{\sqrt{N}}\left(\sum_{k=1}^{n}\epsilon_k + (Nt-n)\epsilon_{n+1}\right) \end{aligned}$$

Note that $\frac{n}{N} \le t < \frac{n+1}{N}$ is the same as $n \le Nt < n+1$ which means that $n = \lfloor Nt \rfloor$, hence, in general we can write for all $t \ge 0$:

$$Y_t^N = \frac{1}{\sqrt{N}}\left(\sum_{k=1}^{\lfloor Nt \rfloor}\epsilon_k + (Nt - \lfloor Nt \rfloor)\epsilon_{\lfloor Nt \rfloor + 1}\right).$$

By definition $(Y_t^N)_{t\ge 0}$ is now a stochastic process in continuous time with continuous paths. This means, for almost all ω, we have that $t \mapsto Y_t^N(\omega)$ is a continuous function, i.e., an element of the space of continuous functions on $[0,\infty)$ which is denoted by $\mathcal{C}[0,\infty)$. The result (without proof) is now the following.

Theorem 5.2.1 (Donsker)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a sequence of iid random variables $(\epsilon_k)_{k\ge 1}$ with mean* 0 *and variance* 1. *Define $(Y_t^N)_{t\ge 0}$ as above. Then $(Y_t^N)_{t\ge 0}$ converges to $(W_t)_{t\ge 0}$ weakly.*

A few words to clarify the statement of the theorem.

(i) Let $\tilde{\Omega} = C[0,\infty)$.

(ii) On $\tilde{\Omega} = C[0,\infty)$ one can construct a probability measure P^* (the **Wiener measure**) such that the **canonical process** $(W_t)_{t\ge 0}$ is a Brownian motion. Let $\omega = \omega(t)_{t\ge 0} \in C[0,\infty)$. Then the canonical process is given as $W_t(\omega) := \omega(t)$.

(iii) $(Y_t^N)_{t\geq 0}$ defines a probability measure P^N on $\tilde{\Omega} = C[0,\infty)$. Indeed,

$$Y^N : (\Omega, \mathcal{F}) \to (C[0,\infty), \mathcal{B}(C[0,\infty))$$

measurable, that means we consider the random variable $\omega \mapsto Y^N(\omega, \cdot)$, which maps ω to a continuous function in t (i.e. the path). Let $A \in \mathcal{B}(C[0,\infty))$, then

$$P^N(A) := P(\{\omega : Y^N(\omega, \cdot) \in A\}).$$

(iv) So, Donsker's theorem says that on $(\tilde{\Omega}, \tilde{\mathcal{F}}) = (C[0,\infty), \mathcal{B}(C[0,\infty)))$ we have that $P^N \xrightarrow{\text{w}} P^*$ (weak convergence).

Remark 5.2.2. By the Central Limit Theorem it is easy to see that, for fixed $t \geq 0$, $Y_t^N \to Z$ weakly where $Z \sim \mathcal{N}(0,t)$, that means:

$$P(Y_t^N \leq z) \to \Phi^t(z),$$

for $N \to \infty$, where Φ^t is the distribution function of $\mathcal{N}(0,t)$. Prove this as an exercise. Of course, this is only a very small part of what one has to prove for Donsker's Theorem. There one has to prove the weak convergence of the whole processes Y^N to the standard Brownian motion on the path space $\mathcal{C}[0,\infty)$.

5.3 BM as a Gaussian process

Definition 5.3.1. *A stochastic process* $(X_t)_{t\geq 0}$ *is called* **Gaussian process** *if for all* n *and (finite) families of time points* $\{t_1, ..., t_n\}$ *the random vector*

$$\begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_n} \end{pmatrix}$$

has an n*-variate normal distribution. A Gaussian process is called* **centered** *if* $E[X_t] = 0$ *for all* t*. The* **covariance function** *of the Gaussian process is given by* $\gamma(s,t) = Cov(X_s, X_t)$*.*

The next theorem gives an alternative characterization of BM to Definition 3.1.8.

