Mathematical finance - lecture notes
(master course)

March 3, 2025

Abstract

These lecture notes build on the work of a number of different authors. Besides the books mentioned below we borrow extensively from lecture notes of Daniel Bartl and Michael Kupper.

1 Martingales and Arbitrage theory in discrete time

Throughout this section we fix a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t=0}^T)$. $X = (X_t)_{t=0}^T$ will denote an adapted process. We shall also assume that $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_T = \mathcal{F}$.

We will use the process X as a model for the asset price. Specifically, the random variable X_t denotes the price of the asset under consideration at time t.

Trades in the asset are modelled through predictable processes. We thus make the following definitions:

Definition 1.1. *A process $H = (H_t)_{t=1}^T$ is called* predictable *or a* trading strategy *if H_t is $\mathcal{F}_{t-1}$-measurable for $t = 1, \ldots, T$.*

The economic interpretation of a trading strategy $H = (H_t)_{t=1}^T$ is that H_t denotes the number of shares that we own from time $t-1$ to time t. As usual the σ-algebra $\mathcal{F}_t$ models the amount of information available at time t. The assumption that H_t should be $\mathcal{F}_{t-1}$-measurable corresponds to the fact that we can use only information available at time $t-1$ to determine how many shares we want to buy at time $t-1$.

If we own H_t shares from time $t-1$ to time t our wealth changes by

$$H_t(X_t - X_{t-1}).$$

Of course such changes in our wealth accumulate over time. This leads to the following

Definition 1.2. *Let H be a trading strategy. The* wealth process / value process / gains from trading process *is given by*

$$V_t := (H \cdot X)_t := \sum_{k=1}^{t} H_k (X_k - X_{k-1}). \tag{1.1}$$

We write S for the set of all values that can be achieved by trading in X wrt a bounded strategy, i.e.

$$S := \{(H \cdot X)_T : H \text{ predictable and bounded.}\}.$$

A fundamental assumption for models of financial markets is that they do not allow for riskless profits, so called *arbitrage opportunities*.

Definition 1.3. *A trading strategy H is an* arbitrage opportunity *if*

$$\mathbb{P}((H \cdot X)_T \geq 0) = 1, \quad \mathbb{P}((H \cdot X)_T > 0) > 0.$$

We say that a market satisfies the no arbitrage *assumption (NA) if there exists no bounded arbitrage strategy.*

Remark 1.4. *Assume that $\mathbb{Q}$ is another measure on $(\Omega, \mathcal{F})$ which is equivalent to $\mathbb{P}$. Then X satisfies (NA) wrt $\mathbb{P}$ iff it satisfies (NA) wrt $\mathbb{Q}$.*

The following result justifies the restriction to bounded trading strategies. Moreover it yields that the arbitrage can be looked in already in *one* period.

Proposition 1.5. *The following are equivalent:*

(i) There exists an arbitrage opportunity.

(ii) There exists $t \in \{1, \dots, T\}$ and $\eta \in L^0(\mathcal{F}_{t-1}; \mathbb{R}^d)$ such that

$$\begin{aligned} \eta \cdot (X_t - X_{t-1}) \geq 0 \ \mathbb{P}\text{-}a.s. \ \text{and} \\ \mathbb{P}(\eta \cdot (X_t - X_{t-1}) > 0) > 0. \end{aligned} \tag{1.2}$$

(iii) There exists $t \in \{1, \dots, T\}$ and $\eta \in L^\infty(\mathcal{F}_{t-1}; \mathbb{R}^d)$ satisfying (1.2).

(iv) There exists a bounded arbitrage opportunity.

Proof.

(a) We show that (i) implies (ii). Let H be an arbitrage opportunity with gains process $V_t = (H \cdot X)_t$. Define

$$t := \inf\{k \in \{0, \dots, T\} : V_k \geq 0 \ \mathbb{P}\text{-as and } \mathbb{P}(V_k > 0) > 0\}.$$

By assumption, $t \leq T$ and either $V_{t-1} = 0$ or $\mathbb{P}(V_{t-1} < 0) > 0$. If $V_{t-1} = 0$ then

$$H_t \cdot (X_t - X_{t-1}) = V_t - V_{t-1} = V_t$$

so that $\eta := H_t$ satisfies (1.2). If $\mathbb{P}(V_{t-1} < 0) > 0$, define $\eta := H_t \mathbb{1}_{\{V_{t-1}<0\}}$, which is $\mathcal{F}_{t-1}$-measurable. Hence,

$$\eta \cdot (X_t - X_{t-1}) = (V_t - V_{t-1})\mathbb{1}_{\{V_{t-1}<0\}} \geq -V_{t-1}\mathbb{1}_{\{V_{t-1}<0\}}$$

which shows that η satisfies (1.2).

(b) We show that (ii) implies (iii). Fix $\eta \in L^0(\mathcal{F}_{t-1})$ such that (1.2) holds. Since $\mathbb{P}(\bigcup_{c\in\mathbb{N}}\{|\eta| \leq c\}) = 1$, continuity of $\mathbb{P}$ implies

$$\mathbb{P}\left(\eta \cdot (X_t - X_{t-1}) > 0, |\eta| \leq c\right) > 0$$

for some $c > 0$. But then $\eta \mathbb{1}_{\{|\eta|\leq c\}} \in L^\infty(\mathcal{F}_{t-1})$ satisfies (1.2).

(c) Suppose that (iii) holds, i.e. there exists $t \in \{1, \ldots, T\}$ and $\eta \in L^\infty(\mathcal{F}_{t-1})$ such that (1.2) holds. Then

$$H_s := \begin{cases} \eta, & s = t, \\ 0, & s \neq t. \end{cases}$$

defines a bounded arbitrage strategy.

(d) It remains to show that (iv) implies (i), but this is obvious.

□

A first goal of this lecture is to understand what type of models satisfy NA. The answer is tightly linked to the concept of martingales:

Definition 1.6. *Assume that X is an integrable process. X is a* martingale *if for $0 \leq s \leq t \leq T$*

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s.$$

X is called sub-martingale *if $\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s$ for all $0 \leq s \leq t \leq T$. X is called* super-martingale *if $\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s$ for all $0 \leq s \leq t \leq T$.*

A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ is called martingale measure *(for X) if X is a $\mathbb{Q}$-martingale, i.e. if $\mathbb{E}_{\mathbb{Q}}[X_t|\mathcal{F}_s] = X_s$ for all $0 \leq s \leq t \leq T$.*

If furthermore, $\mathbb{Q} \sim \mathbb{P}$, $\mathbb{Q}$ is called an equivalent martingale measure. *The set of all equivalent martingale measures will be denoted by $\mathcal{M}$.*

Apparently X is a martingale if $\mathbb{E}[X_n|\mathcal{F}_{n-1}] = X_{n-1}$ for $n = 1, \ldots, T$.

Lemma 1.7. *If H is a bounded trading strategy and X is a martingale, then the process $((H \cdot X)_t)_{t=0}^T$ is a martingale as well.*

X is a martingale iff $\mathbb{E}(H \cdot X)_T = 0$ for all bounded trading strategies H.

Proof. Exercise. □

Theorem 1.8. *For a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ with $\mathbb{Q} \ll \mathbb{P}$, the following are equivalent:*

(i) The measure $\mathbb{Q}$ is a martingale measure.

(ii) Whenever H is a bounded trading strategy, then $\mathbb{E}_{\mathbb{Q}}(H \cdot X) = 0$.

(iii) Whenever H is a bounded trading strategy, then the corresponding gains process $(H \cdot X)$ is a $\mathbb{Q}$-martingale.

(iv) Whenever H is a trading strategy such that $\mathbb{E}_{\mathbb{Q}}[(H \cdot X)_T^-] < \infty$, then the corresponding gains process $(H \cdot X)$ is a $\mathbb{Q}$-martingale (and in particular, $(H \cdot X)_T$ is $\mathbb{Q}$-integrable).

Proof. The first equivalences are a simple consequence of Lemma 1.7. To establish (iv) is technically more subtle and we refer the reader to [3]. □

Lemma 1.9. *If X is a martingale, then X satisfies NA. More generally, if there exists an equivalent martingale measure for X, then X satisfies NA.*

Proof. Exercise. □

1.1 The Fundamental Theorem of Asset Pricing (FTAP)

Remarkably the converse of Lemma 1.9 is true as well:

Theorem 1.10. *Assume that X satisfies (NA). Then there exists an equivalent probability measure $\mathbb{Q} \sim \mathbb{P}$ such that $X_t \in L^1(\mathbb{Q})$ and X is a $\mathbb{Q}$-martingale.*

In fact Theorem 1.10 is a relatively non trivial result first proved by Dalang, Morton and Willinger [1]. Typical approaches to this result require a non negligible amount of functional analysis. We refer to [2] for a modern account. Here we will content ourselves with a proof under the additional assumption that Ω has only finitely many elements.

Assume from now until the end of Section 1.1 that we work under the following

Assumption 1.11. *Assume that $\Omega = \{\omega_1, \ldots, \omega_N\}$, where $\omega_i \neq \omega_j$ for $i \neq j$ and that $\mathbb{P}(\{\omega\}) > 0$ for all $\omega \in \Omega$. Moreover we assume that $\mathcal{F}_T = \mathcal{F}$ is the power set of Ω and that $\mathcal{F}_0 = \{\emptyset, \Omega\}$.*

It follows that all random variables on Ω are bounded and hence the vector space of all random variables consists of

$$V := L^\infty := L^\infty(\mathbb{P}).$$

Moreover, making the identification $\mathbb{1}_{\omega_i} = (\underbrace{0,\dots,0}_{i-1},1,\underbrace{0,\dots,0}_{N-i})$ we find that $V \cong \mathbb{R}^N$. The dual space V^d is of course again isomorphic to $\mathbb{R}^N$. In the present context, it can also be identified with $L^1 := L^1(\Omega)$, or, more importantly with the vector space $\mathcal{SM}$ of all *signed measures* on Ω. Specifically, any linear functional on V is given by a mapping

$$Z \mapsto \langle Z, \sigma\rangle := \int Z\, d\sigma,$$

for a signed measure $\sigma \in \mathcal{SM}$.

We also write L^+ for the cone of all non-negative random variables and $\mathcal{SM}^+$ for the cone of non-negative measures.

In the proof of Theorem 1.10 we will need the following version of the Hahn-Banach theorem:

Theorem 1.12. *Assume that $A, B \subseteq \mathbb{R}^N$ are closed convex sets and that B is compact. Then there exist $y \in (\mathbb{R}^N)^d$ and $\alpha, \beta \in \mathbb{R}$ such that*

$$\forall a \in A, b \in B, \quad \langle a, y\rangle \le \alpha < \beta \le \langle b, y\rangle.$$

If A is a subspace, then $\langle a, y\rangle = 0$ for all $a \in A$ and in particular one can choose $\alpha = 0$.

Proof of Theorem 1.10. (NA) is equivalent to $S \cap L^+ = \{0\}$. Denoting

$$B := \left\{ \sum_{i=1}^N \lambda_i \mathbb{1}_{\omega_i} : \sum_{i=1}^N \lambda_i = 1, \lambda_i \ge 0 \right\},$$

we thus have

$$B \cap S = 0.$$

By Theorem 1.12 applied to the subspace S and the convex compact set B, there exists $\sigma \in \mathcal{SM}_+$ such that

$$\forall G \in S, \forall X \in B, \quad \langle G, \sigma\rangle \le 0 < \langle X, \sigma\rangle.$$

Since $\sigma(\{\omega\}) = \langle \mathbb{1}_\omega, \sigma\rangle > 0$ for $\omega \in \Omega$, we have that σ is a positive measure and in fact $\sigma \sim \mathbb{P}$. Setting $\mathbb{Q} := \sigma/\sigma(\Omega)$ we obtain a probability measure satisfying $\mathbb{Q} \sim \mathbb{P}$.

Since S is a subspace, we have for $G \in S$, $\langle G, \sigma\rangle = 0$, hence also $0 = \langle G, \mathbb{Q}\rangle = \mathbb{E}_{\mathbb{Q}}[G]$. By Lemma 1.7 this yields that $\mathbb{Q}$ is a martingale measure as desired. $\square$

1.2 European contingent claims

Throughout this section we assume that there is an equivalent martingale measure, i.e. $\mathcal{M} \neq \emptyset$.

Definition 1.13. *A non-negative random variable C on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a European contingent claim. A European contingent claim is called a derivative of the underlying asset X if C is measurable with respect to $\sigma(X) = \sigma(X_0, \dots, X_T)$.*

Remark 1.14. *By a basic result of measure theory, a contingent claim C is a derivative of X iff there exists a Borel-measurable function $f \geq 0$ such that $C = f(X_0, \dots, X_T)$.*

Remark 1.15. *A European contingent claim has the interpretation of an asset which yields at time T the amount C, depending on the scenario of the market evolution. Here, T is called the expiration date or the maturity of C.*

Example 1.16.

(i) European call option: $C^{\text{call}} = (X_T - K)^+$. The owner of a European call option has the right but not the obligation to buy the asset i at time T for a fixed strike price K.

(ii) European put option: $C^{\text{put}} = (K - X_T)^+$. The owner of an European put option has the right but not the obligation to sell the asset i at time T for a fixed strike price K.

(iii) Path dependent contingent claim, e.g. up-and-out call:

$$C^{\text{call}}_{\text{u\&o}} = \begin{cases} (X_T - K)^+, & \max_{t \in \{0,\dots,T\}} X_t < B, \\ 0, & \textit{otherwise.} \end{cases}$$

Definition 1.17. *A contingent claim C is called attainable (replicable) if there exist $a \in \mathbb{R}$ and a trading strategy H such that $a + (H \cdot X)_T = C$ $\mathbb{P}$-a.s. In this case (a, H) is called a replicating strategy. The value a is interpreted as the* initial capital / initial endowment *and $V_t := a + (H \cdot X)_t$ is the corresponding* value process.

Remark 1.18. *The concept of replication plays a central role in mathematical finance.*

Let C be a contingent claim with replicating strategy (a, H), i.e. $C = a + (H \cdot X)_T$. We assume that a rational financial agent is able to figure out such a trading strategy. But this implies that she should be indifferent *between owning the claim C and the initial capital a: If she owns a Euros at initial time and would like to switch this for a payoff C at the terminal time T, all she has to do is to invest in the market according to the strategy H.*

In this sense a is the 'fair value' or 'fair price' of the contingent claim C.

If there exists a strategy H such that $C = (H \cdot X)_T$, then we say that C is attainable at price 0. *We denote by*

$$S := \{(H \cdot X)_T : H \textit{ predictable}\}$$

the set of all claims that are attainable at price 0.

Theorem 1.19. *Let C be an attainable claim. Then $\mathbb{E}_{\mathbb{Q}}[C] < \infty$ for every $\mathbb{Q} \in \mathcal{M}$. The value process of any replicating strategy (a, H) satisfies $V_t = \mathbb{E}_{\mathbb{Q}}[C|\mathcal{F}_t]$ $\mathbb{P}$-a.s. for all $t \in \{0, \dots, T\}$ and $\mathbb{Q} \in \mathcal{M}$. In particular, the value process V is an $\mathcal{M}$-martingale, i.e. a $\mathbb{Q}$-martingale for every $\mathbb{Q} \in \mathcal{M}$.*

Proof. Let $V_t = a + (H \cdot X)_t$ be the value process to the trading strategy H with $V_T = C \geq 0$. Theorem 1.8 shows that for every $\mathbb{Q} \in \mathcal{M}$ the value process V is a $\mathbb{Q}$-martingale, that is $V_t = \mathbb{E}_{\mathbb{Q}}[V_T|\mathcal{F}_t]$ $\mathbb{P}$-a.s. for every $t \in \{0, \dots, T\}$. In particular, $\mathbb{E}_{\mathbb{Q}}[C] = \mathbb{E}_{\mathbb{Q}}[V_T] = V_0 \in \mathbb{R}$. □

The proof of Theorem 1.19 is more subtle than it might seem: to be able to apply Theorem 1.8 it was necessary to guarantee some integrability of the negative part which is the reason for making our standing assumption that European contingent claims are non negative. (Of course this assumption is met in all practical instances.) Clearly we do not need to care for such subtleties under the simplifying Assumption 1.11.

Remark 1.20. *In the setup of Theorem 1.19 we have:*

(i) The value process V does not depend on the replicating strategy (a, H).

(ii) We have $\mathbb{E}_{\mathbb{Q}_1}[V_T|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}_2}[V_T|\mathcal{F}_t]$ $\mathbb{P}$-a.s. for all $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{M}$.

1.3 Complete markets

Definition 1.21. *An arbitrage-free market model is called complete, if every contingent claim is attainable.*

Apparently, complete market models are particularly convenient from a theoretical perspective. In fact, complete markets are most commonly used in practice for their simplicity (even though this condition is not met in reality).

The following result characterises *completeness* in terms of martingale measures. (It is thus analogous to the FTAP which characterises absence of arbitrage in terms of martingales.)

Theorem 1.22 (Second fundamental theorem of asset pricing)**.** *An arbitrage-free market model is complete if and only if $|\mathcal{M}| = 1$.*

As in the case of the first FTAP, one direction of the proof is (essentially) trivial:

Proof of Theorem 1.22, easy part. If the market is complete, then for every $A \in \mathcal{F}_T$, the contingent claim $C := \mathbb{1}_A$ is attainable. By Theorem 1.19, the function

$$\mathcal{M} \to [0, 1],\ \mathbb{Q} \mapsto \mathbb{E}_{\mathbb{Q}}[\mathbb{1}_A] = \mathbb{Q}(A)$$

is constant so that $|\mathcal{M}| = 1$. □

For the converse direction, we work again under the simplifying Assumption 1.11. In a first step we provide an analogue of Lemma 1.7. While Lemma 1.7 characterizes martingales in terms of trading strategies, we now establish that trading strategies can be characterized in terms of martingales.

Recall that S denotes the set of all claims that are attainable at price 0. The following lemma provides a fundamental characterization of S in terms of martingale measures.

Lemma 1.23. *Assume that the market model satisfies NA. Let Z be a bounded random variable such that $\mathbb{E}_{\mathbb{Q}}[Z] = 0$ for every martingale measure $\mathbb{Q}$. Then $Z \in S$, i.e. there exists a trading strategy H such that $(H \cdot X)_T = Z$.*

Proof. Given a vector space W with dual space W^d and $K \subseteq W$ we write

$$K^\perp = \{y \in W^d : \langle x, y\rangle = 0 \text{ for all } x \in K\}.$$

Denote the linear space generated by a subset K of W by $\operatorname{span}(K)$. Assuming that W is finite dimensional, it is well known from linear algebra that $(K^\perp)^\perp = \operatorname{span}(K)$. (In fact this is a simple consequence of the Hahn-Banach theorem.)

Recall that $S = \{(H \cdot X)_T : H \text{ is a trading strategy}\}$ and that $\mathcal{SM}, \mathcal{SM}^+$ denote the sets of signed and non negative measures on Ω, respectively. We write $\mathcal{SM}^{++}$ for the set of all 'strictly postive' measures, i.e. non-negative measures that are equivalent to $\mathbb{P}$.

