Introduction to Probability Theory and Stochastic Processes for Finance[*] Lecture Notes

Fabio Trojani

Department of Economics, University of St. Gallen, Switzerland

[*]Correspondence address: Fabio Trojani, Swiss Institute of Banking and Finance, University of St. Gallen, Rosenbergstr. 52, CH-9000 St. Gallen, e-mail: Fabio.Trojani@unisg.ch.

Contents

1 Introduction to Probability Theory **4**

1.1 The Binomial Model 4

1.1.1 The Risky Asset 4

1.1.2 The Riskless Asset 4

1.1.3 A Basic No Arbitrage Condition 5

1.1.4 Some Basic Remarks 5

1.1.5 Pricing Derivatives: a first Example 5

1.2 Finite Probability Spaces 7

1.2.1 Measurable Spaces 7

1.2.2 Probability measures 11

1.2.3 Random Variables 14

1.2.4 Expected Value of Random Variables Defined on Finite Measurable Spaces 15

1.2.5 Examples of Probability Spaces and Random Variables with Finite Sample Space 16

1.3 General Probability Spaces 17

1.3.1 Some First Examples of Probability Spaces with non finite Sample Spaces . 18

1.3.2 Continuity Properties of Probability Measures 20

1.3.3 Random Variables 21

1.3.4 Expected Value and Lebesgue Integral 25

1.3.5 Some Further Examples of Probability Spaces with uncountable Sample Spaces 28

1.4 Stochastic Independence 29

2 Conditional Expectations and Martingales **33**

2.1 The Binomial Model Once More 33

2.2 Sub Sigma Algebras and (Partial) Information 34

2.3 Conditional Expectations 36
2.3.1 Motivation 36
2.3.2 Definition and Properties 37
2.4 Martingale Processes 41

3 Pricing Principles in the Absence of Arbitrage **44**
3.1 Stock Prices, Risk Neutral Probability Measures and Martingales 45
3.2 Self Financing Strategies, Risk Neutral Probability Measures and Martingales . . . 46
3.3 Existence of Risk Neutral Probability Measures and Derivatives Pricing 48
3.4 Uniqueness of Risk Neutral Probability Measures and Derivatives Hedging 50
3.5 Existence of Risk Neutral Probability Measures and Absence of Arbitrage 52

4 Introduction to Stochastic Processes **52**
4.1 Basic Definitions 52
4.2 Discrete Time Brownian Motion 54
4.3 Girsanov Theorem: Application to a Semicontinuous Pricing Model 57
4.3.1 A Semicontinuous Pricing Model 57
4.3.2 Risk Neutral Valuation in the Semicontinuous Model 58
4.3.3 A Discrete Time Formulation of Girsanov Theorem 60
4.3.4 A Discrete Time Derivation of Black and Scholes Formula 64
4.4 Continuous Time Brownian Motion 66

5 Introduction to Stochastic Calculus **71**
5.1 Starting Point, Motivation 71
5.2 The Stochastic Integral 73
5.2.1 Some Basic Preliminaries 74
5.2.2 Simple Integrands 75

5.2.3 Squared Integrable Integrands 81
5.2.4 Properties of Stochastic Integrals 84
5.3 Itô's Lemma 85
5.3.1 Starting Point, Motivation and Some First Examples 85
5.3.2 A Simplified Derivation of Itô's Formula 88
5.4 An Application of Stochastic Calculus: the Black-Scholes Model 93
5.4.1 The Black-Scholes Market 93
5.4.2 Self Financing Portfolios and Hedging in the Black-Scholes Model 93
5.4.3 Probabilistic Interpretation of Black-Scholes Prices: Girsanov Theorem once more 95

1 Introduction to Probability Theory

1.1 The Binomial Model

We start with the binomial model to introduce some basic ideas of probability theory related to the pricing of contingent claims, basically for the following reasons:

- It is a simple setting where the arbitrage concept and its relation to risk neutral pricing can be explained
- It is a model used in practice where binomial trees are calibrated to real data, for instance to price American derivatives
- It is a simple setting to introduce the concept of conditional expectations and martingales, which are at the hearth of the theory of derivatives pricing.

1.1.1 The Risky Asset

S_t is the price of a risky stock at time $t \in I$, where we start for simplicity with a discrete time index $I = \{0, 1, 2\}$. The dynamics of S_t is defined by

$$S_t = \begin{cases} uS_{t-1} & \text{with probability } p \\ dS_{t-1} & \text{with probability 1-}p \end{cases} ,$$

where $p \in (0, 1)$. We impose for brevity the further condition

$$u = \frac{1}{d} > 1$$

giving a recombining tree.