Theorem 5.3.2 (Characterization of BM as Gaussian process)**.** *A standard Brownian motion is a centered Gaussian process with* $\gamma(s,t) = s \wedge t$*. Conversely, a centered Gaussian process with continuous paths and* $\gamma(s,t) = s \wedge t$ *and* $X_0 = 0$ *is a standard Brownian motion.*

Proof. ($\Rightarrow$) Let $(W_t)_{t\geq 0}$ be a standard Brownian motion. It is centered, because

$$E[W_t] = E[W_t - 0] = E[W_t - W_0] = 0$$

because $W_t - W_0 \sim \mathcal{N}(0, t)$.

Now let's check $\gamma(s, t)$. Without loss of generality, we can assume $s < t$.

$$\begin{aligned}\gamma(s,t) &= \mathrm{Cov}(W_s, W_t) = E[W_s W_t] = E[W_s W_t - W_s^2 + W_s^2] \\ &= E[W_s(W_t - W_s) + W_s^2] = E[W_s(W_t - W_s)] + E[W_s^2] = (*)\end{aligned}$$

and because for a Brownian motion, $W_t - W_s$ is independent of $\mathcal{F}_s$ for $s < t$:

$$(*) = E[W_s]E[W_t - W_s] + E[W_s^2] = 0 + s = s = s \wedge t,$$

as $s < t$.

We still have to show, that for $t_1, ..., t_n$ we have that $\begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_n} \end{pmatrix}$ is n-variate normally distributed. Without loss of generality, let $t_1 < t_2 < ... < t_n$. It suffices to show that there exists a matrix A and iid random variables $Z_1, ..., Z_n \sim \mathcal{N}(0, 1)$ such that

$$\begin{pmatrix} X_{t_1} \\ \vdots \\ X_{t_n} \end{pmatrix} = A \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}.$$

This is left as an exercise (Exercise 4.2 from the collection of all exercises). Hence a standard BM is a Gaussian process with the correct covariation function.

($\Leftarrow$) Suppose $(X_t)_{t\geq 0}$ is a centered gaussian process with $\gamma(s, t) = s \wedge t$ with continuous paths. We have to show, that $(X_t)_{t\geq 0}$ is a Brownian motion. Let $s < t$.

$$X_t - X_s = (-1, 1) \begin{pmatrix} X_s \\ X_t \end{pmatrix},$$

hence $X_t - X_s$ is normally distributed because it's a linear transformation of a random vector that is bivariate normal. Moreover

$$E[X_t - X_s] = E[X_t] - E[X_s] = 0 - 0 = 0$$

$$\begin{aligned}\operatorname{Var}(X_t - X_s) &= E[(X_t - X_s)^2] = E[X_s^2] - 2E[X_tX_s] + E[X_t^2]\\ &= s - 2s + t = t - s\end{aligned}$$

Hence $X_t - X_s \sim \mathcal{N}(0, t-s)$ so the first property of Definition 3.1.8 is satisfied. Let now $u \leq s < t$. We will show that $X_t - X_s$ is independent of X_u. First note that $X_t - X_s$ and X_u are uncorrelated, because

$$E[X_u(X_t - X_s)] = E[X_uX_t] - E[X_uX_s] = u \wedge t - u \wedge s = u - u = 0$$

Then, observe that

$$\begin{pmatrix} X_u \\ X_t - X_s \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} X_u \\ X_s \\ X_t \end{pmatrix},$$

hence the random vector $\begin{pmatrix} X_u \\ X_t - X_s \end{pmatrix}$ is bivariate normal because it is a linear transformation of the multivariate normal vector $\begin{pmatrix} X_u \\ X_s \\ X_t \end{pmatrix}$. Together with the fact that X_u and $X_t - X_s$ are uncorrelated we can now conclude that they are even independent. And so we found the second property of Definition 3.1.8 namely the independence of the past (to be precise we would have to use the properly extended natural filtration here, but that's a technicality).