We claim that

$$\operatorname{span}(S^\perp \cap \mathcal{SM}^{++}) = S^\perp. \tag{1.3}$$

Since $S^\perp$ is a subspace, the inclusion $\operatorname{span}(S^\perp \cap \mathcal{SM}^{++}) \subseteq S^\perp$ is trivial.

To prove the converse assume that $\sigma \in S^\perp$. Let $\mathbb{Q}$ be an equivalent martingale measure and recall that $\mathbb{Q} \in S^\perp$. Pick $\alpha \in \mathbb{R}_+$ such that $\sigma + \alpha\mathbb{Q} \in \mathcal{SM}^{++}$. Then we also have $\sigma + \alpha\mathbb{Q} \in S^\perp \cap \mathcal{SM}^{++}$. Since $\mathbb{Q}$ itself is also an element of $S^\perp \cap \mathcal{SM}^{++}$, we have $\sigma \in \operatorname{span}(S^\perp \cap \mathcal{SM}^{++})$, establishing (1.3).

Assume now that $\mathbb{E}_Q[Z] = 0$ for all equivalent martingale measures $\mathbb{Q}$. Then we also have $\mathbb{E}_\sigma[Z] = 0$ for all $\sigma \in S^\perp \cap \mathcal{SM}^{++}$, hence

$$Z \in (S^\perp \cap \mathcal{SM}^{++})^\perp = (S^\perp)^\perp = S.$$

□

Proof of Theorem 1.22, interesting part. Write $\mathbb{Q}$ for the unique equivalent martingale measure and let C be a contingent claim. We need to prove that C is attainable.

Set $a := \mathbb{E}_{\mathbb{Q}}C$. Then, trivially,

$$\mathbb{E}_{\tilde{\mathbb{Q}}}(C - a) = 0$$

for all equivalent martingale measures $\tilde{\mathbb{Q}}$. By Lemma 1.23 there is a trading strategy H such that $C - a = (H \cdot X)_T$. Hence (a, H) is a replicating strategy for C. □

The following result is important in view of our subsequent development of the theory of derivative pricing in continuous time.

Theorem 1.24 (Martingale representation theorem). *For $\mathbb{Q} \in \mathcal{M}$, the following statements are equivalent:*

(i) We have $\mathcal{M} = \{\mathbb{Q}\}$ (the market is complete).

(ii) Every $\mathbb{Q}$-martingale M has the representation

$$M_t = M_0 + \sum_{k=1}^{t} H_k \cdot (X_k - X_{k-1})$$

for some predictable process H.

Proof. Exercise. □

1.4 Pricing by no arbitrage

We extend our previous definition of arbitrage to a market which contains a financial derivative. The purpose is to define the notion of 'arbitrage free price'.

During the entire section we assume that the market model is free of arbitrage.

Definition 1.25. *Let C be a financial derivative and $\pi^C \in \mathbb{R}$. We say that the price π^C introduces arbitrage if there exist $\alpha \in \mathbb{R}$ and a trading strategy H such that*

$$\mathbb{P}((H \cdot X)_T + \alpha(C - \pi^C) \geq 0) = 1, \quad \mathbb{P}((H \cdot X)_T + \alpha(C - \pi^C) > 0) > 0.$$

Otherwise we call π^C an arbitrage free price.

Remark 1.26. *If C is attainable through a strategy (a, H) then a is the unique arbitrage free price.*

Proof. Exercise. □

Theorem 1.27. *The set of arbitrage-free prices of a claim C is non-empty and given by*

$$\Pi(C) = \{\mathbb{E}_{\mathbb{Q}}[C] : \mathbb{Q} \in \mathcal{M} \text{ such that } \mathbb{E}_{\mathbb{Q}}[C] < \infty\}.$$

Proof. We refer to [3] for a proof. Under the simplifying Assumption 1.11 it is a relatively straight forward application of Lemma 1.23 □

2 Brownian Motion and foundations of stochastic analysis

Brownian motion is the most important stochastic process in continuous time. It will serve as the fundamental building block of all continuous time models considered in this lecture. In this section we sketch its construction and some of its basic properties.

Throughout this section $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ denotes a filtered probability space in continuous time. Whenever we consider a stochastic process, we implicitly assume that it is adapted wrt. this basis. We will also need that $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough to support a RV X that is continuously distributed (i.e. $\mathbb{P}(X = a) = 0$ for all $a \in \mathbb{R}$).

2.1 Construction of (pre) Brownian motion

Definition 2.1. *A process $B = (B_t)_{t\geq 0}$ is called a (standard) Brownian motion if it satisfies the following:*

1. *'start in 0': $B_0 = 0$.*
2. *'Gaussian increments': for $0 \leq s \leq t$ we have $B_t - B_s \sim N(0, t - s)$.*
3. *'independent increments': given $t_0 \leq \ldots \leq t_n$, the increments $B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent.*
4. *'continuous paths': for almost all ω the path $t \mapsto B_t(0)$ is continuous.*

If the process B satisfies only properties (1)-(3), then B is called a pre Brownian motion.

A process that satisfies (4) is called a *continuous process*.

If instead of (1) we have that $B_0 \sim a, a \in \mathbb{R}$, we say that B is a Brownian motion started in a. More generally, if Y is a random variable independent of $(B_t - B_0)$ for $t \geq 0$ and $B_0 = Y$ then we say that B is a Brownian motion started in Y.

If B is a Brownian motion and $B_t - B_s$ is independent of $\mathcal{F}_s$ for $0 \leq s \leq t$, then we say that B is a Brownian motion wrt. $(F_t)_{t\geq 0}$. This property is satisfied automatically if we take $(F_t)_{t\geq 0}$ to be the filtration generated by B. (Warning: It is often assumed that B is a Brownian motion wrt. the underlying filtration without making this point explicit.)

Next we sketch the proof that a Brownian motion exists. Our starting point is the following alternative characterization of a pre Brownian motion:

Lemma 2.2. *A stochastic process is a pre Brownian motion iff it satisfies*

1. *$B_0 = 0$.*
2. *$\mathbb{E}[B_s B_t] = s \wedge t$ for $0 \leq s \leq t$.*

3. For all $t_0 \leq \ldots \leq t_n$, $(B_{t_0}, \ldots, B_{t_n})$ is centered Gaussian.

Proof. Exercise using the properties of Gaussian RV. □

An important point of this characterisation is that properties (1) and (2) are properties of the Hilbert space $L^2(\Omega, \mathbb{P})$: (1) asserts that $\|B_0\|_1 = 0$, (2) is equivalent to $\langle B_s, B_t \rangle = s \wedge t$.

We also make the important comment that one can easily find an example of a Hilbert space and a family of elements, such that these properties are satisfied: Indeed we can just take

$$V = L^2(\mathbb{R}_+, \lambda), f_t := \mathbb{1}_{[0,t)}, t \geq 0. \tag{2.1}$$

Then V is a Hilbert space, $\|f_0\|_0 = 0$, and

$$\langle f_s, f_t \rangle = \int \mathbb{1}_{[0,s)} \mathbb{1}_{[0,t)} \, d\lambda = s \wedge t.$$

The idea behind our construction of a pre Brownian motion is to embed this space V into a 'large' space consisting entirely of Gaussian random variables. To formalize this, we start with the following definition.

Definition 2.3. *A Gaussian space is a closed subspace $\Gamma \subseteq L^2(\Omega, \mathbb{P})$ such that $(X_0, \ldots, X_n)$ is centered Gaussian for all $X_1, \ldots, X_n \in \Gamma$.*

Lemma 2.4. *If $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough to support a RV X that is continuously distributed, then $L^2(\Omega, \mathbb{P})$ contains an infinite dimensional Gaussian space.*

Sketch of proof. It is not hard to see that $(\Omega, \mathcal{F}, \mathbb{P})$ also supports a sequence $X_1, X_2, \ldots$ of iid RV satisfying $X_i \sim N(0,1), i \geq 1$. It then can be shown (using properties of multivariate Gaussians) that the closed spaced generated by $X_1, X_2, \ldots$ is Gaussian. □

Using the above ingredients, we can now establish the existence of a pre Brownian motion.

Theorem 2.5. *There exists a pre Brownian motion.*

Proof. Let $\Gamma \subseteq L_2(\Omega, \mathbb{P})$ be an infinite dimensional Gaussian space. Let $V, (f_t)_{t \geq 0}$ be as in (2.1). Since V is a separable Hilbertspace, there exists an isometry ϕ which embeds V into Γ. Set $B_t = \phi(f_t)$ for $t \geq 0$. Then $\|B_0\|_0 = 0$ and

$$\mathbb{E}[B_s B_t] = \langle f_s, f_t \rangle = s \wedge t$$

for $0 \leq s \leq t$. Hence $B = (B_t)_{t \geq 0}$ is a pre Brownian motion. □

Naturally we could hope that every pre Brownian motion automatically has continuous paths. This is not the case. Indeed, if B is a Brownian motion satisfying then there exists a process $\tilde{B} = (\tilde{B}_t)_{t \geq 0}$ satisfying $B_t = \tilde{B}_t$, $\mathbb{P}$-a.s. for every t such that the path $t \mapsto B_t(\omega)$ is *not* continuous for every $\omega \in \Omega$. However, $\tilde{B}$ is still a pre Brownian motion. In fact the process $\tilde{B}$ can even be chosen in such a way that $t\omega B_t(\omega)$ is $\mathbb{P}$-a.s. nowhere continuous.

These considerations motivate the following notions of 'similarity' for stochastic processes.

Definition 2.6. *Let $X = (X_t)_{t\geq 0}$ and $X' = (X'_t)_{t\geq 0}$ be stochastic processes on a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{g\geq 0})$.*

X and X' are indistinguishable *if there exists a (measurable) set $\Omega_1, \mathbb{P}(\Omega_1) = 1$ such that*

$$X_t(\omega) = X'_t(\omega) \text{ for all } \omega \in \Omega_1, t \geq 0.$$

X and X' are modifications *(of each other) if*

$$X_t(\omega) = X'_t(\omega) \ \mathbb{P}\text{-a.s. for all } t \geq 0.$$

Let $X = (X_t)_{t\geq 0}$ and $X' = (X'_t)_{t\geq 0}$ be stochastic processes defined on stochastic bases $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{g\geq 0})$ and $(\Omega', \mathcal{F}', \mathbb{P}', (\mathcal{F}'_t)_{g\geq 0})$, resp. Then X is a version *of X' if X and X' have the same finite dimensional distributions.*

Apparently 'indistinguishabilty' is stronger than 'being modifications' which in turn is stronger than 'being versions'. If process X, X' are indistinguishable and X is continuous, then X' is continuous as well. In contrast, the modification of a continuous process is not continuous in general.

Notably the notions of modification and indistinguishability coincide for continuous processes:

Proposition 2.7. *Assume that X, X' are continuous processes that are modifications of each other. Then X and X' are indistinguishable.*

Proof. Exercise. □

The notion of 'version' allows us to formalize a uniqueness property of (pre) Brownian motion:

Proposition 2.8 (Uniqueness of pre Brownian motion)**.** *Assume that B, B' are Brownian motions (potentially) on different probability spaces. Then B is a version of B'.*

Proof. Exercise. □

The final step of constructing a Brownian motion will consist in defining an adequate modification of a pre Brownian motion which does have continuous paths.

2.2 The Kolmogorov continuity theorem - final step of construction of Brownian motion

The Kolmogorov extension theorem is a very useful tool, since it provides existence results for stochastic processes.

Definition 2.9. *A function $f\colon [0, \infty) \to \mathbb{R}^d$ is Hölder continuous with exponent $\gamma > 0$, if there exists a constant $c > 0$ such that $|f(t) - f(s)| \leq c|t - s|^\gamma$ for all $s, t \in [0, \infty)$.*

Theorem 2.10 (Kolmogorov continuity theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a d-dimensional process $(X_t)_{t\in[0,T]}$. If there exist $\alpha, \varepsilon, C > 0$ such that*

$$\mathbb{E}[|X_t - X_s|^\alpha] \le C|t-s|^{1+\varepsilon}$$

for all $s, t \in [0,T]$, then there exists a continuous modification $(\tilde{X}_t)_{t\in[0,T]}$ which is locally Hölder continuous with exponent $\gamma \in (0, \frac{\varepsilon}{\alpha})$. That is, there exists a positive random variable h and a constant $\delta > 0$ such that

$$\mathbb{P}\left(\left\{\omega \in \Omega : \sup_{\substack{0<t-s<h(\omega)\\ s,t\in[0,T]}} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t-s|^\gamma} \le \delta\right\}\right) = 1.$$

Important note: The proof of the Kolmogorov continuity theorem is rather technical. I have included it in the lecture notes for completeness but I suggest you might skip it.

Proof. For simplicity, assume $d = 1$ and $T = 1$. For every $n \in \mathbb{N}$ define $\mathcal{D}_n := \{k2^{-n} : k \in \{0, \dots, 2^n\}\}$. Let $\mathcal{D} := \bigcup_{n\in\mathbb{N}} \mathcal{D}_n$ be the set of dyadic rationals in $[0,1]$. Fix $\gamma \in [0, \frac{\varepsilon}{\alpha})$. For $n \in \mathbb{N}$ define

$$A_n := \left\{\max_{k\in\{1,\dots,2^n\}} \left|X_{k2^{-n}} - X_{(k-1)2^{-n}}\right| \ge 2^{-\gamma n}\right\}.$$

By Chebyshev's inequality

$$\begin{aligned}
\mathbb{P}(A_n) &= \mathbb{P}\left(\bigcup_{k=1}^{2^n} \left|X_{k2^{-n}} - X_{(k-1)2^{-n}}\right| \ge 2^{-\gamma n}\right) \\
&\le \sum_{k=1}^{2^n} \mathbb{P}\left(\left|X_{k2^{-n}} - X_{(k-1)2^{-n}}\right| \ge 2^{-\gamma n}\right) \\
&\le \sum_{k=1}^{2^n} \frac{C2^{-n(1+\varepsilon)}}{2^{-\gamma\alpha n}} \\
&= C2^{-n(\varepsilon-\gamma\alpha)}.
\end{aligned}$$

Since $\gamma\alpha < \varepsilon$, we obtain $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$. The Borel-Cantelli lemma implies that $\mathbb{P}(\bigcap_{n\in\mathbb{N}} \bigcup_{m\ge n} A_m) = 0$. Hence, there exists a set $\Omega^* \in \mathcal{F}$ with $\mathbb{P}(\Omega^*) = 1$ and a random variable $n^* \colon \Omega \to \mathbb{N}_0$ such that for all $\omega \in \Omega^*$ and $n \ge n^*(\omega)$ we have

$$\max_{k\in\{1,\dots,2^n\}} \left|X_{k2^{-n}}(\omega) - X_{(k-1)2^{-n}}(\omega)\right| < 2^{-\gamma n}. \tag{2.2}$$

Fix $\omega \in \Omega^*$ and $n \ge n^*(\omega)$. We claim that for every $m > n$ one has

$$|X_t(\omega) - X_s(\omega)| \le 2\sum_{j=n+1}^{m} 2^{-\gamma j} \tag{2.3}$$

for all $s, t \in \mathcal{D}_m$ with $0 < t - s < 2^{-n}$.

Indeed, for $m = n+1$ it follows that $t = k2^{-m}$, $s = (k-1)2^{-m}$ and (2.3) follows from (2.2). By induction, suppose (2.3) is valid for $m \in \{n+1, \dots, M-1\}$. Fix $s, t \in \mathcal{D}_M$ with $s < t$ and define

$$\begin{aligned} t^1 &:= \max\{u \in \mathcal{D}_{M-1} : u \leq t\}, \\ s^1 &:= \min\{u \in \mathcal{D}_{M-1} : u \geq s\} \end{aligned}$$

so that $s \leq s^1 \leq t^1 \leq t$, $s^1 - s \leq 2^{-M}$ and $t - t^1 \leq 2^{-M}$. From (2.2) we have $|X_{s^1}(\omega) - X_s(\omega)| \leq 2^{-\gamma M}$, $|X_t(\omega) - X_{t^1}(\omega)| \leq 2^{-\gamma M}$ and from (2.3) with $m = M-1$ we have $|X_{t^1}(\omega) - X_{s^1}(\omega)| \leq 2\sum_{j=n+1}^{M-1} 2^{-\gamma j}$. Hence,

$$\begin{aligned} |X_t(\omega) - X_s(\omega)| &\leq |X_t(\omega) - X_{t^1}(\omega)| + |X_{t^1}(\omega) - X_{s^1}(\omega)| + |X_{s^1}(\omega) - X_s(\omega)| \\ &\leq 2 \cdot 2^{\gamma M} + 2 \sum_{j=n+1}^{M-1} 2^{-\gamma j} \\ &= 2 \sum_{j=n+1}^{M} 2^{-\gamma j}. \end{aligned}$$

Finally, we show that $(X_t(\omega))_{t \in \mathcal{D}}$ is Hölder continuous for every $\omega \in \Omega^*$. For $s, t \in \mathcal{D}$ with $0 < t - s < h(\omega) := 2^{-n^*(\omega)}$ we choose $n \geq n^*(\omega)$ such that $2^{-(n+1)} \leq t - s < 2^{-n}$. It follows from (2.3) that

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^{\infty} 2^{-\gamma j} \leq \delta |t-s|^\gamma$$

for all $0 < t - s < h(\omega)$, where $\delta := \frac{2}{1-2^{-\gamma}}$. That $(X_t(\omega))_{t \in [0,1]}$ is Hölder continuous follows from standard approximation results. □

Corollary 2.11. *Let B' be a pre Brownian motion. Then B has a continuous modification. In particular there exists a Brownian motion on every probability space which supports a continuously distributed random variable.*

Proof. Exercise. □

2.3 Properties of Brownian motion

In the following lemma we collect some simple properties of Brownian motion.

Lemma 2.12. *1. Let $B = (B_t)_{t \geq 0}$ be a Brownian motion. Then $B' = (1/aB_{a^2 t})_{t \geq 0}$ is a also a Brownian motion for $a > 0$.*

2. Let $B = (B_t)_{t \geq 0}$ be a Brownian motion. Then $B' = (-B_t)_{t \geq 0}$ is a Brownian motion as well.

3. *Let B be a Brownian motion wrt. a filtration $(\mathcal{F}_t)_{t\geq 0}$. Then B is a martingale.*

4. *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion. Then $\limsup_{t\to\infty} B_t = +\infty, \liminf_{t\to-\infty} B_t = -\infty$.*

Proof. Exercise $\square$

In classical analysis we mainly encounter continuously differentiable function which in particular admit a finite *total variation* on compact intervals. This is in stark contrast to the behaviour of typical paths of Brownian motion.