1.1.2 The Riskless Asset

B_t is the price at time t of a riskless money account. $r > 0$ is the riskless interest rate on the money account, implying

$$B_t = (1 + r)\, B_{t-1}$$

for any $t = 1, 2$. For simplicity we impose the normalization $B_0 = 1$.

1.1.3 A Basic No Arbitrage Condition

A necessary condition for the absence of arbitrage opportunities in our model is

$$d < 1 + r < u \quad . \tag{1}$$

Example 1 *In the sequel we will often use a numerical example with parameters* $S_0 = 4$*,* $u = 1/d = 2$*,* $r = 0.25$*.*

1.1.4 Some Basic Remarks

Notice that to any trajectory TT, TH, HT, HH, in the tree we can associate the corresponding values of S_1 and S_2. Thus, from the perspective of time 0, both S_1 and S_2 are random entities whose value depends on which event/trajectory will be realized in the model. To fully describe the random behaviour of S_1 and S_2 we can make use of the space $\Omega = \{TT, TH, HT, HH\}$ of all random sequences that can be realized on the tree. Basically, Ω contains all the information about the single outcomes that can be realized in our model.

Definition 2 *(i) The set* Ω *of all possible outcomes in a random experiment is called the sample space. (ii) Each single event* $\omega \in \Omega$ *is called an outcome of the random experiment.*

Example 3 *In the above two period model we had* $\Omega = \{TT, TH, HT, HH\}$ *and* $\omega = TT$ *or* $\omega = TH$ *or* $\omega = HT$ *or* $\omega = HH$*.*

Exercise 4 *Give the sample space and all single outcomes in a binomial tree with three periods.*

1.1.5 Pricing Derivatives: a first Example

Definition 5 *An European call option with strike price* K *and maturity* $T \in I$ *is the right to buy at time* T *the underlying stock for the price* K*. We denote by* c_t *the price of the European call option at time* t*.*

From the definition we immediately have for the pay-off at maturity of the call option:

$$c_T = \begin{cases} S_T - K & S_T > K \\ 0 & S_T \leq K \end{cases} \quad ,$$

or, more compactly:

$$c_T = (S_T - K)^+ \quad ,$$

where $(x)^+ := \max(x, 0)$ is the positive part of x.

Remark 6 *Notice that c_T depends on $\omega \in \Omega$ only through $S_T(\omega)$. The goal in any pricing model is to determined the time 0 price (as for instance the price c_0) of a derivative payoff falling at a later time T, say (as for instance the pay-off $c_T = (S_T - K)^+$).*

Assumption 7 *To illustrate the main ideas we start with $T = 1$.*

Definition 8 *A (perfect) hedging portfolio for c_T with value V_0 at time 0 is a position in Δ_0 stock and $V_0 - \Delta_0 S_0$ money accounts (recall the normalization $B_0 = 1$), such that*

$$\begin{cases} c_1(H) = \Delta_0 S_1(H) + (V_0 - \Delta_0 S_0)(1 + r) \\ c_1(T) = \Delta_0 S_1(T) + (V_0 - \Delta_0 S_0)(1 + r) \end{cases} \tag{2}$$

Remark 9 *A (perfect) hedging portfolio replicates exactly the future pay-off of the derivative to be hedged. Therefore, it is a vehicle to fully eliminate the risk intrinsic in the randomness of the future value of a derivative.*

Proposition 10 *(i) For $T = 1$, the quantity Δ_0 is given by*

$$\Delta_0 = \frac{c_1(H) - c_1(T)}{S_1(H) - S_1(T)} \quad . \tag{3}$$

Δ_0 is called the "delta" of the hedging portfolio. (ii) The risk neutral valuation formula follows:

$$c_0 = V_0 = \frac{1}{1 + r} \left[\widetilde{p} c_1(H) + (1 - \widetilde{p}) c_1(T) \right] \quad ,$$

where

$$\widetilde{p} = \frac{1 + r - d}{u - d} \quad .$$

Proof. (i) Compute the difference between the first and the second equation in (2) and solve for Δ_0. (ii) Insert Δ_0 given by (3) in one of the two equations in (2) and solve for V_0. Absence of arbitrage then implies $V_0 = c_0$. ■