For the continuity of the paths there is nothing to prove because this was assumed. $\square$

5.4 Markov property and passage times, reflection principle, running maximum and minimum

Brownian motion is a Markov process with (uncountable!) state space $\mathbb{R}$. In this case the Markov property means the following

Definition 5.4.1. *Brownian motion satisfies the Markov property. That means, for all $t \geq 0$ and $u > 0$, it holds that*

$$P(W_{t+u} \leq x | \mathcal{F}_t) = P(W_{t+u} \leq x | W_t) \quad a.s.,$$

where $P(W_{t+u} \leq x|\mathcal{F}_t) = E[\mathbb{1}_{\{W_{t+u} \leq x\}}|\mathcal{F}_t]$ and $P(W_{t+u} \leq x|W_t) = E[\mathbb{1}_{\{W_{t+u} \leq x\}}|\sigma(W_t)]$. Hence the conditional distribution of W_{t+u} with respect to the whole past $\mathcal{F}_t$ only depends on the information of time t, i.e., the σ-algebra that is generated by the random variable W_t (the current position of the particle).

Recall the definition of the first passage time of level a of BM, see Example 3.1.9:

$$T_a(\omega) = \inf\{t > 0 : W_t(\omega) = a\}.$$

We already know that T_a is a stopping time. We are now interested in the distribution of T_a, i.e., in $P(T_a \leq t)$. We use a heuristic argument (proof follows later), the so-called **reflection principle**. Take $a > 0$. It is clear that

$$P(T_a \leq t) = P(T_a \leq t, W_t > a) + P(T_a \leq t, W_t < a), \tag{5.2}$$

where we can use strict inequalities in both cases because $P(W_t = a) = 0$ anyway. Observe that if $W_t(\omega) > a$ the Brownian motion (that starts in 0) must have reached the level a at some time before t because it cannot jump over the level a (continuity of the paths). Hence for such an ω it follows that $T_a(\omega) \leq t$, therefore $\{\omega : T_a(\omega) \leq t, W_t(\omega) > a\} = \{W_t(\omega) > a\}$. Hence

$$P(T_a \leq t, W_t > a) = P(W_t > a).$$

For the second part of equation 5.2 observe from the picture below that for every path with $W_t(\omega) < a$ we can find the reflected path (in red) which satisfies $W_t(\omega) > a$ and hence, heuristically

$$P(T_a \leq t, W_t < a) = P(T_a \leq t, W_t > a)$$

which is $= P(W_t > a)$ by the same argument as above.

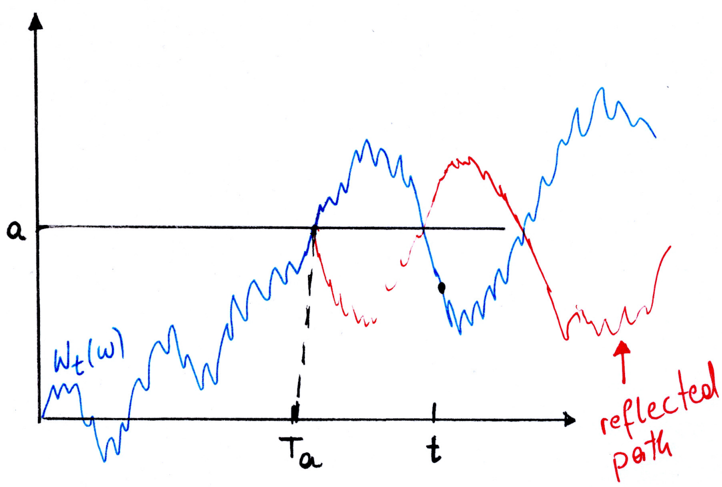

Figure 1

So we get from equation 5.2 that

$$P(T_a \leq t) = P(W_t > a) + P(W_t > a) = 2P(W_t > a).$$

Hence we found in a heuristic way the following lemma.

Lemma 5.4.2 (Distribution of T_a)**.** $P(T_a \leq t) = 2P(W_t > a)$.

For the proof of the lemma that will follow later we need the strong Markov property of BM which we state here without a proof.