Definition 2.13. *Let $(X_t)_{t\geq 0}$ be a real-valued stochastic process. If there exists a stochastic process $(\langle X\rangle_t)_{t\in[0,1]}$, such that for all $t \in [0,1]$ and all sequences $(\mathcal{D}_n)_{n\in\mathbb{N}}$ of partitions of $[0,t]$, i.e.*

$$\mathcal{D}_n := \{0 = t_0 < t_1 < \cdots < t_n = t\}$$

with $|\mathcal{D}_n| := \mu(\mathcal{D}_n) := \sup_{i\in\{1,\ldots,n\}}|t_i - t_{i-1}| \to 0$ as $n \to \infty$ it holds

$$\langle X\rangle_t = \lim_{n\to\infty} \sum_{t_i\in\mathcal{D}_n} (X_{t_i} - X_{t_{i-1}})^2$$

in probability, then $(X_t)_{t\geq 0}$ is said to have finite quadratic variation. Let $(X_t)_{t\geq 0}$ be a real-valued stochastic process. If there exists a stochastic process $(\langle X\rangle_t)_{t\geq 0}$, such that for all $t \geq 0$ and all sequences $(\mathcal{D}_n)_{n\in\mathbb{N}}$ of partitions of $[0,t]$, i. e.

$$\mathcal{D}_n := \{0 = t_0^n < t_1^n < \cdots < t_{n_k}^n = t\}$$

with $|\mathcal{D}_n| := \sup_{i\in\{1,\ldots,k_n\}}|t_i^n - t_{i-1}^n| \to 0$ as $n \to \infty$ it holds

$$\langle X\rangle_t = \lim_{n\to\infty} \sum_{i=1}^{k_n} (X_{t_i^n} - X_{t_{i-1}^n})^2$$

in probability, then $(X_t)_{t\geq 0}$ is said to have finite quadratic variation.

Theorem 2.14. *A Brownian motion $(B_t)_{t\geq 0}$ has finite quadratic variation and it holds $\langle B\rangle_t = t$ for every $t \geq 0$.*

Proof. Let $t \geq 0$ and $(\mathcal{D}_n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[0,t]$. Then

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{i=1}^{k_n}\left(\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right)\right)^2\right] \\
&= \sum_{i,j=1}^{k_n} \mathbb{E}\left[\left(\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right)\left(\left(B_{t_j^n} - B_{t_{j-1}^n}\right)^2 - (t_j^n - t_{j-1}^n)\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{k_n} \mathbb{E}\left[\left(\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2 - (t_i^n - t_{i-1}^n)\right)^2\right] \\
&= \sum_{i=1}^{k_n} \mathbb{E}\left[\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^4 - (t_i^n - t_{i-1}^n)^2\right] \\
&\leq c \sum_{i=1}^{k_n} (t_i^n - t_{i-1}^n)^2 \\
&\leq c|\mathcal{D}_n| \sum_{i=1}^{k_n} (t_i^n - t_{i-1}^n) \\
&= c|\mathcal{D}_n|t \\
&\to 0
\end{aligned}
$$

as $n \to \infty$. Since L^2-convergence implies convergence in probability, the assertion follows. □

A notable consequence of this result is that for almost all ω the function $t \mapsto B_t(\omega)$ has infinite total variation on every interval of positive length.

Lemma 2.15. *Let B be Brownian motion. Then $M_t := B_t^2 - t$ defines a martingale.*

Proof. Exercise. □

More generally, the quadratic variation process exists for continuous martingales and satisfies a counterpart to Lemma 2.15.

Theorem 2.16 (Quadratic variation)**.** *Let $(M_t)_{t\geq 0}$ be a continuous martingale satisfying $\mathbb{E}M_t^2 < \infty$ for $t \geq 0$. Then there exists a continuous process $(\langle M \rangle_t)_{t\geq 0}$ with the following properties:*

(i) $\langle M \rangle_0 = 0$ *and* $\langle M \rangle$ *is increasing.*

(ii) $M^2 - \langle M \rangle$ *is a martingale.*

(iii) *For all $t \geq 0$ and every sequence $(\mathcal{D}_n)_{n\in\mathbb{N}}$ of subdivisions of $[0, T]$ with $\lim_{n\to\infty} \mu(\mathcal{D}_n) = 0$ we have*

$$\langle M \rangle_t = \lim_{n\to\infty} \sum_{\substack{t_i \in \mathcal{D}_n \\ t_i \leq t}} (M_{t_{i+1}\wedge t} - M_{t_i})^2$$

in L^2.

Proof. Actually the proof is quite technical. Even in a stochastic analysis lecture it would be tempting skip it... □

2.4 The Ito-integral

In this section we discuss the definition of a integrals of the form $\int_0^T H_t\,dX_t$, where H and X are stochastic processes. Such integrals are of fundamental importance not just for stochastic analysis but also for mathematical finance. The reason is that they represent the continuous time counterpart of the process $(H \cdot X)$ encountered in the discrete time setup. We will come back to this finance interpretation at a later stage.

As discussed in the previous section the paths of Brownian motion have infinite total variation on every finite interval. As a consequence we can not form a derivative $\frac{dB_t(\omega)}{dt}$ and likewise it does not work to define an stochastic integral in a 'pathwise sense': Given an arbitrary (say bounded continuous) stochastic process H there is not reason why the Riemann-Stiltjes integral (or the Lebesgue-Stiltjes integral)

$$\int_0^T H_t(\omega)\,dB_t(\omega) \tag{2.4}$$

should exist. Indeed, for this we would need precisely that the function $t \mapsto B_t(\omega)$ has finite total variation on the interval $[0, T]$.

The same problem occurs not only for Brownian motion but virtually for all (continuous) martingales which are not identically 0. A way out of this is to define integrals in slightly different way. It turns out that a natural class of processes for which one can define a useful stochastic integral consists in so called *semi martingales*:

Definition 2.17. *A continuous adapted process X is a semimartingale*[1] *if there exist a continuous martingale M and a continuous process A such that almost surely $\omega \mapsto A_t(\omega)$ has almost surely finite variation on every finite interval with*

$$X = M + A. \tag{2.5}$$

Remark 2.18. 1. *Apparently every martingale is a semi martingale and in particular Brownian motion is a semi martingale.*

2. *The* semi-martingale decomposition *in* (2.5) *is unique if we demand in addition that* $A_0 = 0$.

3. *The celebrated Doob-Meyer Theorem implies that every sub-martingale that is sufficiently bounded is a semi-martingale.*

[1] Actually, this definition is not quite precise. Rigorously the following, wider definition should be used: A continuous process X is a semimartingale if it is *locally* the sum of a martingale and a finite variation process. I.e. if there exists a sequences of stopping times T^n, martingales M^n, and finite variation processes A^n such that $\lim_n T^n = \infty$ a.s. and $X_t(\omega) = M_t^n(\omega) + A_t^n(\omega)$ for $t \leq T^n(\omega)$.

Theorem 2.19. *Assume that X is a semi-martingale and that H is a continuous bounded adapted process.*

There exists a continuous adapted stochastic process $I =: (H \cdot X) =: \int_0^\cdot H_s \, dX_s$, such that for all $t \geq 0$ and all sequences $(\mathcal{D}_n)_{n\in\mathbb{N}}$ of partitions of $[0,t]$, i.e.

$$\mathcal{D}_n := \{0 = t_0^n < t_1^n < \cdots < t_{n_k}^n = t\}$$

with $|\mathcal{D}_n| := \sup_{i\in\{1,\ldots,k_n\}} |t_i^n - t_{i-1}^n| \to 0$ as $n \to \infty$ it holds

$$I_t = \lim_{n\to\infty} \sum_{i=1}^{k_n} H_{t_i^n}(X_{t_i^n} - X_{t_{i-1}^n})$$

in probability.

Definition 2.20. *In the context of Theorem 2.19, the process given by $I_t = (H \cdot X)_t =: \int_0^t H_s \, dX_s$ is called the* Ito-integral *or* stochastic integral *of H wrt. X.*

It is straightforward to see that the Ito-integral is linear in both the integrand (H in the above definition) as well as the integrator (X in the above definition).

Note also that if $t \mapsto A_t(\omega)$ has almost surely finite variation and is continuous in t, then the Ito-integral $I_T = \int_0^T H_s \, dA_s$ satisfies

$$I_T(\omega) = \int_0^T H_t(\omega) \, dA_t(\omega), \tag{2.6}$$

where the right hand side of (2.6) denotes the usual Riemann-Stieltjes integral of the function $t \mapsto H_t(\omega)$ wrt. the continuous, finite variation function $t \mapsto A_t(\omega)$.

Next we turn to the definition of stochastic integrals for so called simple integrands:

Definition 2.21. *A process $H = (H_t)_{t\in[0,T]}$ is called simple if*

$$H_t(\omega) = \sum_{i=1}^n H^i I_{(s_i, s_{i+1}]},$$

where $0 \leq s_1 \leq \ldots \leq s_n$ and each H^i is $\mathcal{F}_{s_i}$-measurable and bounded. Let X be a continuous adapted process.

For simple integrands we set

$$(H \cdot X)_u := \int_0^u H_s \, dX_s := I_u := \int_0^u H_t \, dX_t := \sum_{i=1}^n H^i (X_{s_{i+1}\wedge u} - X_{s_i \wedge u}).$$

The Ito-integral for simple integrands has a clear interpretation in mathematical finance terms: as in the discrete time setup, X_t stands for the price of a financial asset at time t, while H_t denotes the number of shares we hold at time t. Then $(H \cdot X)_T$ represents exactly gains/losses accumulated from trading until time T.

The definition of the stochastic integral for simple processes connects to our previous definition through appropriate convergence results for stochastic integrals. Here we will only mention a version of the *dominated convergence Theorem of stochastic integration*.

To state it, we need a definition.

Definition 2.22. *A process L is called* locally bounded *if there exists a sequence of stopping times $\tau_n,\ n \geq 1, \lim_{n\to\infty} \tau_n = \infty$ such that $(t,\omega) \mapsto L_t^{\tau_n}(\omega) := L_{t\wedge\tau_n(\omega)}(\omega)$ is bounded for every n.*

Clearly every bounded process is locally bounded. We also have:

Lemma 2.23. *Let L be a continuous process such that L_0 is bounded. Then L is locally bounded.*

Sketch of proof. The idea is to use that the sequence of stopping times given by

$$\tau_n := \inf\{t \geq 0 : |L_t| \geq n\}.$$

□

Using this notion we can formulate a version of the dominated convergence Theorem for stochastic integrals:

Theorem 2.24. *Let $H, H_n, n \geq 1$ be adapted processes, which are simple or continuous. Assume that L is a locally bounded process such that $|H|, |H_n| \leq L$ and that X is a continuous semi martingale. Then*

$$\lim_{n\to\infty} (H_n \cdot X)_T = (H \cdot X)_T$$

in probability.

Proof. The proof of this result is beyond the scope of our lecture. □

In view of Theorem 2.24 it is natural to try to extend the Ito-integral to all locally bounded integrands that can be approximated by simple functions.

Definition 2.25. *A process that is the pointwise limit of simple functions is called* predictable.

Lemma 2.26. *Every (left-)continuous adapted process is predictable.*

Proof. Exercise. □

If H is a predictable process which is locally bounded, the stochastic integral $\int_0^T H_s\, dX_s$ can be defined by approximating with H with simple functions. Theorem 2.24 then applies also in the case where $H, H_n, n \geq 1$ are only predictable but not necessarily continuous or simple.

2.5 Alternative construction of the Ito-integral for Brownian motion

In this section we will sketch an alternative definition of the stochastic integral

$$\int H_t \,\mathrm{d}B_t,$$

i.e. of the particular case where the integrator is Brownian motion. We fix $T > 0$ and concentrate on the interval $[0, T]$.

Definition 2.27.

(i) A function $H \colon \Omega \times [0, T] \to \mathbb{R}$ is measurable, if it is $\mathcal{F} \otimes \mathcal{B}([0, T])$-measurable. It is adapted, if $H_t(\cdot) = H(\cdot, t)$ is $\mathcal{F}_t$-measurable for every $t \in [0, T]$.

(ii) Let

$$\begin{aligned} \mathcal{H}^2 &:= \mathcal{H}^2([0, T]) \\ &:= \left\{ H \colon \Omega \times [0, T] \to \mathbb{R} \text{ measurable, adapted, } \mathbb{E}\left[\int_0^T H_s^2(\cdot)\,\mathrm{d}s\right] < \infty \right\}. \end{aligned}$$

(iii) Let

$$\begin{aligned} \mathcal{H}_0^2 := \Bigg\{ & H = \sum_{i=0}^{n-1} a_i \mathbb{1}_{(t_i, t_{i+1}]} : 0 = t_0 < t_1 < \cdots < t_n = T, a_i \in L^2(\Omega, \mathcal{F}_{t_i}, \mathbb{P}), \\ & \text{for every } i \in \{0, \ldots, n-1\}, n \in \mathbb{N} \Bigg\}. \end{aligned}$$

(iv) For $H = \sum_{i=0}^{n-1} a_i \mathbb{1}_{(t_i, t_{i+1}]} \in \mathcal{H}_0^2$ let

$$(H \cdot B)_T := I(H) := \int_0^T H(\cdot, s)\,\mathrm{d}B_s := \sum_{i=0}^{n-1} a_i (B_{t_{i+1}} - B_{t_i}).$$

Lemma 2.28 (Ito's isometry)**.** *For $H \in \mathcal{H}_0^2$ we have*

$$\|I(H)\|_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = \|H\|_{L^2(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}([0,T]), \mathbb{P} \otimes \lambda)}.$$

Proof. Let $H = \sum_{i=0}^{n-1} a_i \mathbb{1}_{(t_i, t_{i+1}]} \in \mathcal{H}_0^2$. With Fubini's theorem we obtain

$$\|H\|_2^2 = \sum_{i=0}^{n-1} \mathbb{E}[a_i^2](t_{i+1} - t_i).$$

Moreover,

$$\begin{aligned}
\|I(H)\|_2^2 &= \mathbb{E}[I(H)^2] \\
&= \sum_{i,j=0}^{n-1} \mathbb{E}\left[a_i a_j (B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})\right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}\left[\mathbb{E}\left[a_i^2 (B_{t_{i+1}} - B_{t_i})^2 \middle| \mathcal{F}_{t_i}\right]\right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}[a_i^2]\mathbb{E}\left[(B_{t_{i+1}} - B_{t_i})^2\right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}[a_i^2](t_{i+1} - t_i),
\end{aligned}$$

where we have used the independence of the increments of B. □

The following lemma is need to prove Proposition 2.30 below in which we obtain that the $\mathcal{H}_0^2$ is dense in $\mathcal{H}^2$ with respect to the L^2-norm. Both the proof of the lemma and the proposition are quite technical. I put the proofs here for completeness but *strongly suggest* that you omit them.

Lemma 2.29. *Let $H\colon \Omega \otimes [0,T] \to \mathbb{R}$ be measurable and bounded. Then*

$$\lim_{h \downarrow 0} \mathbb{E}\left[\int_0^T \left|H(\cdot,t) - H(\cdot,(t-h)^+)\right|^2 \,\mathrm{d}t\right] = 0.$$

Proof.

(i) For $t \in [0,T]$ and $n \in \mathbb{N}$ let

$$\begin{aligned}
g(\cdot,t) &:= \int_0^t H(\cdot,s)\,\mathrm{d}s, \\
H_n(\cdot,t) &:= n\left(g(\cdot,t) - g\left(\cdot,\left(t-\tfrac{1}{n}\right)^+\right)\right)
\end{aligned}$$

so that $|H_n(\cdot,t)| \le n\int_{(t-\frac{1}{n})^+}^t |H(\cdot,s)|\,\mathrm{d}s \le \|H\|_\infty$. Let

$$\begin{aligned}
A &:= \left\{(\omega,t) \in \Omega \times [0,T] : \lim_{n\to\infty} H_n(\omega,t) \ne H(\omega,t)\right\}, \\
A_\omega &:= \{t \in [0,T] : (\omega,t) \in A\}
\end{aligned}$$

for $\omega \in \Omega$. By the theorem of Lebesgue we have $\lambda(A_\omega) = 0$ for every $\omega \in \Omega$, hence $\mathbb{P} \otimes \lambda(A) = 0$. Therefore, by dominated convergence

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T |H_n(\cdot,t) - H(\cdot,t)|^2 \,\mathrm{d}t\right] = 0.$$

(ii) For $\varepsilon > 0$ we can find $n \in \mathbb{N}$ and $h_0 > 0$ such that

$$\mathbb{E}\left[\int_0^T |H_n(\cdot,t) - H(\cdot,t)|^2 \,\mathrm{d}t\right] < \varepsilon$$

and

$$\mathbb{E}\left[\int_0^T \left|H_n(\cdot,(t-h)^+) - H(\cdot,(t-h)^+\right|^2 \,\mathrm{d}t\right] \leq \varepsilon$$

for all $0 < h < h_0$. Using the triangle inequality we obtain

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \left|H(\cdot,t) - H(\cdot,(t-h)^+\right|^2 \,\mathrm{d}t\right]^{\frac{1}{2}} \\
&\leq \mathbb{E}\left[\int_0^T |H(\cdot,t) - H_n(\cdot,t)|^2 \,\mathrm{d}t\right]^{\frac{1}{2}} + \mathbb{E}\left[\int_0^T \left|H_n(\cdot,t) - H_n(\cdot,(t-h)^+)\right|^2 \,\mathrm{d}t\right]^{\frac{1}{2}} \\
&\quad + \mathbb{E}\left[\int_0^T \left|H_n(\cdot,(t-h)^+) - H(\cdot,(t-h)^+)\right|^2 \,\mathrm{d}t\right]^{\frac{1}{2}} \\
&\leq 2\varepsilon + \mathbb{E}\left[\int_0^T \left|H_n(\cdot,t) - H_n(\cdot,(t-h)^+)\right|^2 \,\mathrm{d}t\right]^{\frac{1}{2}}.
\end{aligned}$$

We let $h \downarrow 0$ and conclude from the continuity of $t \mapsto H_n(\cdot,t)$ and the dominated convergence theorem that

$$\lim_{h\downarrow 0} \mathbb{E}\left[\int_0^T \left|H_n(\cdot,t) - H_n(\cdot,(t-h)^+)\right|^2 \,\mathrm{d}t\right]^{\frac{1}{2}} = 0.$$

□

Proposition 2.30. *The set $\mathcal{H}_0^2$ is dense in $\mathcal{H}^2$ with respect to the L^2-norm.*

Proof.

(i) Let $H \in \mathcal{H}^2$. For $n \in \mathbb{N}$ define $H_n := (-n) \vee (H \wedge n)$ which is bounded and adapted. By dominated convergence $\|H_n - H\|_2 \to 0$ as $n \to \infty$.