Remark 11 *(i) The price $V_0 = c_0$ does not depend on the binomial probability p. (ii) Under the given conditions (cf. (1)) one has $\widetilde{p} \in (0,1)$. Therefore the identity*

$$c_0 = \frac{1}{1+r}\left[\widetilde{p} c_1(H) + (1-\widetilde{p})\, c_1(T)\right]$$

says that the price c_0 is a discounted expectation of the call future random pay-offs, computed using the risk adjusted probabilities $\widetilde{p}$ and $(1-\widetilde{p})$. More compactly, we could thus write

$$c_0 = \frac{\widetilde{E}(c_1)}{1+r} \quad ,$$

where $\widetilde{E}$ denotes expectations under $\widetilde{p}$, $1-\widetilde{p}$. This is a so called risk adjusted (or risk neutral) valuation formula.

Exercise 12 *(i) For the case $T = 1$ and for the model parameters in Example 1 compute the numerical value of c_0. (ii) For the case $T = 2$ compute recursively the hedging portfolio of the derivative, starting from $\Delta_1(H)$, $\Delta_1(T)$, $V_1(H)$, $V_1(T)$, and finishing with Δ_0 and V_0.*

1.2 Finite Probability Spaces

In the sequel we let $\Omega \neq \emptyset$ be a given sample space.

1.2.1 Measurable Spaces

Let $\mathcal{F}$ be the family of all subsets of Ω; $\mathcal{F}$ is an example of a so called sigma algebra, a concept that we define in the sequel.

Definition 13 *(i) A sigma algebra $\mathcal{G} \subset \mathcal{F}$ is a family of subsets of Ω such that:*

1. $\emptyset \in \mathcal{G}$

2. If $A \in \mathcal{G}$ then it follows $A^c \in \mathcal{G}$

3. If $(A_i)_{i \in \mathbb{N}} \subset \mathcal{G}$ is a countable sequence in $\mathcal{G}$, then it follows

$$\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{G}$$

(ii) The couple $(\Omega, \mathcal{G})$ is called a measurable space.

Example 14 *(i) $\mathcal{F}$ is a sigma algebra, the finest one on Ω. Indeed, $\emptyset \in \mathcal{F}$. Moreover, for any set $A \in \mathcal{F}$ the complement A^c is a subset of Ω, i.e. is in $\mathcal{F}$. The same holds for any (not only for a countable) union of sets in $\mathcal{F}$. (ii) The subfamily $\mathcal{G} := \{\emptyset, \Omega\}$ is the coarsest sigma algebra on Ω. (iii) In the setting of the binomial model of Example 1, it is easy to verify (please do it!) that the subfamily*

$$\mathcal{G} := \{\emptyset, \Omega, \{HT, HH\}, \{TT, TH\}\} \quad ,$$

is a sigma algebra, the sigma algebra generated by the first period price movements in the model.

Remark 15 *We make use of sigma algebras to model different information sets at the disposal of the investor in doing her portfolio choices. For instance, in the setting of the binomial model of Example 1, the information available at time 0 (before observing prices) can be modelled by the trivial information set*

$$\mathcal{G}_0 := \{\emptyset, \Omega\} \quad .$$

That is, at time 0 investors only know that the possible realized outcome ω has to be an element of the sample space Ω. At time 1 investors can observe S_1. Thus, depending on the value of S_1 they will know at time 1 that either

$$\omega \in \{HT, HH\} \quad \text{(if and only if } S_1(\omega) = S_0 u\text{)} \quad ,$$

or

$$\omega \in \{TT, TH\} \quad \text{(if and only if } S_1(\omega) = S_0 d\text{)} \quad .$$

Thus at time 1 investors do not have full information about ω, since they still do not know the direction of the price movement in period 2. However, they can determine to which specific event of their information set ω belongs. The larger (smaller) this set, the preciser (the rougher) the information on the realized outcome ω. For instance, while at time 0 investors only know that the outcome will be an element of the sample space, at time 1 they know that the outcome implies either an upward or a downward price movement in the first period. Based on these considerations a natural sigma algebra $\mathcal{G}_1$ to model investors price information at time 1 is

$$\mathcal{G}_1 := \{\emptyset, \Omega, \{HT, HH\}, \{TT, TH\}\} \quad ,$$

(verify that $\mathcal{G}_1$ is indeed a sigma algebra). Similarly, by observing only the price S_2 investors will know at time 2 that either

$$\omega = HH \quad \textit{(if and only if } S_2(\omega) = S_0 u^2\textit{)} \quad ,$$

or

$$\omega = TT \quad \textit{(if and only if } S_2(\omega) = S_0 d^2\textit{)} \quad ,$$

or

$$\omega \in \{TH, HT\} \quad \textit{(if and only if } S_2(\omega) = S_0 du\textit{)} \quad .$$