Theorem 5.4.3 (Strong Markov property of BM)**.** *The BM satisfies the strong Markov property. That means, for every stopping time τ, which satisfies $P(\tau < \infty) = 1$, it holds that $(\widehat{W}_t)_{t\geq 0}$, given by*

$$\widehat{W}_t := W_{\tau+t} - W_\tau,$$

is a standard Brownian motion independent of

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \textit{ for all } t\}.$$

Proof. Can be found in any standard book on BM. We skip it here because it goes beyond the scope of the lecture. □

In order to apply the strong Markov property for the stopping time T_a we have to show that $P(T_a < \infty) = 1$ first. It will follow from the next lemma which shows that Brownian paths take arbitrarily large positive values and arbitrarily small negative values with probability 1. From this it will follow that, with probability 1, the paths cross each level a in a finite time.

Lemma 5.4.4. $P(\sup_{t\geq 0} W_t = +\infty$ *and* $\inf_{t\geq 0} W_t = -\infty) = 1$.

Before we look at the proof note that the sup as well as the inf are well-defined, there are no measurability issues here. Indeed, the continuity of the paths allows to define both expressions over rational numbers, because, for example, $\sup_{t\geq 0} W_t = \sup_{q\geq 0, q\in\mathbb{Q}} W_q$ a.s. And the countable sup, inf are measurable.

Proof of Lemma 5.4.4. Let $Z = \sup_{t\geq 0} W_t$. As $W_0 = 0$ we obviously have that $Z \geq 0$ a.s. In the exercises you will prove that the stochastic process X^c_t, $t \geq 0$, defined by $X^c_t = cW_{\frac{t}{c^2}}$ is a standard BM as well, see Exercise 4.1 d). In particular this holds for all $c > 0$. Hence

$$Z = \sup_{t\geq 0} X^c_t = c \sup_{t\geq 0} W_{\frac{t}{c^2}} = cZ \quad \text{a.s.,}$$

where the last equation holds because, for fixed $c > 0$, the sup over all t/c^2 is the same as the sup over all t. So we see, that for all $c > 0$ we get

$$Z = cZ \quad \text{a.s.},$$

which can only hold if $Z = 0$ or $= +\infty$ a.s., i.e.,

$$P(Z = 0 \text{ or } Z = +\infty) = 1. \tag{5.3}$$

Our aim is to show that Z is $+\infty$ with probability 1. Suppose $p = P(Z = 0)$. Then

$$\begin{aligned} p = P(Z = 0) &\leq P(W_1 \leq 0 \text{ and } W_u \leq 0 \text{ for all } u \geq 1) \\ &\leq P(W_1 \leq 0 \text{ and } \sup_{t \geq 0}(W_{1+t} - W_1 = 0)), \end{aligned} \tag{5.4}$$

where the inequality in (5.4) is explained as follows: if $W_u \leq 0$ for all $u \geq 1$ then $W_{1+t} \leq 0$ for all $t \geq 0$. Therefore

$$0 \geq \sup_{t \geq 0}(W_1 + W_{1+t} - W_1) = W_1 + \sup_{t \geq 0}(W_{1+t} - W_1), \tag{5.5}$$

but by the Markov property $(W_{1+t} - W_1)_{t \geq 0}$ is a standard Brownian motion again and hence $\sup_{t \geq 0}(W_{1+t} - W_1) = Z$ a.s. and we know already that $P(Z = 0 \text{ or } \infty) = 1$. By the upper bound of 0 in the inequality in (5.5) we see that in this case it cannot be $+\infty$ and hence it follows that $\sup_{t \geq 0}(W_{1+t} - W_1) = 0$ a.s, which explains (5.4). Let's proceed from there, we saw

$$\begin{aligned} p &\leq P(W_1 \leq 0 \text{ and } \sup_{t \geq 0} W_{1+t} - W_1 = 0) \\ &= P(W_1 \leq 0) P(\sup_{t \geq 0} W_{1+t} - W_1 = 0) \\ &= \frac{1}{2} P(Z = 0) = \frac{1}{2} p, \end{aligned} \tag{5.6}$$

where, for (5.6), we use the Markov property for the deterministic $\tau = t$ to get the independence of the two events, such that the probability of the intersection is the product of the probabilities.