(ii) By (i) we can assume that H is measurable, bounded and adapted. We have to show that there exists $(H_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}_0^2$ with $\|H_n - H\|_2 \to 0$. To that end, for all $n \in \mathbb{N}$ and $t \in \mathbb{R}$ define

$$\varphi_n(t) := \sum_{j\in\mathbb{Z}} \frac{j-1}{2^n} \mathbb{1}_{\left(\frac{j-1}{2^n}, \frac{j}{2^n}\right]}(t).$$

Since $t - \frac{1}{2^n} \leq \varphi_n(t-s) + s < t$ and φ_n only takes discrete values, the process

$$H_{n,s}(\cdot,t) := H(\cdot,(s + \varphi_n(t-s))^+)$$

is adapted and a member of $\mathcal{H}_0^2$. Moreover, we claim that

$$\mathbb{E}\left[\int_0^T \int_0^1 |H_{n,s}(\cdot,t) - H(\cdot,t)|^2 \,\mathrm{d}s\,\mathrm{d}t\right] \to 0 \tag{2.7}$$

as $n \to \infty$. Indeed, for $n \in \mathbb{N}$ we have

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \int_0^1 |H_{n,s}(\cdot,t) - f(\cdot,t)|^2 \,\mathrm{d}s\,\mathrm{d}t\right] \\
&= \mathbb{E}\left[\int_0^T \int_0^1 \left|H\left(\cdot,(s+\varphi_n(t-s))^+\right) - H(\cdot,t)\right|^2 \,\mathrm{d}s\,\mathrm{d}t\right] \\
&= \sum_{j\in\mathbb{Z}} \mathbb{E}\left[\int_0^T \int_{\left[t-\frac{j}{2^n},t-\frac{j-1}{2^n}\right)\cap[0,1]} \left|H\left(\cdot,\left(s+\tfrac{j-1}{2^n}\right)^+\right) - H(\cdot,t)\right|^2 \,\mathrm{d}s\,\mathrm{d}t\right] \\
&\leq 2^n(T+1)\mathbb{E}\left[\int_0^T \int_0^{2^{-n}} \left|H\left(\cdot,(t-h)^+\right) - H(\cdot,t)\right|^2 \,\mathrm{d}h\,\mathrm{d}t\right] \\
&= 2^n(T+1)\int_0^{2^{-n}} \mathbb{E}\left[\int_0^T \left|H\left(\cdot,(t-h)^+\right) - H(\cdot,t)\right|^2 \,\mathrm{d}t\right] \mathrm{d}h \\
&\to 0
\end{aligned}$$

as $n \to \infty$ by Lemma 2.29. This shows (2.7).

Finally according to (2.7), there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $(\omega,t,s) \mapsto H_{n_k,s}(\omega,t)$ converges to f $\mathbb{P}\otimes\lambda\otimes\lambda$-as. Fubini's theorem implies $H_{n_k,s} \to H$ $\mathbb{P}\otimes\lambda$-as for λ-almost all $s \in [0,1]$. Hence, we may choose $s \in [0,1]$ such that $H_{n_k,s} \to H$ $\mathbb{P}\otimes\lambda$-as. By dominated convergence we get

$$\mathbb{E}\left[\int_0^T |H_{n_k,s}(\cdot,t) - H(\cdot,t)|^2 \,\mathrm{d}t\right] \to 0$$

as $k \to \infty$. Since $(H_{n_k,s})_{k\in\mathbb{N}}$ is a sequence in $\mathcal{H}_0^2$, the proof is complete.

□

Thus we can extend the stochastic integral from $\mathcal{H}_0^2$ to $\mathcal{H}^2$. Indeed, for $H \in \mathcal{H}^2$ let $(H_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}_0^2$ be a sequence such that $\|H_n - H\|_2 \to 0$. In particular, $(H_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}^2$. By Lemma 2.28 we get

$$\|I(H_m) - I(H_n)\|_2 = \|I(H_m - H_n)\|_2 = \|H_m - H_n\|_2$$

showing that $(I(H_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\Omega,\mathcal{F},\mathbb{P})$. Since $L^2(\Omega,\mathcal{F},\mathbb{P})$ is complete, we can define the stochastic integral

$$I(H) := \lim_{n\to\infty} I(H_n) \in L^2(\Omega,\mathcal{F},\mathbb{P}).$$

Remark 2.31. *The stochastic integral I is well-defined. Suppose there exist two sequences $(H_n^1)_{n\in\mathbb{N}}$ and $(H_n^2)_{n\in\mathbb{N}}$ converging to $f \in L^2$ in L^2. Then, by Lemma 2.28 we have*

$$\left\|I(H_n^1) - I(H_n^2)\right\|_2 = \left\|H_n^1 - H_n^2\right\|_2 \leq \left\|H_n^1 - f\right\|_2 + \left\|H - H_n^2\right\|_2 \to 0$$

as $n \to \infty$.

Theorem 2.32 (Ito's isometry)**.** *For $H \in \mathcal{H}^2$ we have $\|H\|_2 = \|I(H)\|_2$.*

Proof. Let $(H_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}_0^2$ such that $\|H_n - H\|_2 \to 0$. By definition of $I(H)$, we have $\|I(H_n) - I(H)\|_2 \to 0$, so that Lemma 2.28 yields

$$\|H\|_2 = \lim_{n\to\infty} \|H_n\|_2 = \lim_{n\to\infty} \|I(H_n)\|_2 = \|I(H)\|_2 .$$

□

2.6 Ito's formula

In this section we discuss the "chain rule of stochastic calculus", the celebrated Ito-formula.

For reference, let's recall the usual chain rule:

$$(f \circ g)'(t) = f'(g)(t) \cdot g'(t) \quad \Longleftrightarrow \quad \frac{df(g)}{dt} = \frac{df}{dg}\frac{dg}{dt} = f'(g)\frac{dg}{dt}. \tag{2.8}$$

Our goal is to derive a similar formula for the case where g is replaced by Brownian motion.

As an intermediate step, we consider the case where g is not necessarily integrable but still has finite variation. Formally we can multiply the right hand side of (2.8) with dt to obtain

$$df(g) = f'(g)dg. \tag{2.9}$$

To give (2.9) a rigorous meaning we need to write it in integrated form:

$$f(g(T)) - f(g(0)) = \int_0^T 1\, df(g) = \int_0^T f'(g(t))\, dg(t). \tag{2.10}$$

The validity of (2.10) is a basic result for the Riemann-Stieltjes integral. We can ask ourselves whether (2.10) remains true when we replace the finite variation function $t \mapsto g(t)$ with Brownian motion $t \mapsto B_t$. Remarkably, this is not the case, rather there is an additional correction term:

Theorem 2.33 (Ito's formula)**.** *Let $f \in C^2(\mathbb{R})$.*[2] *Then*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s$$

for every $t \in [0, T]$.

[2]Let $C^2(\mathbb{R})$ denote the space of twice differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$ with continuous second derivative.

Usually Ito's formula is given in a shorthand version similar to (2.9). It then reads

$$df(B_t) = f'(B_t)\, dB_t + \frac{1}{2} f''(B_t) dt. \tag{2.11}$$

We will first give a heuristic derivation of (2.11). The key idea is that, informally, Theorem 2.14 on the quadratic variation of Brownian motion asserts that

$$dB_t^2 = dt \tag{2.12}$$

. To estimate $df(B_t)$ we can then use Taylor's formula to obtain

$$\begin{aligned} df(B_t) &= f(B_{t+dt}) - f(B_t) = f((B_t) + dB_t) - f(B_t) && (2.13)\\ &= f'(B_t)dB_t + \frac{1}{2} f''(B_t)dB_t^2 + \frac{1}{6} f'''(B_t)dB_t^3 + \dots && (2.14)\\ &= f'(B_t)dB_t + \frac{1}{2} f''(B_t)dB_t^2. && (2.15) \end{aligned}$$

Note that all terms including $dB_t^n, n \geq 3$ are of higher order than dt an thus do not contribute to sum. However the term dB_t^2 is of the same order as dt (in fact equal to it) and thus leads to an important contribution. This is the decisive difference to the case of finite variation functions.

In the remainder of this section, we give a rigorous proof of Ito's formula. I think that it doesn't add that much to the heuristic discussion above, feel free to skip it. The proof of Ito's formula is based on the following result.

Theorem 2.34. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be continuous, $t_i := \frac{i}{n} T$ for $i \in \{0, \dots, n\}$. Then*

$$\sum_{i=1}^{n} f(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) \to \int_0^T f(B_s)\, \mathrm{d}B_s$$

in probability as $n \to \infty$.

Proof.

(i) For $m \in \mathbb{N}$ let

$$\tau_m := \inf\{t \geq 0 : |B_t| \geq m\} \wedge T.$$

Then $(\tau_m)_{m \in \mathbb{N}}$ is a localizing sequence for $f(B)$. Further, there exists a continuous function $f_m\colon \mathbb{R} \to \mathbb{R}$ with compact support and $f|_{[-m,m]} = f_m|_{[-m,m]}$. By construction

$$\int_0^{\cdot} f(B_s)\, \mathrm{d}B_s = \int_0^{\cdot} f_m(B_s)\, \mathrm{d}B_s$$

on $[0, \tau_m]$.

(ii) Let $m \in \mathbb{N}$ be fixed. For $n \in \mathbb{N}$ let

$$\phi_n(\omega, s) := \sum_{i=1}^{n} f_m(B_{t_{i-1}}(\omega)) \mathbb{1}_{(t_{i-1}, t_i]}(s).$$

By construction, $\phi_n \in \mathcal{H}_0^2$. We claim $\phi_n \to f_m(B)$ in $L^2(\Omega \times [0,T])$. Indeed,

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T |\phi_n(\cdot, s) - f_m(B_s)|^2 \, \mathrm{d}s\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left|f_m(B_{t_{i-1}}) - f_m(B_s)\right|^2 \, \mathrm{d}s\right] \\
&\leq \frac{T}{n} \sum_{i=1}^{n} \mathbb{E}\left[\sup_{s \in (t_{i-1}, t_i]} \left|f_m(B_{t_{i-1}}) - f_m(B_s)\right|^2\right].
\end{aligned}$$

Since f_m is uniformly continuous, we have

$$\mu(h) := \sup\{|f_m(x) - f_m(y)| : |x - y| \leq h\} \to 0$$

for $h \to 0$. For $i \in \{1, \ldots, n\}$ let $M_i := \sup_{s \in (t_{i-1}, t_i]} \left|B_s - B_{t_{i-1}}\right|$. Then

$$\begin{aligned}
\mathbb{E}\left[\sup_{s \in (t_{i-1}, t_i]} \left|f_m(B_{t_{i-t}}) - f_m(B_s)\right|^2\right] &\leq \mathbb{E}\left[\mu(M_i)^2\right] \\
&\leq \mathbb{E}\left[\mu\left(\sup_{i \in \{1, \ldots, n\}} M_i\right)^2\right] \\
&\to 0
\end{aligned}$$

as $n \to \infty$, since B is pathwise uniformly continuous, and μ is bounded so that we can apply the dominated convergence theorem. Hence,

$$\mathbb{E}\left[\int_0^T |\phi_n(\cdot, s) - f_m(B_s)|^2 \, \mathrm{d}s\right] \leq T\mathbb{E}\left[\mu\left(\sup_{i \in \{1, \ldots, n\}} M_i\right)^2\right] \to 0$$

as $n \to \infty$. This implies the claim for f_m, that is

$$\begin{aligned}
\int_0^T f_m(B_s) \, \mathrm{d}B_s &= \lim_{n \to \infty} \int_0^T \phi_n(\cdot, s) \, \mathrm{d}B_s \\
&= \lim_{n \to \infty} \sum_{i=1}^{n} f_m(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})
\end{aligned}$$

in L^2.

(iii) For $\varepsilon > 0$ let

$$A_n(\varepsilon) := \left\{ \left| \sum_{i=1}^n f(B_{t_{i-1}})\big(B_{t_i} - B_{t_{i-1}}\big) - \int_0^T f(B_s)\,\mathrm{d}B_s \right| > \varepsilon \right\}.$$

We have to show $\lim_{n\to\infty} \mathbb{P}(A_n(\varepsilon)) = 0$. For $m \in \mathbb{N}$ we have

$$\int_0^t f(B_s)\,\mathrm{d}B_s = \int_0^t f_m(B_s)\,\mathrm{d}B_s$$

on $\{\tau_m = T\}$. Hence

$$\begin{aligned}\mathbb{P}(A_n(\varepsilon)) &= \mathbb{P}(A_n(\varepsilon) \cap \{\tau_m < T\}) + \mathbb{P}(A_n(\varepsilon) \cap \{\tau_m = T\}) \\ &\le \mathbb{P}(\tau_m < T) + \mathbb{P}(A_n(\varepsilon) \cap \{\tau_m = T\}) \\ &\to 0\end{aligned}$$

as $n, m \to \infty$ by (ii). This yields

$$\sum_{i=1}^n f(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) \xrightarrow{\mathbb{P}} \int_0^T f(B_s)\,\mathrm{d}B_s$$

as $n \to \infty$.

□

Proof of Theorem 2.33.

(i) Assume f to posses compact support. Fix $t \in [0, T]$ and define $t_i := \frac{i}{n}t$ for $i \in \{1, \dots, n\}$. Using Taylor's theorem we obtain

$$\begin{aligned}&f(B_t) - f(0) \\ &= \sum_{i=1}^n (f(B_{t_i}) - f(B_{t_{i-1}})) \\ &= \sum_{i=1}^n f'(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) + \frac{1}{2}\sum_{i=1}^n f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 \\ &\qquad + \sum_{i=1}^n h(B_t, B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 \\ &=: A_n + B_n + C_n,\end{aligned}$$

where $x \mapsto h(x, y)$ is continuous and $\lim_{x\to y} h(x, y) = h(y, y) = 0$. We will show that A_n and B_n converge to the respective integrals and the remainder term C_n converges to zero as $n \to \infty$.

(ii) Since f' is continuous it follows from Theorem 2.34 that $A_n \xrightarrow{\mathbb{P}} \int_0^t f'(B_s)\,\mathrm{d}B_s$ as $n \to \infty$.

(iii) We claim that for f with compact support we have $B_n \xrightarrow{\mathbb{P}} \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s$. Indeed,

$$\begin{aligned} B_n &= \frac{1}{2}\sum_{i=1}^n f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 \\ &= \frac{1}{2}\sum_{i=1}^n f''(B_{t_{i-1}})(t_i - t_{i-1}) + \frac{1}{2}\sum_{i=1}^n f''(B_{t_{i-1}})\left((B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\right) \\ &=: B_n' + B_n''. \end{aligned}$$

Since f' is continuous, we have $B_n' \to \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s$ $\mathbb{P}$-as. Moreover, it holds

$$\begin{aligned} \mathbb{E}\left[(B_n'')^2\right] &= \frac{1}{4}\sum_{i,j=1}^n \mathbb{E}\left[f''(B_{t_{i-1}})\left((B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\right)\right. \\ &\qquad \left.\cdot f''(B_{t_{j-1}})\left((B_{t_j} - B_{t_{j-1}})^2 - (t_j - t_{j-1})\right)\right] \\ &= \frac{1}{4}\sum_{i=1}^n \mathbb{E}\left[f''(B_{t_{i-1}})^2\left((B_{t_i} - B_{t_{i-1}})^2 - (t_i - t_{i-1})\right)^2\right] \\ &\le \frac{1}{4}\left\|f''\right\|_\infty^2 \sum_{i=1}^n \mathbb{V}\mathrm{ar}\left[(B_{t_i} - B_{t_{i-1}})^2\right] \\ &= \frac{1}{4}\left\|f''\right\|_\infty^2 \mathbb{V}\mathrm{ar}[X^2]\sum_{i=1}^n (t_i - t_{i-1})^2 \\ &\to 0, \end{aligned}$$

where $X \sim \mathcal{N}(0,1)$, since

$$\sum_{i=1}^n (t_i - t_{i-1})^2 \le \sup_{j\in\{1,\ldots,n\}}(t_j - t_{j-1})\sum_{i=1}^n (t_i - t_{i-1}) \to 0$$

as $n \to \infty$. Hence $B_n'' \to 0$ in L^2 and therefore in probability.

(iv) We claim that for f with compact support we have $C_n \xrightarrow{\mathbb{P}} 0$. In fact with Hölder's inequality

$$\begin{aligned} \mathbb{E}[|C_n|] &\le \mathbb{E}\left[\sum_{i=1}^n \left|h(B_{t_i}, B_{t_{i-1}})\right|(B_{t_i} - B_{t_{i-1}})^2\right] \\ &\le \sum_{i=1}^n \mathbb{E}\left[h(B_{t_i}, B_{t_{i-1}})^2\right]^{\frac{1}{2}} \mathbb{E}\left[(B_{t_i} - B_{t_{i-1}})^4\right]^{\frac{1}{2}} \\ &= \sum_{i=1}^n (t_i - t_{i-1})\mathbb{E}[X^4]^{\frac{1}{2}}\mathbb{E}\left[h(B_{t_i}, B_{t_{i-1}})^2\right]^{\frac{1}{2}}, \end{aligned}$$

where $X \sim \mathcal{N}(0,1)$. By continuity of $x \mapsto h(x,y)$ and $\lim_{x \to y} h(x,y) = 0$, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|x-y| \leq \delta$ implies $h(x,y) \leq \varepsilon$. Therefore

$$\begin{aligned}
\mathbb{E}\left[h(B_{t_i}, B_{t_{i-1}})^2\right] &= \mathbb{E}\left[h(B_{t_i}, B_{t_{i-1}})^2 \mathbb{1}_{\{|B_{t_i} - B_{t_{i-1}}| \leq \delta\}}\right] \\
&\quad + \mathbb{E}\left[h^2(B_{t_i}, B_{t_{i-1}}) \mathbb{1}_{\{|B_{t_i} - B_{t_{i-1}}| > \delta\}}\right] \\
&\leq \varepsilon^2 + \|h\|_\infty^2 \, \mathbb{P}\left(\left|B_{t_i} - B_{t_{i-1}}\right| > \delta\right) \\
&\leq \varepsilon^2 + \|h\|_\infty^2 \, \frac{t}{n\delta^2}.
\end{aligned}$$

Consequently,

$$\mathbb{E}[|C_n|] \leq \sum_{i=1}^n (t_i - t_{i-1}) \mathbb{E}[X^4] \frac{1}{2} \left(\varepsilon^2 + \|h\|_\infty^2 \tfrac{t}{n\delta^2}\right)^{\frac{1}{2}} \to t\mathbb{E}[X^4]^{\frac{1}{2}} \varepsilon^2$$

as $n \to \infty$. Since ε was arbitrary, we conclude $C_n \to 0$ in L^1 and therefore also $C_n \xrightarrow{\mathbb{P}} 0$. This shows Theorem 2.33 for $f \in C^2(\mathbb{R})$ with compact support.

(v) For $m \in \mathbb{N}$ define

$$\tau_m := \inf\{t \geq 0 : |B_t| \geq m\} \wedge T.$$

Then $(\tau_m)_{m \in \mathbb{N}}$ is a localizing sequence of stopping times. For $m \in \mathbb{N}$ there exists $f_m \in C^2(\mathbb{R})$ with compact support and $f|_{[-m,m]} = f_m|_{[-m,m]}$. By the steps (i)-(iv) Ito's formula holds on $[0, \tau_m]$. Letting $\tau_m \uparrow T$ yields the claim.

$\square$

Example 2.35. *Let $f(x) = \frac{1}{2}x^2$ for $x \in \mathbb{R}$. Then*

$$\frac{1}{2}B_t^2 = f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s = \int_0^t B_s\,\mathrm{d}B_s + \frac{1}{2}t.$$

This shows that $\frac{1}{2}(B_t^2 - t) = \int_0^t B_s\,\mathrm{d}B_s$ is a martingale.