On the other hand, by observing the prices S_1 and S_2 investors will know at time 2

$$\omega = HH \quad \textit{(if and only if } S_2(\omega) = S_0 u^2\textit{)} \quad ,$$

or

$$\omega = TT \quad \textit{(if and only if } S_2(\omega) = S_0 d^2\textit{)} \quad ,$$

or

$$\omega = TH \quad \textit{(if and only if } S_1(\omega) = S_0 d \textit{ and } S_2(\omega) = S_0 du\textit{)} \quad ,$$

or

$$\omega = HT \quad \textit{(if and only if } S_1(\omega) = S_0 u \textit{ and } S_2(\omega) = S_0 du\textit{)} \quad .$$

Based on these considerations a natural sigma algebra $\mathcal{G}_2$ to model investors price information up to time 2 is the smallest one containing the system of subsets of Ω given by

$$\mathcal{E}_2 := \{\emptyset, \Omega, \{HT\}, \{HH\}, \{TT\}, \{TH\}\} \quad .$$

We denote this sigma algebra by $\mathcal{G}_2 = \sigma(\mathcal{E}_2)$. Finally, the sigma algebra representing the information obtained by observing only the price S_2 is

$$\mathcal{G}_3 = \{\emptyset, \Omega, \{HH\}, \{TH, HT, TT\}, \{TT\}, \{TH, HT, HH\}, \{TH, HT\}, \{TT, HH\}\}$$

Notice that while the relation $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{G}_2$ implies an information set growing over time, we do not have $\mathcal{G}_1 \subset \mathcal{G}_3$ (why?). Therefore, the sequence of sigma algebras $\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_3$ is not consistent with the idea of an investor's information set growing over time.

Exercise 16 *(Borel sigma algebra on $\mathbb{R}$) Let $\Omega := \mathbb{R}$ and denote by $\mathcal{T}$ the set of all open intervals in $\mathbb{R}$*

$$\mathcal{T} = \{(a, b) \quad | \quad a \leq b,\ a, b \in \mathbb{R}\} \quad .$$

1. *Show with a simple counterexample that $\mathcal{T}$ is not a sigma algebra on $\mathbb{R}$.*

2. *We know that there does exist a sigma algebra over $\mathbb{R}$ containing $\mathcal{T}$ (which one?). Thus, there also exists a "minimal sigma algebra" containing $\mathcal{T}$, the so-called Borel sigma algebra over $\mathbb{R}$ (denoted by $\mathcal{B}(\mathbb{R})$) which has to be of the form*

$$\mathcal{B}(\mathbb{R}) = \bigcap_{\substack{\mathcal{G} \text{ is } \sigma-algebra \text{ over } \mathbb{R} \\ \mathcal{T} \subset \mathcal{G}}} \mathcal{G}$$

   To show that $\mathcal{B}(\mathbb{R})$ is indeed a sigma algebra over $\mathbb{R}$ it is thus sufficient to show that intersections of sigma algebras are sigma algebras. Do this, by verifying the corresponding definition.

3. *Show, using simple set operations, that the events $(-\infty, a)$, (a, ∞), $[a, b]$, $(a, b]$, $\{a\}$, where $a \leq b$, are elements of $\mathcal{B}(\mathbb{R})$.*

4. Show that any countables subset $\{a_i\}_{i\in\mathbb{N}}$ *of* $\mathbb{R}$ *is an element of* $\mathcal{B}(\mathbb{R})$.

As mentioned, a natural way to model a growing amount of information over time is through increasing sequences of sigma algebras. This is the next definition.

Definition 17 *Let* $(\Omega, \mathcal{G})$ *be a measurable space. A sequence* $(\mathcal{G}_i)_{i=0,1,\dots,n}$ *of sigma algebras over* Ω *such that*

$$\mathcal{G}_0 \subset \mathcal{G}_1 \subset \dots \subset \mathcal{G}_n \subset \mathcal{G} \quad ,$$

is called a filtration.

Example 18 *In Remark 15 the sequence* $(\mathcal{G}_i)_{i=0,1,2}$ *is a filtration, while the sequence* $(\mathcal{G}_i)_{i=0,1,3}$ *is not.*

1.2.2 Probability measures

For the whole section let $(\Omega, \mathcal{G})$ be a measurable space.