So we get for p which is a probability ($0 \leq p \leq 1$) the equation

$$p \leq \frac{1}{2} p.$$

This is only possible if $p = 0$ and hence $P(Z = 0) = 0$. So it follows by (5.3) that, indeed,

$$P(Z = +\infty) = P(\sup_{t \geq 0} W_t = +\infty) = 1.$$

Now, for the inf, observe that by Exercise 4.1 b) (Symmetry) we have that $\tilde{W}_t = -W_t$ is a standard BM as well. By our result above we get

$$\begin{aligned}
1 &= P(\sup_{t\geq 0} \tilde{W}_t = +\infty) \\
&= P(\sup_{t\geq 0}(-W_t) = +\infty) \\
&= P(-\inf_{t\geq 0} W_t = +\infty) \\
&= P(\inf_{t\geq 0} W_t = -\infty).
\end{aligned}$$

Building the intersection of the two probability 1 sets we see that we proved the lemma as

$$P(\sup_{t\geq 0} W_t = +\infty) \text{ and } P(\inf_{t\geq 0} W_t = -\infty) = 1.$$

□

As a consequence we can now prove the **recurrence property** of Brownian motion. Brownian motion is a Markov process with the continuous (i.e., not countable) state space $\mathbb{R}$. Although we did not define a Markov process with a continuous state space in the lecture on Markov processes you might have an idea what recurrence means. Indeed, it can be characterized by the passage times of all levels $a \in \mathbb{R}$, namely by $P(T_a < +\infty) = 1$ for all a.

Corollary 5.4.5. *$P(T_a < +\infty) = 1$ for all $a \in \mathbb{R}$.*

Proof. Let $a \geq 0$. Suppose it would hold that $P(T_a = +\infty) > 0$. Let $\omega \in \{\omega : T_a(\omega) = +\infty\}$. Then it holds that $W_t(\omega) \neq a$ for all $t > 0$. As $W_0 = 0$ and the paths are continuous with probability 1 (we assume ω to be in this probability 1 set w.l.o.g.) it follows that $W_t(\omega) < a$ for all $t > 0$. This implies that $\sup_{t\geq 0} W_t(\omega) \leq a$, hence if $P(T_a = +\infty) > 0$ then

$$P(\sup_{t\geq 0} W_t \leq a) > 0,$$

which is a contradiction to Lemma 5.4.4, hence $P(T_a < +\infty) = 1$ for all $a \geq 0$. By using the inf instead of the sup we get the result for $a < 0$ as well. □

Using Lemma 5.4.2 we can also find a density for the random variable T_a.

Theorem 5.4.6 (Density of T_a)**.** *Suppose $a \neq 0$. Then T_a has the following density*

$$f(t) = \frac{|a|}{\sqrt{2\pi}} t^{-\frac{3}{2}} e^{-\frac{a^2}{2t}}.$$

This is the density of an Inverse Gamma distribution with parameters $\frac{1}{2}$, $\frac{a^2}{2}$ *(Notation:* $IG(\frac{1}{2}, \frac{a^2}{2})$*). This implies that* $E[T_a] = +\infty$.

Proof. Use Lemma 5.4.2 to see that

$$P(T_a \le t) = 2P(W_t \ge a) = 2\int_a^{+\infty} \sqrt{\frac{1}{2\pi t}} e^{-\frac{x^2}{2t}},$$

and from that find the density f as an exercise. □

Note that the fact that $E[T_a] = +\infty$ means that although Brownian motion is recurrent (comes back to each a infinitely often, or, in other words, crosses each level a in finite time with probability 1) it takes infinitely long, in average, to get there. This is an analogous property to the null-recurrence of the symmetric random walk on $\mathbb{Z}$ in discrete time that we saw in the lecture on Markov chains.

Let us now come back to the proof of Lemma 5.4.2 where we just gave a heuristic argument. We need the following theorem.