2.7 Generalizations

Let $f \in C^{1,2}([0,\infty) \times \mathbb{R})$. By convention we write $f_t = \frac{\partial f}{\partial t}$, $f_x = \frac{\partial f}{\partial x}$ and $f_{xx} = \frac{\partial^2 f}{\partial x x}$.

Theorem 2.36. *Let $f \in C^{1,2}([0,\infty) \times \mathbb{R})$. Then*

$$f(t, B_t) = f(0,0) + \int_0^t f_t(s, B_s) + \frac{1}{2} f_{xx}(s, B_s)\,\mathrm{d}s + \int_0^t f_x(s, B_s)\,\mathrm{d}B_s$$

for every $t \in [0, \infty)$.

In short we would express this as

$$df(t,B_t) = \left[f_t(t,B_t) + \frac{1}{2}f_{xx}(t,B_t)\right]\,dt + f_x(t,B_t)\,dB_t. \tag{2.16}$$

As above, it is not hard to argue this on an intuitive level using Taylor's formula. We have:

$$\begin{aligned}df(t,B_t) &= f(t+dt,B_{t+dt}) - f(B_t) = \\ &= f_t(t,B_t)dt + f_{xx}(t,B_t)dB_t + \frac{1}{2}f_{xx}(t,B_t)dB_t^2 + f_{tx}(t,B_t)dtdB_t + \frac{1}{6}f_{txx}(t,B_t)dB_t^2 + \dots \\ &= f_t(t,B_t)dt + f_x(t,B_t)dB_t + \frac{1}{2}f_{xx}(t,B_t)dB_t^2.\end{aligned}$$

As before we use here that all terms of order $dtdB_t, dB_t^3$ and higher do not contribute at the level of dt. We thus obtain (2.16) based on $dB_t^2 = dt$.

We do not discuss a rigorous version of this argument which in any case would be very similar to the one in the previous section.

The Ito-formula in (2.16) is again just a very special case of a much more general Ito formula for semimartingales.

To state this general version of the Ito-formula we need the concept of quadratic co-variation:

Definition 2.37. *Let $(X_t)_{t\geq 0}, (Y_t)_{t\geq 0}$ be a real-valued stochastic processes. A stochastic process $(\langle X,Y\rangle_t)_{t\in[0,1]}$, such that for all $t \in [0,1]$ and all sequences $(\mathcal{D}_n)_{n\in\mathbb{N}}$ of partitions of $[0,t]$ with $|\mathcal{D}_n| := \mu(\mathcal{D}_n) := \sup_{i\in\{1,\dots,n\}}|t_i - t_{i-1}| \to 0$ as $n\to\infty$ it holds*

$$\langle X,Y\rangle_t = \lim_{n\to\infty}\sum_{t_i\in\mathcal{D}_n}(X_{t_i} - X_{t_{i-1}})(Y_{t_i} - Y_{t_{i-1}})$$

in probability, is called the quadratic co-variation process of X,Y.

If X,Y are continuous semi-martingales, $\langle X,Y\rangle$ exists. Apparently we have $\langle X,X\rangle = \langle X\rangle$. It is also easy to see that

$$\langle X,Y\rangle = \frac{1}{4}(\langle X+Y\rangle - \langle X-Y\rangle). \tag{2.17}$$

Often the quadratic covariation process is defined simply through (2.17). (A definition of this type is called 'definition by polarization'.)

Using this concept we have the following version of Ito's formula: Assume that $f:\mathbb{R}^2\to\mathbb{R}$ is C^2 and that X,Y are continuous semi-martingales. Then we have

$$\begin{aligned}df(X_t,Y_t) = f_x(X_t,Y_t)dX_t + f_y(X_t,Y_t)dY_t + \\ 1/2 f_{xx}(X_t,Y_t)d\langle X\rangle_t + f_{xy}(X_t,Y_t)d\langle X,Y\rangle_t + 1/2 f_{yy}(X_t,Y_t)d\langle Y\rangle_t.\end{aligned} \tag{2.18}$$

Naturally, an analogue of (2.18) for functions in more than two variables is valid as well. We omit it due to the more complicated notations. Likewise we omit the integral form of (2.18) and its proof.

In the next section we will describe a class of semi-martingales for which we can calculate the quadratic covariation process explicitly.

2.8 Ito-processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $B = (B_t)_{t\in[0,T]}$ a Brownian motion with standard filtration $(\mathcal{F}_t)_{t\in[0,T]}$.

Definition 2.38. *A continuous stochastic process $X = (X_t)_{t\in[0,T]}$ is called Ito process, if*

$$X_t = X_0 + \int_0^t a(\cdot, s)\,\mathrm{d}s + \int_0^t b(\cdot, s)\,\mathrm{d}B_s,$$

where $X_0 \in \mathbb{R}$, $a, b\colon \Omega \times [0,T] \to \mathbb{R}$ are $\mathcal{F} \otimes \mathcal{B}([0,T])$-measurable and adapted such that $\int_0^T |a(\cdot, s)|\,\mathrm{d}s < \infty$ and $\int_0^T |b(\cdot, s)|^2\,\mathrm{d}s < \infty$ $\mathbb{P}$-as.

Ito-processes are a relatively tractable class of stochastic processes that is sufficiently general to cover many important applications. In this section we collect basic results concerning Ito-processes as integrators and the quadratic variation of Ito-processes. We will omit the respective proofs but emphasize that they usually follow a rather basic scheme: First one proves the results for the case where the 'coefficients' a, b are simple processes which is fairly straightforward. Then one establishes the general case through limiting arguments.

Proposition 2.39. *Let $X_t = X_0 + \int_0^t a(\cdot, s)\,\mathrm{d}s + \int_0^t b(\cdot, s)\,\mathrm{d}B_s$ be an Ito process. Then for $f\colon \Omega \times [0,T] \to \mathbb{R}$ measurable and adapted with $\int_0^T |f(\cdot, s)a(\cdot, s)|\,\mathrm{d}s < \infty$ and $\int_0^T |f(\cdot, s)b(\cdot, s)|^2\,\mathrm{d}s < \infty$ $\mathbb{P}$-as the Ito integral is given by*

$$\int_0^t f(\cdot, s)\,\mathrm{d}X_s = \int_0^t f(\cdot, s)a(\cdot, s)\,\mathrm{d}s + \int_0^t f(\cdot, s)b(\cdot, s)\,\mathrm{d}B_s.$$

Proposition 2.40. *Let X be an Ito process with representation*

$$\begin{aligned} X &= X_0^1 + \int_0^\cdot a_1(\cdot, s)\,\mathrm{d}s + \int_0^\cdot b_1(\cdot, s)\,\mathrm{d}B_s \\ &= X_0^2 + \int_0^\cdot a_2(\cdot, s)\,\mathrm{d}s + \int_0^\cdot b_2(\cdot, s)\,\mathrm{d}B_s. \end{aligned}$$

Then $X_0^1 = X_0^2$, $a_1 = a_2$ and $b_1 = b_2$ $\mathbb{P} \otimes \lambda$-as.

Theorem 2.41 (Quadratic variation of an Ito process)**.** *Let*

$$X = \int_0^\cdot a(\cdot, s)\,\mathrm{d}s + \int_0^\cdot b(\cdot, s)\,\mathrm{d}B_s$$

be an Ito process. Then

$$\langle X \rangle_t = \int_0^t b^2(\cdot, s)\,\mathrm{d}s.$$

Let

$$X^i = \int_0^\cdot a^i(\cdot, s)\,\mathrm{d}s + \int_0^\cdot b^i(\cdot, s)\,\mathrm{d}B_s, i = 1, 2$$

be Ito processes. Then

$$\langle X^1, X^2\rangle_t = \int_0^t b^1(\cdot, s)b^2(\cdot, s)\,\mathrm{d}s.$$

Informally we would express this result as

$$d\langle X^1, X^2\rangle = b_t^1 b_t^2\,\mathrm{d}t \tag{2.19}$$

We conclude this section by giving the Ito-formula for Ito-processes. Importantly this is just a special case of (2.18) (in particular it is recommended to only memorize (2.18) and the 'rule' given in (2.19)).

Theorem 2.42. *Let $f \in C^{1,2}([0,\infty) \times \mathbb{R})$ and $X = \int_0^\cdot a(\cdot, s)\,\mathrm{d}s + \int_0^\cdot b(\cdot, s)\,\mathrm{d}B_s$ be an Ito process. Then we have*

$$\begin{aligned} f(t, X_t) &= f(0,0) + \int_0^t f_t(s, X_s)\,\mathrm{d}s + \int_0^t f_x(s, X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f_{xx}(s, X_s)\,\mathrm{d}\langle X\rangle_s \\ &= f(0,0) + \int_0^t \left(f_t(s, X_s) + f_x(s, X_s)a(\cdot, s) + \frac{1}{2}f_{xx}(s, X_s)b^2(\cdot, s)\right)\mathrm{d}s \\ &\quad + \int_0^t f_x(s, X_s)b(\cdot, s)\,\mathrm{d}B_s \end{aligned}$$

for every $t \in [0, T]$.

2.9 Introduction to stochastic differential equations

We consider stochastic differential equations (SDE) of the form

$$\begin{cases} \mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \\ X_0 = x_0, \end{cases}$$

where $x_0 \in \mathbb{R}$. This equation should be interpreted as an informal way of expressing the corresponding integral equation

$$X_t = x_0 + \int_0^t \mu(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s.$$

In view of our goals later on, the following example is the most important part of this section:

Example 2.43 (Geometric Brownian motion)**.** *We consider the SDE*

$$\begin{cases} \mathrm{d}X_t = \mu X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}B_t, \\ X_0 = x_0, \end{cases}$$

where $x_0, \mu \in \mathbb{R}$ *and* $\sigma > 0$*. We start by making the ansatz* $X_t = f(t, B_t)$*. By Ito's formula we obtain*

$$\begin{aligned} \mathrm{d}X_t &= f_t(t, B_t)\,\mathrm{d}t + f_x(t, B_t)\,\mathrm{d}B_t + \tfrac{1}{2} f_{xx}(t, B_t)\,\mathrm{d}t \\ &= \left(f_t(t, B_t) + \tfrac{1}{2} f_{xx}(t, B_t)\right)\,\mathrm{d}t + f_x(t, B_t)\,\mathrm{d}B_t \\ &\overset{!}{=} \mu X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}B_t \\ &= \mu f(t, B_t)\,\mathrm{d}t + \sigma f(t, B_t)\,\mathrm{d}B_t. \end{aligned}$$

By comparison of coefficients we get $\mu f = f_t + \frac{1}{2} f_{xx}$ *and* $\sigma f = f_x$*. From the second equation we obtain*

$$f(t, x) = \exp(\sigma x + g(t))$$

for $(t, x) \in [0, \infty) \times \mathbb{R}$*. Plugging in* f *into the first equation yields*

$$\mu f = g' f + \tfrac{1}{2}\sigma^2 f,$$

so that for instance

$$g(t) = \left(\mu - \tfrac{1}{2}\sigma^2\right) t$$

for $t \in [0, \infty)$*. Hence,*

$$X_t = x_0 \exp\left(\sigma B_t + \left(\mu - \tfrac{1}{2}\sigma^2\right) t\right)$$

which is the Black-Scholes model, the price process of a financial asset with drift μ *and volatility* σ*.*

For completeness we state (without proof) the most important criterion for existence and uniqueness of solutions to SDEs. The Lipschitz conditions appearing therein should be familiar from the theory of deterministic ODEs.

Theorem 2.44 (Existence and uniqueness of solutions)**.** *Let* $\mu, \sigma \colon [0, \infty) \times \mathbb{R} \to \mathbb{R}$ *such that*

$$\begin{aligned} |\mu(t, x) - \mu(t, y)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 &\leq K|x - y|^2, \\ |\mu(t, x)|^2 + |\sigma(t, x)|^2 &\leq K\left(1 + |x|^2\right) \end{aligned}$$

hold for all $t \in [0, T]$*,* $x, y \in \mathbb{R}$ *for some* $K > 0$*. Then the SDE*

$$\begin{cases} \mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \\ X_0 = x_0, \end{cases}$$

has a unique continuous, adapted solution with $\sup_{t \in [0,T]} \mathbb{E}\left[|X_t|^2\right] < \infty$*.*

2.10 Stochastic integral and martingales

Recall that a process $H = (H_t)_{t\in[0,T]}$ is simple if

$$H_t(\omega) = \sum_{i=1}^{n} H^i I_{(s_i,s_{i+1}]},$$

where $0 \le s_1 \le \ldots \le s_n$ and each H^i is $\mathcal{F}_{s_i}$-measurable and that for simple integrands the stochastic integral is given by

$$(H \cdot X)_u := \int_0^u H_s\, dX_s := I_u := \int_0^u H_t\, dX_t := \sum_{i=1}^{n} H^i (X_{s_{i+1}\wedge u} - X_{s_i \wedge u}). \tag{2.20}$$

From this definition it is straight forward (Exercise) to see that if M is a continuous martingale and H is a bounded simple process, then $(H \cdot M)_t, t \in [0,T]$ is a continuous martingale.

Using limiting arguments, this can be extended to the case of general H. For instance one can prove:

Proposition 2.45. *Let H be a bounded predictable process and let M be a continuous martingale. Then $(H \cdot M)_t, t \in [0,T]$ is a continuous martingale.*

(More generally, if H is locally bounded and M is a continuous martingale, then $(H \cdot M)$ is continuous local martingale.)

As usual, such results are simpler to prove in the case where we are integrating against Brownian motion. In this case we have the following:

Theorem 2.46. *Let B be a Brownian motion and $\varphi \in \mathcal{H}^2([0,T])$. Then*

$$X_t := \int_0^t \varphi_s\, \mathrm{d}B_s$$

is a continuous martingale and $\mathbb{E}[X_T^2] < \infty$.

Proof. Exercise. □

We also note that an Ito-process is a martingale only if the "dt-part" vanishes: the Ito process

$$X_t = X_0 + \int_0^t a(\cdot, s)\, \mathrm{d}s + \int_0^t b(\cdot, s)\, \mathrm{d}B_s,$$

is a martingale if and only if $a(\cdot,\cdot) = 0$, $\mathbb{P} \otimes \lambda$-a.s. (Exercise.) In view of this, the term $\int_0^t a(\cdot, s)\, \mathrm{d}s$ is called *drift part*, while $\int_0^t b(\cdot, s)\, \mathrm{d}B_s$ is called *martingale part*.

The above comments should are probably not very surprising. In contrast the following represents a rather remarkable converse of Theorem 2.46:

Theorem 2.47 (martingale representation theorem). *Let $(B_t)_{t=0}^T$ be Brownian motion and write $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$ for the filtration generated by B. Let $(X_t)_{t\in[0,T]}$ be a martingale adapted to $(\mathcal{F}_t)_{t\in[0,T]}$ and $\mathbb{E}[X_T^2] < \infty$. Then there exists a unique $\varphi \in \mathcal{H}^2([0,T])$ such that*

$$X_t = X_0 + \int_0^t \varphi_s \, \mathrm{d}B_s \tag{2.21}$$

for every $t \in [0,T]$.

Below we will prove an important special case of Theorem 2.47. Before going into this, we make some important comments: First of all $X_0 = \mathbb{E}X_T$ since $\mathcal{F}_0$ is the trivial σ-algebra. Moreover, uniqueness of ϕ is a straight forward consequence of Ito's isometry.

Next we claim that it is sufficient for Theorem 2.47 to show that for each $X \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ there exists $\varphi \in \mathcal{H}^2([0,T])$ such that

$$X_T = \mathbb{E}[X_0] + \int_0^T \varphi_t \, \mathrm{d}B_t. \tag{2.22}$$

Indeed, (2.21) follows from (2.22) by applying the conditional expectation operator $\mathbb{E}[\cdot|\mathcal{F}_t]$.

We will now prove (2.22) in an important and instructive case:

Assume that $X = f(B_T)$, where $f \in C^2(\mathbb{R})$ is such that $\mathbb{E}[f(B_T)^2] < \infty$. We define the martingale $(X_t)_{t\in[0,T]}$ through $X_t = \mathbb{E}[f(B_T)|\mathcal{F}_t]$. The crucial point is to notice that there exists a function $f(t,b), (t,b) \in [0,T] \times \mathbb{R}$ such that

$$\mathbb{E}[f(B_T)|\mathcal{F}_t] = f(t, B_t). \tag{2.23}$$

To see this, note first that $\mathbb{E}[f(B_T)|\mathcal{F}_t] = \mathbb{E}[f(B_t + (B_T - B_t))|\mathcal{F}_t] = \mathbb{E}[f(B_t + (B_T - B_t))|B_t]$ since $(B_T - B_t)$ is independent of $\mathcal{F}_t$. Moreover $\mathbb{E}[f(B_t + (B_T - B_t))|B_t] = f(t, B_t)$, where

$$f(t,b) := \int f(b+y) \, d\gamma_{T-t}(y) \tag{2.24}$$

and γ_{T-t} denotes the centered Gaussian with variance $T - t$. (Exercise.)

Since $f \in C^2$, we can apply Ito's formula to the process $X_t = f(t, B_t)$ to obtain

$$dX_t = f_x(t, B_t) \, dB_t + [f_{xx}(t, B_t) + f_t(t, B_t)] \, dt. \tag{2.25}$$

Next we note that the drift term $[f_{xx}(t, B_t) + f_t(t, B_t)]$ vanishes. This can either be shown directly from the definition of $f(t,b)$ or, more elegantly, by noticing that the drift term vanishes necessarily since $X_t = \mathbb{E}[f(B_T)|\mathcal{F}_t]$ is a martingale by definition. It follows that (3.12) asserts (in integral form) that

$$X_t = X_0 + \int_0^t f_x(u, B_u) \, dB_u \tag{2.26}$$

as required.

Summing up, in the particular case where X_T is given in the form $f(B_T)$ we have not only established the martingale representation theorem, but we have also found an explicit representation of the required integrand φ.

We note that the approach presented here can in fact be used to establish Theorem 2.47 in the general case: In the first step, one iterates the above idea to provide an explicit representation in the case where $X = f(B_{t_1}, \ldots, B_{t_n})$ for $0 \le t_1 \le \ldots \le t_n \le T$. Then, in the second step, one uses that the set of all X of this form is dense in $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$.

The next two results highlight the particular role the Brownian motion has in stochastic analysis.

Theorem 2.48 (Levy's characterization of Brownian motion)**.** *Let M be a continuous (local) martingale starting at $M_0 = 0$ and assume that $\langle M \rangle_t = t$ for $t \ge 0$. Then M is Brownian motion.*

Sketch of proof. What we should prove that differences of the form $M_{t_{i+1}} - M_{t_i}$ are Gaussian with variance $t_{i+1} - t_i$. Instead of this we just show that $M_t \sim N(0, t)$. (The general case would follow using the same idea.) To do this we will calculate the *moment generating function*

$$\lambda \mapsto \mathbb{E} \exp(\lambda M_t) = \phi_t(\lambda).$$

To show that $M_t \sim N(0, t)$ we need to prove that ϕ_t is the moment generating function of the appropriate normal distribution, i.e. that $\phi_t(\lambda) = \frac{1}{2}\lambda^2 t$. To do this we consider the process

$$X_t := \exp(\lambda M_t - \tfrac{1}{2}\lambda^2 t).$$

By Ito's formula, we have

$$dX_t = \lambda X_t \, dM_t + \tfrac{1}{2}\lambda^2 d\langle M_t \rangle - \tfrac{1}{2}\lambda^2 \, dt = \lambda X_t \, dM_t.$$

Ignoring issues of boundedness, we thus obtain that X is a martingale hence we have

$$\mathbb{E} \exp(\lambda M_t - \tfrac{1}{2}\lambda^2 t) = \mathbb{E} X_t = \mathbb{E} X_0 = 1$$

which yields $\mathbb{E} \exp(\lambda M_t) = \frac{1}{2}\lambda^2 t$ as required.

In order to give the real proof, one usually works with the characteristic functions instead of the moment generating function since this avoids (as usual) problems associated to boundedness. I went with the moment generating function to avoid considering complex numbers. $\square$

Let X be a stochastic process, let $(\tau_t)_{t \ge 0}$ be a family of stopping times such that for $s < t$ we have $\tau_s \le \tau_t$. Then $Y_t := X_{\tau_t}, t \ge 0$ is called a *time-change* of X.

In essence, the following results says that every continuous martingale looks like Brownian motion up to a time-change.

Corollary 2.49. *Let M be a continuous martingale such that $M_0 = 0$ and $\langle M \rangle_\infty = \infty$. Set*

$$\tau_t := \inf\{u \geq 0 : \langle M \rangle_u = t\}.$$

Then $B_t := M_{\tau_t}$ is a Brownian motion.

Sketch of Proof. To establish this, one proves that a time-change of a martingale is still a martingale and that

$$\langle B \rangle_t = \langle M \rangle_{\tau_t} = t.$$

Then Levy's characterization theorem applies. □

2.11 Girsanov's Theorem

In this section we present the final ingredient from stochastic analysis that is important for our considerations in mathematical finance.

A basic viewpoint on this result is the following: Consider a Brownian motion with drift, say

$$X_t := B_t + \mu \cdot t, \quad t \geq 0,$$

where μ is a positive constant. We are interested to determine in which sense X behaves is different from a Brownian motion.

First of all note that the paths of X still have the same quadratic variation as the paths of B since the finite variation part is irrelevant for this.

A crucial difference between X and B is the following: At time t the paths of B are (on average) close to 0, paths which are at level $\pm\mu t$ are relatively rare. In contrast, typical paths of X at time t are at height μt, while paths which are at level 0 (or $2\mu t$) are rare. The idea behind Girsanov's theorem is that one can turn X into a Brownian motion by changing the probability of individual paths: I.e., to make X look like a Brownian motion, paths which end at 0 should get a much higher weight while paths that end at μt should get a much smaller weight. In Girsanov's Theorem one changes the underlying measure accordingly, so that X becomes a Brownian motion.

Theorem 2.50 (Girsanov)**.** *For $T > 0$ let $\mu \in \mathcal{H}^2([0,T])$ be bounded. Consider*

$$X_t := B_t + \int_0^t \mu_s \, \mathrm{d}s$$

and

$$M_t := \exp\left(-\int_0^t \mu_s \, \mathrm{d}B_s - \frac{1}{2}\int_0^t \mu_s^2 \, \mathrm{d}s\right)$$

for $t \in [0,T]$. Define $\mathbb{Q}$ by $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} := M_T$.

(i) The process $(M_t)_{t\in[0,T]}$ is a $\mathbb{P}$-martingale.

(ii) The process $(X_t M_t)_{t\in[0,T]}$ is a $\mathbb{P}$-martingale.

(iii) The process $(X_t)_{t\in[0,T]}$ is a $\mathbb{Q}$-Brownian motion.

(As common by now, I advise to only skim over the proof.)

Proof.

(i) Ito's formula applied to the Ito process

$$F_t := -\int_0^t \mu_s \,\mathrm{d}B_s - \frac{1}{2}\int_0^t \mu_s^2 \,\mathrm{d}s$$

yields

$$\begin{aligned}\mathrm{d}M_t &= \exp(F_t)\,\mathrm{d}F_t + \tfrac{1}{2}\exp(F_t)\,\mathrm{d}\langle F\rangle_t \\ &= -M_t\mu_t\,\mathrm{d}B_t - \tfrac{1}{2}M_t\mu_t^2\,\mathrm{d}t + \tfrac{1}{2}M_t\mu_t^2\,\mathrm{d}t \\ &= -\mu_t M_t\,\mathrm{d}B_t.\end{aligned}$$

Hence, M is a positive local martingale and therefore a supermartingale (by Fatou's lemma). Moreover, for $t\in[0,T]$ and every $p>1$ we have

$$\begin{aligned}M_t^p &= \exp\left(-\int_0^t p\mu_s\,\mathrm{d}B_s - \frac{1}{2}\int_0^t (p\mu_s)^2\,\mathrm{d}s\right)\exp\left(\frac{p(p-1)}{2}\int_0^t \mu_s^2\,\mathrm{d}s\right) \\ &= S_t \exp\left(\frac{p(p-1)}{2}\int_0^t \mu_s^2\,\mathrm{d}s\right),\end{aligned}$$

where S is a supermartingale. Therefore, with $|\mu|\le c$ we obtain

$$\mathbb{E}\left[\int_0^T \mu_s^2 M_s^2\,\mathrm{d}s\right] \le c^2\int_0^T \mathbb{E}[M_s^2]\,\mathrm{d}s \le c^2 T\exp(c^2 T).$$

Hence, $\mu M\in\mathcal{H}^2([0,T])$ so that M is a $\mathbb{P}$-martingale. In particular, we have $\mathbb{E}[M_T] = M_0 = 1$.

(ii) Let $Y := XM$. It holds

$$\begin{aligned}\mathrm{d}X_t &= \mathrm{d}B_t + \mu_t\,\mathrm{d}t, \\ \mathrm{d}M_t &= -\mu_t M_t\,\mathrm{d}B_t\end{aligned}$$

so that $\mathrm{d}\langle X,M\rangle_t = -\mu_t M_t\,\mathrm{d}t$. By the product formula we get

$$\begin{aligned}\mathrm{d}Y_t &= X_t\,\mathrm{d}M_t + M_t\,\mathrm{d}X_t + \mathrm{d}\langle X,M\rangle_t \\ &= -\mu_t M_t X_t\,\mathrm{d}B_t + M_t\,\mathrm{d}B_t + M_t\mu_t\,\mathrm{d}t - \mu_t M_t\,\mathrm{d}t \\ &= M_t(1-\mu_t X_t)\,\mathrm{d}B_t.\end{aligned}$$

Therefore Y is a local martingale. Moreover, using Hölder's inequality and (i) we have

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T M_s^2(1-\mu_s X_s)^2 \,\mathrm{d}s\right] \\
&\le \mathbb{E}\left[\int_0^T M_s^2(1+\mu_s^2+X_s^2+\mu_s^2X_s^2)\,\mathrm{d}s\right] \\
&\le (1+c^2)\left(\mathbb{E}\left[\int_0^T M_s^2\,\mathrm{d}s\right]+\mathbb{E}\left[\int_0^T M_s^4\,\mathrm{d}s\right]^{\frac{1}{2}}\mathbb{E}\left[\int_0^T X_s^4\,\mathrm{d}s\right]^{\frac{1}{2}}\right) \\
&< \infty.
\end{aligned}$$

Hence, $M(1-\mu X) \in \mathcal{H}^2([0,T])$ so that Y is a $\mathbb{P}$-martingale.

(iii) Let $0 \le s \le t \le T$ and $A \in \mathcal{F}_s$. Then

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[X_t \mathbb{1}_A] &= \mathbb{E}[M_T X_t \mathbb{1}_A] \\
&= \mathbb{E}[M_t X_t \mathbb{1}_A] \\
&= \mathbb{E}[M_s X_s \mathbb{1}_A] \\
&= \mathbb{E}[M_T X_s \mathbb{1}_A] \\
&= \mathbb{E}_{\mathbb{Q}}[X_s \mathbb{1}_A].
\end{aligned}$$

Hence X is a $\mathbb{Q}$-martingale. By Levy's characterization Theorem, X is a $\mathbb{Q}$-Brownian motion.

□

3 The financial models of Bachelier and Black-Scholes

3.1 The Bachelier model

Louis Bachelier (1870-1946). We assume that the asset follows the dynamics

$$X_t = x_0 + mt + \sigma B_t,$$

where $m \in \mathbb{R}$ is the drift parameter, $\sigma > 0$ is called the volatility and $(B_t)_{t\in[0,\infty)}$ is a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ with standard Brownian filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$. It is our goal to find a replicating strategy for a contingent claim of the form $f(X_T)$, where $f \colon \mathbb{R} \to [0,\infty)$. As usual we are particularly interested in European Call / Put options.

Example 3.1.

(i) Let $f \colon \mathbb{R} \to [0,\infty)$, $x \mapsto (x-K)^+$, then $(X_T-K)^+$ is a European call option with strike price $K > 0$.

(ii) Let $f\colon \mathbb{R} \to [0,\infty)$, $x \mapsto (K-x)^+$, then $(K-X_T)^+$ is a European put option with strike price $K>0$.

According to Girsanov's theorem,

$$B_t^* := B_t + \frac{m}{\sigma} t$$

defines a $\mathbb{P}^*$-Brownian motion, where $\frac{d\mathbb{P}^*}{d\mathbb{P}} = \exp(-\frac{m}{\sigma} B_T - \frac{m^2}{2\sigma^2} T)$. In particular,

$$X_t = x_0 + \sigma B_t^*$$

is a $\mathbb{P}^*$-martingale, that is $\mathbb{P}^*$ is an *equivalent martingale measure*.

The principle idea is now to apply the martingale representation theorem to the $\mathbb{P}^*$-martingale $M_t := \mathbb{E}_{\mathbb{P}^*}[f(X_T)|\mathcal{F}_t]$. Since our payoff function has a very simple form, we will even obtain explicit formulas:

Define $g(b^*) := f(x_0 + \sigma b^*)$ and

$$g(t, b^*) := \mathbb{E}_{\mathbb{P}^*}[g(B^*)|\mathcal{F}_t]$$

so that $g(t, B_t^*) = M_t$. As in (3.11) we have

$$g(t,b) = \int g(b+y)\, d\gamma_{T-t}(y) \tag{3.1}$$

and

$$M_t = M_0 + \int_0^t g_b(s, B_s^*) dB_s^*. \tag{3.2}$$

In terms of X we can rewrite this as

$$M_t = M_0 + \int_0^t \tfrac{1}{\sigma} g_b(s, (X_s - x_0)/\sigma) dX_s^*.$$

Noting that $M_0 = \mathbb{E}_{\mathbb{P}^*}[f(X_T)] =: p$ we thus obtain

$$f(X_T) = p + (H \cdot X)_T, \tag{3.3}$$

where $H_t = \frac{1}{\sigma} g_b(t, (X_t - x_0)/\sigma)$.

Remark 3.2. *In the above treatise we have been imprecise at two points:*

1. *To arrive at (3.11), we would like the function $f(t,x)$, or, $g(t,b)$, resp. to be C^2. This is not the case if take f to be a European put / call function. However this can be easily overcome: The convolution with respect to a Gaussian in (3.1) is 'smoothing' and hence $g(t,b)$* is *C^2 for $t<T$. It is then not hard to see that this is enough for our arguments to go through.*

2. *We would like the hedging relation (3.3) to hold with respect to $\mathbb{P}$ but our derivation was with respect to the measure $\mathbb{P}^*$. In principle there could be problem here since the stochastic integral was defined with a fixed underlying probability measure in mind. However, this is not an issue since the stochastic integral remains the same as long as one switches to an* equivalent *probability. This is trivial as long as one considers only simple integrands. Moreover this remains true when passing to limits in probability (which remains unaltered under changes to an equivalent probability measure) which allows us to conclude.*

Example 3.3. *Consider again the European call option $f(X_T) = (X_T - K)^+$. Since f is increasing and 1-Lipschitz, it follows that g is increasing and σ-Lipschitz. Since these properties are preserved under convolution with a Gaussian, $g(t, .)$ is also increasing and σ-Lipschitz. As the convolution operator smoothes, $g(t, .)$ is differentiable for $t \in [0, T)$ and the derivative satisfies $g_b(t, .) \in [0, \sigma]$. Considering the hedging strategy obtained in (3.2), we find that*

$$H_t(X_t) = \tfrac{1}{\sigma} g_b(t, (X_t - x_0)/\sigma) \in [0, 1].$$

In particular, it is no "short-selling" is necessary to hedge a European call option in the Bachelier-model.

Note also that convexity of f implies that $g_b(t, \cdot)$ and then also $H_t(\cdot)$ is increasing.

Theorem 3.4. *In the Bachelier model, the fair price of a contingent claim $f(X_T)$ is given by $\mathbb{E}_{\mathbb{P}^*}[f(X_T)]$.*

More generally then Theorem 3.4 we have:

Theorem 3.5. *Let $G \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}^*)$ be a contingent claim. Then there exists a unique pair (a, H), $a \in \mathbb{R}$, $H \in \mathcal{H}_2(0, T)$ such that*

$$a + (H \cdot X)_T = G. \tag{3.4}$$

Moreover we have $a = \mathbb{E}_{\mathbb{P}^*}[G]$.

Proof. We apply the martingale representation theorem to G and the $\mathbb{P}^*$-Brownian motion B^* to obtain a strategy $\tilde{H}$ such that

$$\mathbb{E}_{\mathbb{P}^*}[G] + \int_0^T \tilde{H}\, dB^* = G.$$

Setting $H_t := \frac{1}{\sigma}\tilde{H}_t$ for $t \le T$ we arrive at

$$G = \mathbb{E}_{\mathbb{P}^*}[G] + \int_0^T H\, d(\sigma B^*) = \mathbb{E}_{\mathbb{P}^*}[G] + \int_0^T H\, dX.$$

As in the martingale representation theorem, a and H are uniquely determined. □

In reminiscence of Theorem 1.24 a consequence of Theorem 3.14 is the following:

Theorem 3.6. *The measure $\mathbb{P}^*$ is the only equivalent martingale measure.*

Idea of proof. For instance one can show that the claims of the form $G = a + (H \cdot X)_T$ with H bounded are dense in $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ and for these claims all equivalent martingale measures $\tilde{\mathbb{P}}$ satisfy $\mathbb{E}_{\tilde{\mathbb{P}}} G = a$. □

3.2 Geometric Brownian motion – the Black-Scholes model

The asset is modeled by

$$X_t = X_0 \exp\left(\sigma B_t + \left(m - \tfrac{1}{2}\sigma^2\right) t\right),$$

that is X is the solution of $\mathrm{d}X_t = X_t(\sigma \, \mathrm{d}B_t + m \, \mathrm{d}t)$. As before, $m \in \mathbb{R}$ is called the drift parameter, $\sigma > 0$ is called the volatility and $(B_t)_{t\in[0,\infty)}$ is a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ with standard Brownian filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$. Again, we want to find a replicating strategy for a contingent claim of the form $G = f(X_T)$, where $f\colon \mathbb{R} \to [0, \infty)$. That is, we would like to have a perfect hedge in the sense that

$$G = f(X_T) = a + \int_t^T H_s \, \mathrm{d}X_s,$$

for some pair (a, H). In fact, as before, it would be highly appreciated if we can also obtain concrete representations of a and H. As in the case of Bachelier's model, our starting point is to appropriately change the underlying probability measure to turn X into a martingale.

According to Girsanov's theorem,

$$B_t^* := B_t + \frac{m}{\sigma} t$$

defines a $\mathbb{P}^*$-Brownian motion, where $\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}} = \exp(-\frac{m}{\sigma} B_T - \frac{m^2}{2\sigma^2} T)$.

Lemma 3.7. *The process*

$$X_t = X_0 \exp\left(\sigma B_t^* - \tfrac{1}{2}\sigma^2 t\right) \tag{3.5}$$

is a $\mathbb{P}^*$-martingale.

Proof. Since $B_t^* - B_s^*$ is independent of $\mathcal{F}_s$, we have

$$\begin{aligned}
\mathbb{E}^*\left[X_t \middle| \mathcal{F}_s\right] &= \mathbb{E}^*\left[X_s e^{\sigma(B_t^* - B_s^*) - \frac{1}{2}\sigma^2(t-s)} \middle| \mathcal{F}_s\right] \\
&= X_s e^{-\frac{1}{2}\sigma^2(t-s)} \mathbb{E}^*\left[e^{\sigma(B_t^* - B_s^*)}\right] \\
&= X_s.
\end{aligned}$$

□

Remark 3.8. *Sometimes the term 'geometric Brownian motion' is reserved to the process (3.5) under the measure $\mathbb{P}^*$, i.e. to the particular case where there is no drift appearing. Sometimes one asks in addition that also $\sigma \equiv 1$.*

As before X is also determined through the SDE

$$dX_t = \sigma X_t \, dB_t^* \tag{3.6}$$

We will now give two (very) slightly different derivations for the price / hedge of a contingent claim $f(X_t)$:

1. As in the case of the Bachelier model we can proceed by simply rewriting all relevant terms using B_t^* instead of X_t.

   I.e. we set

   $$f(X_T) = f\left(\exp^{\sigma B_t^* - 1/2\sigma^2 T}\right) =: g(B_t^*)$$

   and define $g(t,b) := \int g(b+y)\, d\gamma_{T-t}(y)$ as before. Note that

   $$B_t^* = \tfrac{1}{\sigma}\log(X_t) + \tfrac{1}{2}\sigma t, \quad dB_t = \tfrac{dX_t}{\sigma X_t}.$$

   We thus obtain

   $$\begin{aligned} f(X_T) - \mathbb{E}_{\mathbb{P}^*} f(X_T) &= g(B_T^*) - \mathbb{E}_{\mathbb{P}^*} g(B_T^*) & (3.7)\\ &= \int_0^T g'(t, B_t^*)\, dB_t^* & (3.8)\\ &= \int_0^T g'\left(t, \tfrac{1}{\sigma}\log(X_t) + \tfrac{1}{2}\sigma t\right) \tfrac{dX_t}{\sigma X_t}. & (3.9) \end{aligned}$$

   We have thus fund a hedging strategy for the claim $f(X_T)$ as desired. Through some calculations (which are only mildly tedious) one can also eliminate the appearance of $g'(t,b)$ in the above formula.