Theorem 5.4.7 (Reflection principle)**.** *Let* $a \neq 0$ *and* T_a *be the passage time. Define*

$$\widetilde{W}_t = \begin{cases} W_t & \text{for } t \le T_a \\ 2W_{T_a} - W_t = 2a - W_t & \text{for } t > T_a. \end{cases}$$

Then $(\widetilde{W}_t)_{t\ge 0}$ *is again a standard BM.*

Note that the paths of the process $(\widetilde{W}_t)_{t\ge 0}$ exactly look like the following. In Figure 1 (on page 51) that we used for the heuristic argument take the blue path until time T_a and then continue with the red (reflected) path. (Convince yourselves that this is the case.) The proof of Theorem 5.4.7 follows by the strong Markov property of BM and will be postponed for the moment. Let's now use Theorem 5.4.7 to prove Lemma 5.4.2.

Proof of Lemma 5.4.2. Let $a > 0$. As in the heuristic argument, see equation 5.2 we first observe that

$$P(T_a \le t) = P(T_a \le t, W_t > a) + P(T_a \le t, W_t < a). \tag{5.7}$$

Now define the new standard BM $(\widetilde{W}_t)_{t\ge 0}$ as in Theorem 5.4.7. This process has the same properties as the BM $(W_t)_{t\ge 0}$. Hence we get

$$P(T_a \le t, W_t < a) = P(\widetilde{T}_a \le t, \widetilde{W}_t < a),$$

where $\widetilde{T}_a = \inf\{t > 0 : \widetilde{W}_t = a\}$ is the first passage time of $\widetilde{W}_t$. Looking at the definition of $\widetilde{W}_t$ it is clear that this process crosses the level a for the first time exactly at the same time as W_t, hence $\widetilde{T}_a = T_a$ a.s. From this we see that

$$P(\widetilde{T}_a \leq t, \widetilde{W}_t < a) = P(T_a \leq t, \widetilde{W}_t < a).$$

But for $T_a \leq t$ we have that $\widetilde{W}_t = 2a - W_t$ and so we see that

$$P(T_a \leq t, \widetilde{W}_t < a) = P(T_a \leq t, 2a - W_t < a) = P(T_a \leq t, W_t > a) = P(W_t > a),$$

where the last equality follows as we already observed on page 51: if $W_t > a$ the process had to cross the level a at some earlier time, hence this implies $T_a \leq t$ and therefore $\{T_a \leq t, W_t > a\} = \{W_t > a\}$. Now we put everything together in Equation 5.7 and see that

$$P(T_a \leq t) = P(T_a \leq t, W_t > a) + P(T_a \leq t, W_t < a) = 2P(W_t > a).$$

□

The first passage time of Brownian motion is closely connected to the running maximum and the running minimum of Brownian motion. Here is the definition of these stochastic processes.

Definition 5.4.8 (Running max/min)**.** *Define, for each $t \geq 0$,*

$$M_t = \max_{0 \leq s \leq t} W_s \tag{5.8}$$

$$m_t = \min_{0 \leq s \leq t} W_s \tag{5.9}$$

Note that both expressions are well-defined and $\mathcal{F}_t$-measurable (because they can be defined over rational $0 \leq q \leq t$). The adapted stochastic processes $(M_t)_{t \geq 0}$ and $(m_t)_{t \geq 0}$ are the running maximum and running minimum, respectively, of the Brownian motion.

Theorem 5.4.9. *The following statements hold.*

(i) *For each $a > 0$, $P(M_t \geq a) = P(T_a \leq t) = 2P(W_t > a)$.*

(ii) *For each $a < 0$, $P(m_t \leq a) = 2P(W_t < a)$.*

Proof. Regarding (i) observe that if ω is such that $M_t(\omega) \geq a$ then at some point $s \leq t$ the BM must have crossed a, because by the continuity of the paths this is the only way that the running maximum at t can be $\geq a$. Hence

$T_a(\omega) \leq t$. On the other hand, if $T_a(\omega) \leq t$ then at some time $s \leq t$ the BM reached a hence the maximal value $M_t(\omega) \geq a$. So we see that

$$\{\omega : M_t(\omega) \geq a\} = \{\omega : T_a(\omega) \leq t\}$$

and the first equality of (i) is clear. For the second equality of (i) we refer to Lemma5.4.2.