2. An alternative to the above derivation is to redo the our derivation of the martingale representation theorem directly for the process X instead of Brownian motion:

   We define the $\mathbb{P}^*$-martingale $(M_t)_{t\in[0,T]}$ through $M_t = \mathbb{E}_{\mathbb{P}^*}[f(X_T)|\mathcal{F}_t]$. We have

   $$X_T = X_t \exp^{\sigma(B_T - B_t) - \frac{1}{2}\sigma^2(T-t)}.$$

   Thus

   $$\mathbb{E}_{\mathbb{P}^*}[f(X_T)|\mathcal{F}_t] = f(t, X_t), \tag{3.10}$$

where

$$f(t,x) := \int f\left(x\exp^{\sigma y-\frac{1}{2}\sigma^2(T-t)}\right)\,d\gamma_{T-t}(y) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f\left(xe^{\sigma\sqrt{T-t}y-\frac{1}{2}\sigma^2(T-t)}\right)e^{-\frac{y^2}{2}}\,\mathrm{d}y. \tag{3.11}$$

and γ_{T-t} denotes the centered Gaussian with variance $T-t$ as before.

Applying Ito's formula to the process $M_t = f(t, X_t)$ (and using that M_t is a martingale) we obtain

$$df(t,X_t) = f_x(t,X_t)\,dX_t. \tag{3.12}$$

Summing up we obtain the desired hedging strategy

$$f(X_T) = \mathbb{E}_{\mathbb{P}^*} f(X_T) + \int_0^T f_x(t,X_t)\,dX_t. \tag{3.13}$$

As in the case of the Bachelier model, pricing and hedging works for much more derivatives:

Theorem 3.9. *Let $G \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}^*)$ be a contingent claim. Then there exists a unique pair (a, H) such that*

$$a + (H\cdot X)_T = G, \tag{3.14}$$

and $\tilde{H} \in \mathcal{H}_2(0,T)$, where $\tilde{H}_t := \sigma X_t H_t, t \le T$. Moreover we have $a = \mathbb{E}_{\mathbb{P}^*}[G]$.

Proof. We apply the martingale representation theorem to G and the $\mathbb{P}^*$-Brownian motion B^* to obtain a strategy $\tilde{H}$ such that

$$\mathbb{E}_{\mathbb{P}^*}[G] + \int_0^T \tilde{H}\,dB^* = G.$$

Setting $H_t := \frac{1}{\sigma X_t}\tilde{H}_t$ for $t \le T$ we arrive at

$$G = \mathbb{E}_{\mathbb{P}^*}[G] + \int_0^T H\sigma X_t\,dB^* = \mathbb{E}_{\mathbb{P}^*}[G] + \int_0^T H\,dX.$$

As in the martingale representation theorem, a and H are uniquely determined. □

We conclude this section with some further remarks concerning pricing and hedging:

1. Assume that G is a contingent claim and a, H are so that

$$a + (H\cdot X)_T = G.$$

Setting

$$G_t := \mathbb{E}_{\mathbb{P}^*}[G|\mathcal{F}_t],$$

we then have

$$a + (H \cdot X)_t = G(t)$$

which entails that

$$G(t) + \int_t^T H_s \, dX_s = G.$$

The mathematical finance interpretation of this equation is that, at time t (and given the information available up to time t), G_t is exactly the amount of money that is required to replicate the future payoff G (using the strategy $(H_s)_{s=t}^T$). This implies that the fair price of G at time is given by $G(t)$. We make the important comment that the argument so far does not depend on the fact that we are working with the Black-Scholes model.

Let us now specify to the case where $G = f(X_T)$. Using the notations above we then have

$$f(t, X_t) = G_t = \mathbb{E}_{\mathbb{P}^*}[f(X_T)|\mathcal{F}_t] = \mathbb{E}_{\mathbb{P}^*}[f(X_T)|X_t],$$

and, resp. $f(t,x) = \mathbb{E}_{\mathbb{P}^*}[f(X_T)|X_t = x]$. In particular $f(t,x)$ denotes the fair price at time t of the payoff $f(X_T)$ given the asset price at time t equals x. Recall also that by (3.13) we have $f(X_T) = f(0, X_0) + (H \cdot X)_T$ for the strategy $H = f_x(t, X_t)$. We are now in the position to give a new interpretation of the hedging strategy H: the amount of stocks we should hold at time t is the *derivative* of the value of the contingent claim.

2. ("Put-Call parity"). There is an important relationship between the prices of call- and put options: Observe that

$$(X_T - k)_+ - (k - X_T)_+ = X_T - k.$$

Applying the expectation operator wrt $\mathbb{P}^*$ we thus obtain

$$\text{price(call)} - \text{price(put)} = X_0 - k.$$

Once again we note that this does actually *not* depend on the specific structure of the Black-Scholes model.

3.3 The Call option in the Black-Scholes model

Warning: the calculations in this section are tedious and the whole chapter will not be eaxmined.

Our starting point is formula (3.11) which yields and explicit representation for the $\mathbb{P}^*$-martingale $\mathbb{E}_{\mathbb{P}^*}[f(X_T)|\mathcal{F}_t] = f(t, X_t)$

Defining $\alpha := \sigma\sqrt{T-t}$ we have

$$\begin{aligned} xe^{\alpha y - \frac{1}{2}\alpha^2} \geq K &\iff \log \tfrac{x}{K} + \alpha y - \tfrac{1}{2}\alpha^2 \geq 0 \\ &\iff y \geq \tfrac{1}{\alpha}\left(-\log \tfrac{x}{K} + \tfrac{1}{2}\alpha^2\right) =: z. \end{aligned}$$

Hence, we have

$$\begin{aligned} f(t,x) &= \frac{1}{\sqrt{2\pi}} \int_z^\infty \left(xe^{\alpha y - \frac{1}{2}\alpha^2} - K\right) e^{-\frac{y^2}{2}} \,\mathrm{d}y \\ &= \frac{x}{\sqrt{2\pi}} \int_z^\infty e^{\alpha y - \frac{1}{2}\alpha^2 - \frac{y^2}{2}} \,\mathrm{d}y - K\Phi(-z) \\ &= \frac{x}{\sqrt{2\pi}} \int_z^\infty e^{-\frac{(y-\alpha)^2}{2}} \,\mathrm{d}y - K\Phi(-z) \\ &= \frac{x}{\sqrt{2\pi}} \int_{z-\alpha}^\infty e^{-\frac{u^2}{2}} \,\mathrm{d}u - K\Phi(-z) \\ &= x\Phi(\alpha - z) - K\Phi(-z). \end{aligned}$$

In particular, we obtain the Black-Scholes formula

$$f(t,x) = x\Phi(d_+(t,x)) - K\Phi(d_-(t,x)), \tag{3.15}$$

where

$$d_\pm(t,x) = \frac{\log \frac{x}{K} \pm \frac{1}{2}\alpha^2}{\alpha} = \frac{\log \frac{x}{K} \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}.$$

Of course we are particularly interested in the European call option, i.e. $f^{\text{call}}(x) = (x-k)_+$ (and $f^{\text{call}}(.,.)$ defined accordingly).

Lemma 3.10. *It holds $f_x^{\text{call}}(t,x) = \Phi(d_+(t,x))$.*

Proof. Let $\varphi := \Phi'$, then we have

$$f_x^{\text{call}}(t,x) = \Phi(d_+(t,x)) + x\varphi(d_+(t,x))\frac{\partial}{\partial x}d_+(t,x) - K\varphi(d_-(t,x))\frac{\partial}{\partial x}d_-(t,x).$$

Since $d_+(t,x) = d_-(t,x) + \sigma\sqrt{T-t}$ we have

$$-\frac{1}{2}d_+(t,x)^2 = -\frac{1}{2}d_-(t,x)^2 + \log\frac{K}{x}$$

so that

$$\begin{aligned} \frac{\partial}{\partial x}d_+(t,x) &= \frac{\partial}{\partial x}d_-(t,x), \\ \varphi(d_-(t,x)) &= \frac{x}{K}\varphi(d_+(t,x)) \end{aligned}$$

it follows $f_x^{\text{call}}(t,x) = \Phi(d_+(t,x))$. $\square$

Remark 3.11.

(i) We have $f_x^{\text{call}}(t,x) = \Phi(d_+(t,x)) \geq 0$ *so that* $x \mapsto f^{\text{call}}(t,x)$ *is increasing.*

(ii) Since $x \mapsto d_+(t,x)$ *is increasing, also* $x \mapsto f_x^{\text{call}}(t,x)$ *is increasing, i. e.* $x \mapsto f^{\text{call}}(t,x)$ *is convex.*

3.4 Some notes on more general models

An obvious question at this stage is whether the Black-Scholes model is able to describe the behaviour of financial assets appropriately.

A simple way to test this is to look a the "log-returns", i.e. $\lg S_t$ of a financial asset. The Black-Scholes model postulates that we should observe a dynamic of the form

$$\lg S_t = \sigma B_t + \tilde{\mu} t$$

for some $\tilde{\mu} \in \mathbb{R}$, i.e. that the volatility of the log price is constant. Classical time-series analysis negates that.

Another question would be whether the Black-Scholes formula gives adequate predictions for the prices of financial derivatives. Given a single call option $C_{T,K}$ with time to maturity T and strike price K we can observe its market price $p_{T,K}$ and determine the unique $\sigma = \sigma(T,K)$ such that the Black-Scholes formula reproduces the price $p_{T,K}$. The value $\sigma(T,K)$ is called the *Black-Scholes implied volatility*.

Assuming that the Black-Scholes model is correct, $\sigma(T,K)$ should be constant in time and space. Again, empirical evidence suggest that this is not the case: Generally speaking, $T \mapsto \sigma(T,K)$ is increasing while $K \mapsto \sigma(T,K)$ appears to be convex in K with a minimal value be attained for K close to S_0. [3]

During the last (three or so) decades substantial effort has gone into the question to find models which provide better descriptions of real financial markets. Frequently considered models are the so called *Heston-model* and the *SABR-model*. In the last five years, *rough volatility models* have received specific attention.

In practise, Black-Scholes is still the most frequently used model. Maybe the second most common model is the so called *local volatility model* introduced by Dupire. The principal idea is to modify the Black-Scholes model or the Bachelier model by allowing σ to depend also on time and space: that is on assumes that asset price process under the risk-neutral measure follows the dynamics

$$dS_t = \sigma_{loc}(t, S_t)\, dB_t. \tag{3.16}$$

Interestingly, all sufficiently regular models can, in a certain sense be approximated by a model with dynamics as in (3.17)

[3] Interestingly, it has been argued that the Black-Scholes model was an adequate model for many financial markets until the 1987 stock-crash an is much less adequate since then.

Theorem 3.12 (Gyöngy, '81). *Assume the a process $\tilde{S}$, $\tilde{S}_0 \in \mathbb{R}$ has dynamics given by*

$$d\tilde{S}_t = \sigma(t) dB_t, \tag{3.17}$$

where σ is an adapted process which is bounded away from 0 and ∞. Set

$$\sigma^2_{loc}(t, s) := \mathbb{E}[\sigma^2 | S_t = s].$$

Then there is a process S with $S_0 = \tilde{S}_0$ which satisfies the SDE

$$dS_t = \sigma_{loc}(t, S_t)\, dB_t. \tag{3.18}$$

Bruno Dupire realised that if the prices of all call options $C(T, K), T, K \geq 0$ are known from market data (and satisfy some mild regularity conditions), then there exists a unique function $\sigma_l oc$ such the model given by (3.17) reproduces these prices exactly.

This is reassuring in that we can specify at least some model which is consistent with market data.

4 Risk measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

We interpret given random variables X, Y as potential losses a company might endure. A risk measure is a mapping that assigns to each random variable a real number that aims to quantify the 'amount of risk' associated to the corresponding random variable. For instance we might want to compare the risk associated to X and Y resp. or we might want to evaluate how much reserve we should put aside to protect against a certain amount of risk.

4.1 Mean-variance, Sharpe ratio, value at risk, expected shortfall

Definition 4.1. *Let $X \in L^0(\mathbb{P})$.*

(i) If $X \in L^2(\mathbb{P})$ define the variance of X by $\mathbb{V}\mathrm{ar}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$.

(ii) If $X \in L^2(\mathbb{P})$ define the standard deviation of X by $\sigma(X) := \sqrt{\mathbb{V}\mathrm{ar}[X]} = \|X - \mathbb{E}[X]\|_2$.

(iii) Define the cumulative distribution function of X by

$$F_X : \mathbb{R} \to [0, 1],\ t \mapsto \mathbb{P}(X \leq t).$$

(iv) For $u \in (0, 1)$ a u-quantile of X is any $q \in \mathbb{R}$ satisfying

$$\mathbb{P}(X < q) \leq u \leq \mathbb{P}(X \leq q).$$

Further, we define the upper quantile function by

$$q_X : (0, 1) \to \mathbb{R},\ u \mapsto \inf\{x \in \mathbb{R} : F_X(x) > u\}$$

and note that it holds $q_X(u) = \sup\{x \in \mathbb{R} : F_X(x) \leq u\}$.

Remark 4.2. *For every $x \in \mathbb{R}$ one has*

$$\{u \in (0,1) : u < F_X(x)\} \subseteq \{u \in (0,1) : q_X(u) \leq x\} \subseteq \{u \in (0,1) : u \leq F_X(x)\}.$$

Hence, $\mathbb{P}(X \leq x) = F_X(x) = \lambda(q_X \leq x)$, where λ denotes the Lebesgue measure on $\mathbb{R}$. In particular, $\mathbb{P} \circ X^{-1} = \lambda \circ q_X^{-1}$.

Definition 4.3 (Mean-variance, Markowitz (1952))**.** *For $X \in L^2(\mathbb{P})$ and $\alpha > 0$ define*

$$MV^\alpha(X) := \mathbb{E}[X] - \frac{\alpha}{2} \mathbb{V}\mathrm{ar}[X].$$

Lemma 4.4. *The Mean-variance satisfies the following properties:*

(N) Normalization: It holds $MV^\alpha(0) = 0$.

(D) Distribution-based: For all $X, Y \in L^2(\mathbb{P})$ with $\mathbb{P} \circ X^{-1} = \mathbb{P} \circ Y^{-1}$ it holds $MV^\alpha(X) = MV^\alpha(Y)$.

(T) Translation property: For all $X \in L^2(\mathbb{P})$ and $m \in \mathbb{R}$ it holds $MV^\alpha(X+m) = MV^\alpha(X)+m$.

(C) Concavity: For all $X, Y \in L^2(\mathbb{P})$ and $\lambda \in [0,1]$ it holds

$$MV^\alpha(\lambda X + (1-\lambda)Y) \geq \lambda MV^\alpha(X) + (1-\lambda) MV^\alpha(Y).$$

Proof. We only show (T) and (C). Clearly,

$$\begin{aligned} MV^\alpha(X+m) &= \mathbb{E}[X+m] - \frac{\alpha}{2} \mathbb{V}\mathrm{ar}[X+m] \\ &= \mathbb{E}[X] + m - \frac{\alpha}{2} \mathbb{V}\mathrm{ar}[X] \\ &= MV^\alpha(X) + m. \end{aligned}$$

As the expectation is linear, it suffices to show that $\mathbb{V}\mathrm{ar}$ is convex. Indeed, due to convexity of $\mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ we have

$$\begin{aligned} \mathbb{V}\mathrm{ar}[\lambda X + (1-\lambda)Y] &= \mathbb{E}\left[(\lambda(X - \mathbb{E}[X]) + (1-\lambda)(Y - \mathbb{E}[Y]))^2\right] \\ &\leq \lambda \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] + (1-\lambda)\mathbb{E}[(Y - \mathbb{E}[Y])^2] \\ &= \lambda \, \mathbb{V}\mathrm{ar}[X] + (1-\lambda) \, \mathbb{V}\mathrm{ar}[Y]. \end{aligned}$$

□

Remark 4.5. *Note that the mean variance MV^α is not monotone, i. e. there exist $X, Y \in L^2(\mathbb{P})$ with $X \leq Y$ but $MV^\alpha(X) > MV^\alpha(Y)$.*

Definition 4.6 (Sharpe ratio, Sharp (1966)). *For $X \in L^2_+(\mathbb{P})$ with $\sigma(X) > 0$ define*

$$SR(X) := \frac{\mathbb{E}[X]}{\sigma(X)}.$$

Lemma 4.7. *The Sharpe ratio satisfies the following properties:*

(D) Distribution-based: For all $X, Y \in L^2_+(\mathbb{P})$ with $\sigma(X) > 0$, $\sigma(Y) > 0$ and $\mathbb{P} \circ X^{-1} = \mathbb{P} \circ Y^{-1}$ it holds $SR(X) = SR(Y)$.

(S) Scale-invariance: For every $X \in L^2_+(\mathbb{P})$ with $\sigma(X) > 0$ and $\lambda > 0$ it holds $SR(\lambda X) = SR(X)$.

Proof. Obvious. □

Remark 4.8. *The Sharpe ratio SR is not monotone.*

Definition 4.9 (Value at risk). *For $X \in L^0(\mathbb{P})$ the value at risk at level $\alpha \in (0, 1)$ is defined by*

$$VaR^\alpha(X) := -q_X(\alpha) = \inf\{m \in \mathbb{R} : \mathbb{P}(X + m < 0) \leq \alpha\}.$$

Lemma 4.10. *The value at risk satisfies the following properties:*

(N) Normalization: It holds $VaR^\alpha(0) = 0$.

(M) Monotonicity: For all $X, Y \in L^0(\mathbb{P})$ with $X \leq Y$ it holds $VaR^\alpha(X) \geq VaR^\alpha(Y)$.

(T) Translation property: For all $X \in L^0(\mathbb{P})$ and $m \in \mathbb{R}$ it holds $VaR^\alpha(X + m) = VaR^\alpha(X) - m$.

(P) Positive homogeneity: For all $X \in L^0(\mathbb{P})$ and $\lambda > 0$ it holds

$$VaR^\alpha(\lambda X) = \lambda VaR^\alpha(X).$$

(D) Distribution-based: For all $X, Y \in L^0(\mathbb{P})$ with $\mathbb{P} \circ X^{-1} = \mathbb{P} \circ Y^{-1}$ it holds $VaR^\alpha(X) = VaR^\alpha(Y)$.

Proof. Exercise. □

Remark 4.11. *Note that we do not have $VaR^\alpha(X + Y) \leq VaR^\alpha(X) + VaR^\alpha(Y)$ for all $X, Y \in L^0(\mathbb{P})$. There exists $n \in \mathbb{N}$ and $X_1, \ldots, X_n$ iid such that*

$$VaR^\alpha\left(\sum_{k=1}^n X_k\right) > 0 \geq nVaR^\alpha(X_1).$$

Definition 4.12 (Expected shortfall). *The expected shortfall of $X \in L^1(\mathbb{P})$ at level $\alpha \in (0,1)$ is defined by*

$$ES^\alpha(X) := -\mathbb{E}[X | X \leq q_X(\alpha)] = \mathbb{E}[-X | -X \geq VaR^\alpha(X)].$$

Lemma 4.13. *The expected shortfall satisfies the following properties:*

(N) Normalization: It holds $ES^\alpha(0) = 0$.

(M) Monotonicity: For all $X, Y \in L^1(\mathbb{P})$ with $X \leq Y$ it holds $ES^\alpha(X) \geq ES^\alpha(Y)$.

(T) Translation property: For all $X \in L^1(\mathbb{P})$ and $m \in \mathbb{R}$ it holds $ES^\alpha(X+m) = ES^\alpha(X) - m$.