Regarding (ii) fix $a < 0$. Take the standard Brownian motion $(-W_t)_{t\geq 0}$. Apply (i) to this BM and $-a$ (which is > 0) to see that

$$\begin{aligned} 2P(-W_t > -a) &= P(\max_{0\leq s\leq t}(-W_s) \geq -a) \\ &= P(-\min_{0\leq s\leq t} W_s \geq -a) \\ &= P(-m_t \geq -a) = P(m_t \leq a). \end{aligned}$$

It follows that $2P(W_t < a) = P(m_t \leq a)$ and we proved (ii). □

From the above theorem it follows that the joint distribution of BM and its running maximum (running minimum, respectively) can easily be determined. This is a very helpful tool in many applications.

Theorem 5.4.10 (Joint distribution of BM and running max)**.** *Let $y \geq 0$ and $x \leq y$. Then*

$$P(W_t \leq x, M_t \geq y) = P(W_t \geq 2y - x). \tag{5.10}$$

As a consequence the joint distribution of (W_t, M_t) has the following (bivariate) density

$$f_{(W_t,M_t)}(x, y) = \sqrt{\frac{2}{\pi}}\frac{(2y - x)}{t^{\frac{3}{2}}}e^{-\frac{(2y-x)^2}{2t}}.$$

Proof. We will only prove equation 5.10. To find the bivariate density is left as an exercise. First note that

$$P(W_t \leq x, M_t \geq y) = P(W_t \leq x, T_y \leq t),$$

because $\{M_t \geq y\} = \{T_y \leq t\}$, see the proof of Theorem 5.4.9. Define the Brownian motion $(\widetilde{W}_t)_{t\geq 0}$ as in Theorem 5.4.7 for $a = y$. The proof is now very similar to the proof of Lemma 5.4. Indeed, as this is a standard BM as well we get

$$P(W_t \leq x, T_y \leq t) = P(\widetilde{W}_t \leq x, \widetilde{T}_y \leq t),$$

where again $\widetilde{T}_y$ is the corresponding passage time of $\widetilde{W}$. We already know that $\widetilde{T}_y = T_y$ a.s. Hence we get

$$\begin{aligned} P(W_t \le x, M_t \ge y) &= P(\widetilde{W}_t \le x, \widetilde{T}_y \le t) \\ &= P(\widetilde{W}_t \le x, T_y \le t) \\ &= P(2y - W_t \le x, T_y \le t) \\ &= P(W_t \ge 2y - x, T_y \le t) \\ &= P(W_t \ge 2y - x), \end{aligned}$$

where the last equality follows because if $W_t \ge 2y - x$ then, because $x \le y$, it follows that $W_t \ge y$ and this implies that $T_y \le t$, hence $\{W_t \ge 2y - x, T_y \le t\} = \{W_t \ge 2y - x\}$. $\square$

Bibliography

[1] P. Billingsley. Probability and Measure *Wiley Series in Probability and Statistics*, 1995.

[2] R. C. Dalang, A. Morton and W. Willinger. Equivalent martingale measures and no arbitrage in stochastic securities market models. *Stochastics and Stochastic Reports*, 29:185–201, 1990.

[3] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300:463–520, 1994.

[4] F. Delbaen and W. Schachermayer. The fundamental theorem of asset pricing for unbounded stochastic processes. *Math. Ann.*, 312:215–250, 1998.

[5] F. Delbaen and W. Schachermayer. The mathematics of arbitrage *Springer*, 2006.

[6] H. Föllmer and A. Schied. Stochastic finance *De Gruyter Graduate*, 2016.

[7] J. M. Harrison and S. R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11:215–260, 1981.

[8] I. Karatzas and S. Shreve Brownian motion and stochastic calculus *Springer*, 1998.

[9] D. Williams Probability with martingales *Cambride mathematical textbooks*, 1991.