(P) Positive homogeneity: For all $X \in L^1(\mathbb{P})$ and $\lambda > 0$ it holds

$$ES^\alpha(\lambda X) = \lambda ES^\alpha(X).$$

(D) Distribution-based: For all $X, Y \in L^1(\mathbb{P})$ with $\mathbb{P} \circ X^{-1} = \mathbb{P} \circ Y^{-1}$ it holds $ES^\alpha(X) = ES^\alpha(Y)$.

Proof. Exercise. □

4.2 Coherent, convex and quasi-convex risk measures

Let $\mathcal{X} \subseteq L^0(\mathbb{P})$ be a subspace containing $L^\infty(\mathbb{P})$.

Definition 4.14. *A monetary risk measure on $\mathcal{X}$ is a function $\rho \colon \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ which satisfies (N), (M) and (T), i. e.*

(N) Normalization: It holds $\rho(0) = 0$.

(M) Monotonicity: For all $X, Y \in \mathcal{X}$ with $X \leq Y$ it holds $\rho(X) \geq \rho(Y)$.

(T) Translation property: For all $X \in \mathcal{X}$ and $m \in \mathbb{R}$ it holds $\rho(X+m) = \rho(X) - m$.

A monetary risk measure ρ is called a convex risk measure, if it satisfies (C), i. e.

(C) Convexity: For all $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$ it holds

$$\rho(\lambda X + (1-\lambda)Y) \leq \lambda \rho(X) + (1-\lambda)\rho(Y).$$

A convex risk measure ρ is called a coherent risk measure, if it satisfies (P), i. e.

(P) Positive homogeneity: For all $X \in \mathcal{X}$ and $\lambda > 0$ it holds

$$\rho(\lambda X) = \lambda \rho(X).$$

Lemma 4.15.

(i) Under (P), convexity (C) and subadditivity (SA) are equivalent.

(ii) A monetary risk measure is Lipschitz continuous with respect to $\|\cdot\|_\infty$. In particular, it is real-valued on $L^\infty(\mathbb{P})$.

Proof.

(i) If ρ is convex, then $\rho(X+Y) = 2\rho\left(\frac{1}{2}X + \frac{1}{2}Y\right) \leq 2\left(\frac{1}{2}\rho(X) + \frac{1}{2}\rho(Y)\right) = \rho(X) + \rho(Y)$. Conversely, we obtain $\rho(\lambda X + (1-\lambda)Y) \leq \rho(\lambda X) + \rho((1-\lambda)Y) = \lambda\rho(X) + (1-\lambda)\rho(Y)$.

(ii) Let $X, Y \in L^\infty(\mathbb{P})$, then by (M) and (T) we have

$$\rho(X) \geq \rho(Y + \|X-Y\|_\infty) = \rho(Y) - \|X-Y\|_\infty\,,$$

i. e. $\rho(Y) - \rho(X) \leq \|X-Y\|_\infty$. Analogously, $\rho(X) \leq \rho(Y) + \|X-Y\|_\infty$.

□

Definition 4.16. *The acceptance set of a monetary risk measure ρ is given by*

$$\mathcal{A}_\rho := \{X \in \mathcal{X} : \rho(X) \leq 0\}.$$

Lemma 4.17. *For a monetary risk measure ρ we have*

$$\rho(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}_\rho\}$$

for every $X \in \mathcal{X}$. This represents the capital requirement. The monetary risk measure ρ is convex (coherent) if and only if its acceptance set $\mathcal{A}_\rho$ is convex (a convex cone).

Proof. Exercise. □

Definition 4.18. *A quasi-convex risk measure is a function $\rho\colon \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ satisfying (N), (M) and*

(Q) Quasi-convexity: For all $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$ we have

$$\rho(\lambda X + (1-\lambda)Y) \leq \rho(X) \vee \rho(Y).$$

Remark 4.19. *The condition (M) is interpreted as "more is better than less", whereas (Q) means "averages are better than extremes", i. e. we assume there exists some diversification effect.*

Lemma 4.20. *A function $\rho\colon \mathcal{X} \to \mathbb{R}$ satisfying (T) and (Q) also satisfies (C).*

Proof. For $X, Y \in \mathcal{X}$ and $\lambda \in [0,1]$ we have

$$\begin{aligned}
&\rho(\lambda X + (1-\lambda)Y) - \lambda\rho(X) - (1-\lambda)\rho(Y)\\
&= \rho(\lambda(X + \rho(X)) + (1-\lambda)(Y + \rho(Y)))\\
&\leq \rho(X + \rho(X)) \vee \rho(Y + \rho(Y))\\
&= 0.
\end{aligned}$$

□

4.2.1 Average value at risk

Definition 4.21. *For $X \in L^1(\mathbb{P})$ define the average value at risk at level $\alpha \in (0,1)$ by*

$$AVaR^\alpha(X) := \frac{1}{\alpha}\int_0^\alpha VaR^u(X)\,\mathrm{d}u = -\frac{1}{\alpha}\int_0^\alpha q_X(u)\,\mathrm{d}u.$$

Remark 4.22. *The average value at risk $AVaR^\alpha$ is a coherent risk measure. The properties (N), (M), (T), (P) and (D) are obvious, whereas (C) follows from the following theorem.*

Theorem 4.23. *For $X \in L^1(\mathbb{P})$ we have*

$$\begin{aligned} AVaR^\alpha(X) &= \max_{\mathbb{Q}\in\mathcal{Q}^\alpha} \mathbb{E}_\mathbb{Q}[-X] && (4.1)\\ &= \min_{s\in\mathbb{R}} \left(\tfrac{\mathbb{E}_\mathbb{P}[(s-X)^+]}{\alpha} - s\right), && (4.2)\end{aligned}$$

where

$$\mathcal{Q}^\alpha := \left\{\mathbb{Q} : \left\|\tfrac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right\|_\infty \leq \tfrac{1}{\alpha}\right\}.$$

The minimum in (4.2) is attained by any α-quantile q of X, the maximum in (4.1) by $\overline{\mathbb{Q}}$ defined by

$$\frac{\mathrm{d}\overline{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} := \frac{1}{\alpha}\mathbb{1}_{\{X<q\}} + c\mathbb{1}_{\{X=q\}},$$

where q is any α-quantile of X and $c \in [0,\infty)$ is such that $\overline{\mathbb{Q}}$ is a probability measure.

Proof. Since q_X has the same distribution under λ as X under $\mathbb{P}$, we have for $q = q_X(\alpha)$

$$\begin{aligned} AVaR^\alpha(X) &= -\frac{1}{\alpha}\int_0^\alpha q_X(u)\,\mathrm{d}u\\ &= \frac{1}{\alpha}\int_0^\alpha (q - q_X(u))\,\mathrm{d}u - q\\ &= \frac{1}{\alpha}\mathbb{E}_\lambda[(q-q_X)^+] - q\\ &= \frac{1}{\alpha}\mathbb{E}_\mathbb{P}[(q-X)^+] - q.\end{aligned}$$

For $\mathbb{Q} \in \mathcal{Q}^\alpha$ and $s \in \mathbb{R}$ we have

$$\mathbb{E}_\mathbb{Q}[-X] = \mathbb{E}_\mathbb{Q}[s-X] - s \leq \frac{1}{\alpha}\mathbb{E}_\mathbb{P}[(s-X)^+] - s.$$

On the other hand, for an α-quantile q of X, if

$$\frac{\mathrm{d}\overline{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} := \frac{1}{\alpha}\mathbb{1}_{\{X<q\}} + c\mathbb{1}_{\{X=q\}},$$

then

$$\mathbb{E}_{\bar{\mathbb{Q}}}[-X] = \mathbb{E}_{\bar{\mathbb{Q}}}[q - X] - q = \mathbb{E}_{\mathbb{P}}\left[\tfrac{1}{\alpha}\mathbb{1}_{\{X<q\}}(q - X)\right] - q = \frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}[(q - X)^+] - q.$$

Hence,

$$AVaR^{\alpha}(X) = \frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}[(q - X)^+] - q = \mathbb{E}_{\bar{\mathbb{Q}}}[-X] \geq \mathbb{E}_{\mathbb{Q}}[-X]$$

for every $\mathbb{Q} \in \mathcal{Q}^{\alpha}$ and

$$\mathbb{E}_{\bar{\mathbb{Q}}}[-X] \leq \frac{1}{\alpha}\mathbb{E}_{\mathbb{P}}[(s - X)^+] - s$$

for every $s \in \mathbb{R}$. □

4.2.2 Entropic risk measure

Definition 4.24. *For $X \in L^1(\mathbb{P})$ define the entropic risk measure at level $\alpha > 0$ by*

$$\rho(X) := \frac{1}{\alpha}\log \mathbb{E}_{\mathbb{P}}[\exp(-\alpha X)].$$

Remark 4.25. *The entropic risk measure satisfies (N), (M) and (T), but not (P).*

Theorem 4.26. *For some probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ define the entropy of $\mathbb{Q}$ relative to $\mathbb{P}$ by*

$$H(\mathbb{Q}|\mathbb{P}) := \begin{cases} \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}}{d\mathbb{P}}\right], & \mathbb{Q} \ll \mathbb{P}, \\ +\infty, & \textit{otherwise.} \end{cases}$$

Then for every $X \in L^{\infty}(\mathbb{P})$ we have

$$\rho(X) = \max_{\mathbb{Q} \ll \mathbb{P}} \left(\mathbb{E}_{\mathbb{Q}}[-X] - \tfrac{1}{\alpha}H(\mathbb{Q}|\mathbb{P})\right).$$

Proof. Define $\mathbb{P}_X$ by

$$\frac{d\mathbb{P}_X}{d\mathbb{P}} := \frac{e^{-\alpha X}}{\mathbb{E}_{\mathbb{P}}[e^{-\alpha X}]}.$$

Note that $x \log x \colon (0, \infty) \to \mathbb{R}$ is bounded from below so that the entropy of $\mathbb{Q}$ relative to $\mathbb{P}$ is well-defined. Since $x \log x \colon (0, \infty) \to \mathbb{R}$ is convex, Jensen's inequality yields

$$\mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}}{d\mathbb{P}}\right] = \mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}}\right] \geq \mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] \log \mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = 0$$

with equality if and only if $\mathbb{Q} = \mathbb{P}$. Therefore,

$$\begin{aligned} H(\mathbb{Q}|\mathbb{P}) &= \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}}{d\mathbb{P}}\right] \\ &= \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{Q}}{d\mathbb{P}_X}\right] + \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{P}_X}{d\mathbb{P}}\right] \\ &\geq \mathbb{E}_{\mathbb{Q}}\left[\log \frac{d\mathbb{P}_X}{d\mathbb{P}}\right] \\ &= \alpha\mathbb{E}_{\mathbb{Q}}[-X] - \log \mathbb{E}_{\mathbb{P}}[e^{-\alpha X}] \end{aligned}$$

so that

$$\frac{1}{\alpha}\log \mathbb{E}_{\mathbb{P}}\left[e^{-\alpha X}\right] \geq \mathbb{E}_{\mathbb{Q}}[-X] - \frac{1}{\alpha}H(\mathbb{Q}|\mathbb{P})$$

for every $\mathbb{Q} \ll \mathbb{P}$ with equality if and only if $\mathbb{Q} = \mathbb{P}_X$. □

4.3 Variance optimal hedging

We return to the very basic question of this lecture: what is a reasonable price of a financial derivative? In contrast to our previous approaches, we do not make the assumption that our market model is complete. In this case, it is in general no longer possible to perfectly hedge/replicate a financial derivative. Still, a reasonable approach could be to hedge against at least part (as much as possible) of the risk that comes with buying/selling a financial derivative.

The literature provides a number of suggestions how to interpret this task. We give here a brief introduction to the concept of variance-optimal hedging. The idea is to approximate the financial derivative as well as possible in the sense of L^2-distance. For simplicity, we return to the setup of the first chapter and assume in a addition that asset price $X = (X_t)_{t=0}^T$ is already a martingale under the measure $\mathbb{P}$. (This assumption is a bit rough, but not totally unreasonable from a practical perspective. And without this assumption, the results would be less neat.) In addition, we assume that the underlying probability space is finite. (Note that this not necessary, in fact it is rather simple to guess the appropriate L^2-assumptions one would need to make otherwise.)

Definition 4.27. *We call the pair (a, H) (where $a \in \mathbb{R}$ and H is a hedging strategy) the* variance optimal hedging strategy *for the derivative C if it minimizes the expected squared hedging error*

$$\mathbb{E}(C - a' - (H' \cdot X)_T)^2 \tag{4.3}$$

over all strategies (a', H').

It is imperative to reinterpret Definition (4.27) as the L^2-projection it really is: Write

$$S_{\mathbb{R}} := \{a + (H \cdot X)_T : a \in \mathbb{R}, H \text{ predictable}\}$$

for the space of all derivatives that can be exactly replicated. Then (a, H) is a variance optimal hedging strategy iff $a + (H \cdot X)_T$ is the $L^2(\mathbb{P})$-projection of the claim C.

While the L^2-projection is of course unique, there can be some ambiguity concerning (a, H) (think, e.g. of X being constant in time). We thus obtain only the following statement concerning existence and uniqueness:

Lemma 4.28. *Let C be a financial derivate. There exists a variance optimal hedging strategy. Assume that $(a, H), (a', H')$ are variance optimal hedging strategies. Then $(a, H), (a', H')$ have the same wealth process, i.e.*

$$a + (H \cdot X)_t = a' + (H' \cdot X)_t. \tag{4.4}$$

Specifically we have $a = a' = \mathbb{E}[C]$.

Proof. Existence and uniqueness of the minimizing element $a + (H \cdot X)_t \in S_{\mathbb{R}}$ follow from uniqueness and existence of the L^2-projection. (4.4) then follows by taking conditional expectations. □

The following martingale decomposition plays a crucial role for variance minimal hedging:

Theorem 4.29 (Galchouk-Kunita-Watanabe-decomposition)**.** *Assume that V is a martingale. Then there exists a decomposition*

$$V = V_0 + (H \cdot X)_T + M, \tag{4.5}$$

where H is predictable and M is a martingale which is orthogonal *to X in the sense that MX is a martingale.*

The martingales M and $(H \cdot X)$ in this decomposition are unique.

Proof. Denote by $Y = a + (H \cdot X)_T$ the orthogonal projection of V_T onto $S_{\mathbb{R}}$ and write U for the martingale generated by Y, i.e. $U_t := \mathbb{E}[Y|\mathcal{F}_t]$. Set $M := V - U$. As a difference of martingales, M is a martingale as well.

Since Y is the orthogonal projection, we have

$$\mathbb{E}(V_N - Y)(a' + (H' \cdot X)_T) = 0$$

for $a' \in \mathbb{R}$, H' predictable. Considering $a' = 1, H' \equiv 0$ we obtain that $\mathbb{E}(V_N - Y) = 0$, hence $V_0 = a$.

Let $t \in \{1, \dots, T\}$ and $A \in \mathcal{F}_t$. Define a predictable process H' by

$$H'_s = \begin{cases} I_A & \text{if } s \geq t \\ 0 & \text{else.} \end{cases}$$

Then we have

$$
\begin{aligned}
0 &= \mathbb{E}(V_T - Y)(H' \cdot X)_T && (4.6)\\
&= \mathbb{E}(M_T I_A (X_T - X_{t-1}) && (4.7)\\
&= \mathbb{E}(M_T I_A X_T) - \mathbb{E}(M_T I_A X_{t-1}) && (4.8)\\
&= \mathbb{E}(M_T I_A X_T) - \mathbb{E}(M_{t-1} I_A X_{t-1}). && (4.9)
\end{aligned}
$$

Hence MX is a martingale.

We have seen that $\mathbb{E}(M_T(H' \cdot X)_T) = 0$ for all predictable H' implies that MX is a martingale. Conversely it is easy to see that, if M is a martingale orthogonal to X, then $\mathbb{E}(M(H' \cdot X)_T) = 0$ for all predictable H'. From this it follows that (a, H) is the variance minimal hedge of V_T. In particular M_T is unique. □

As a consequence of the above proof we obtain:

Corollary 4.30. *Denote by*

$$V_t = V_0 + (H \cdot X)_t + M_t$$

the Galchouk-Kunita-Watanabe decomposition of the martingale $V_t := \mathbb{E}[C|\mathcal{F}_t]$. Then (V_0, H) is the variance minimal hedge of the derivative C.

5 Question you should be able to answer after completing this course

1. What is a model of a financial market? (discrete time, finite Ω)
2. What is a financial derivative?
3. What is an arbitrage opportunity (economically and mathematically)? (discrete time, finite Ω)
4. Why is “No Arbitrage” a reasonable assumption in math. finance?
5. How can we characterize martingales in terms of trading strategies? (discrete time, finite Ω
6. How do can one characterize absence of arbitrage in terms martingale measures? How would you prove it? (discrete time, finite Ω)
7. When is a financial derivative attainable? What does this tell us about its price?
8. How can we use martingales to characterize the attainable claims? (discrete time, finite Ω)

9. What is a complete market? What does this have to do with attainability of financial claims? (discrete time, finite Ω)
10. Why is not a sensible why of pricing a financial derivative to take the expectation of its payoff?
11. What could be a sensible pricing rule for general financial derivatives in an incomplete model? (discrete time, finite Ω).
12. Give an example of a martingale / a model with arbitrage / a model without arbitrage / a non complete model and financial derivative that can be replicated / etc.
13. What is a Brownian motion? Alternative definition?
14. What are the important steps in constructing a Brownian motion?
15. Are the paths of Brownian motion continuously differentiable?
16. What is a modification / version of a stochastic process? When are two stochastic processes indistinguishable?
17. What is quadratic variation?
18. Why can't we just consider a Riemann-Stiltjes integral wrt. BM?
19. What are ways of defining a stochastic integral wrt. BM / semi-martingales?
20. What is the Ito-Isometry? Why are we interested in it?
21. What is the Ito-formula? What is the main difference to the usual chain rule?
22. What is the martingale representation theorem?
23. What is the content Levy's characterization of Brownian motion?
24. What does Girsanov's theorem say?
25. What is the Bachelier model? Is it free of arbitrage? Does it admit an equivalent martingale measure? How would you charactrize that equivalent martingale measure? How do we price a financial derivative in the Bachelier model?
26. What is geometric Brownian motion?
27. Is it free of arbitrage? Does it admit an equivalent martingale measure? How would you charactrize that equivalent martingale measure? How do we price a financial derivative in the Bachelier model?

28. What are advantages of the Black-Scholes model over the Bachelier model?
29. What is problematic about the Black-Scholes model?
30. What is “put-call parity”?
31. What is “Black-Scholes-implied volatility”?

References

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

[2] F. Delbaen and W. Schachermayer. *The mathematics of arbitrage.* Springer Finance. Springer-Verlag, Berlin, 2006.

[3] H. Föllmer and A. Schied. *Stochastic finance.* De Gruyter Graduate. De Gruyter, Berlin, 2016. An introduction in discrete time, Fourth revised and extended edition of [MR1925197].