Definition 19 *We say that an event* $A \in \mathcal{G}$ *is realized in a random experiment with sample space* Ω *if* $\omega \in A$.

Example 20 *In the two period binomial model we have*

$$\{TH, TT\} = \{\textit{The stock price drops in the first period}\} \quad .$$

Thus, if a time 1 *we observe* T, $\{TH, TT\}$ *is realized. On the other hand, if we observe* H, *then* $\{TH, TT\}$ *is not realized (i.e.* $A^c = \{HT, HH\}$ *is realized).*

The next step is to assign in a consistent way probabilities to events that can be realized in a random experiment.

Definition 21 *(i) A probability measure on* $(\Omega, \mathcal{G})$ *is a function* $P : \mathcal{G} \rightarrow [0,1]$ *such that:*

1. $P(\Omega) = 1$

2. *For any disjoint sequence* $(A_i)_{i \in \mathbb{N}} \subset \mathcal{G}$ *such that* $A_i \cap A_j = \emptyset$ *for* $i \neq j$ *it follows*

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i) \quad .$$

This property is called sigma additivity.

(ii) We call a triplet $(\Omega, \mathcal{G}, P)$ *a probability space.*

Example 22 *In the two period binomial model we set* $\Omega = \{TT, TH, HT, HH\}$, $\mathcal{G} = \mathcal{F}$, *and define probabilities with the binomial rule*

$$P(HH) = p^2 \quad , \quad P(TT) = (1-p)^2 \quad , \quad P(TH) = P(HT) = p(1-p) \quad .$$

The sigma additivity then implies, for instance

$$P(HT, HH) = P(HH) + P(HT) = p^2 + p(1-p) \quad .$$

More generally, we have, in this finite sample space setting:

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Proposition 23 *Let* $(\Omega, \mathcal{G}, P)$ *be a probability space. We have:*

1. $P(A \backslash B) = P(A) - P(A \cap B)$

2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

3. $P(A^c) = 1 - P(A)$

4. *If* $A \subset B$ *then* $P(A) \leq P(B)$

Proof. 1. $A \backslash B = A \cap B^c$ and $A = (A \cap B) \cup (A \cap B^c)$. By sigma additivity if follows:

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A \cap B) + P(A \backslash B) \quad .$$

2. $A \cup B = (A \backslash B) \cup B$. Therefore, using 1 and by sigma additivity:

$$P(A \cup B) = P(A \backslash B) + P(B) = P(A) + P(B) - P(A \cap B) \quad .$$

3. This is a particular case of 1. with $A = \Omega$ and $B = A$. 4. By 1. we have, under the given assumption:

$$P(B) = P(B \cap A) + P(B \backslash A) = P(A) + P(B \backslash A) \geq P(A) \quad .$$

■

Remark 24 *In Definition 21, the condition 1. for a probability measure implies the condition,*

1'. $P(\emptyset) = 0$.

In fact, a function $\mu : \mathcal{G} \to [0, \infty]$ *satisfying condition 1'. and 2. in Definition 21 is called a measure on the measurable space* $(\Omega, \mathcal{G})$*. Notice, that in this case we can have* $\mu(\Omega) = \infty$.

Exercise 25 *The Lebesgue measure on the measurable space* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *(denoted by* μ_0*) is a measure* $\mu_0 : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ *such that*

$$\mu_0((a, b)) = b - a$$

for any open interval (a, b)*,* $a \leq b$*. It can be shown that Lebesgue measure exists and is unique (we will not prove this, we will just assume it in the sequel). Show the following properties of Lebesgue measure, using the general definition of a measure.*

1. $\mu_0(\emptyset) = 0$, $\mu_0(\mathbb{R}^+) = \infty$
2. $\mu_0(\{a\}) = 0$ *for any* $a \in \mathbb{R}$
3. *For any countable subset* $\{a_i\}_{i \in \mathbb{N}}$ *of* $\mathbb{R}$ *one has* $\mu_0\left(\{a_i\}_{i \in \mathbb{N}}\right) = 0$.

1.2.3 Random Variables

For the whole section let $(\Omega, \mathcal{G})$ be a measurable space such that the cardinality of Ω is finite ($|\Omega| < \infty$). We will extend the concept of a random variable to non finite sample spaces in a later section.

Definition 26 *Let $X : \Omega \to \mathbb{R}$ be a function from Ω to the real line. (i) The sigma algebra*

$$\sigma(X) := \left\{X^{-1}(B) : B \text{ is a subset of } \mathbb{R}\right\} \quad ,$$

where $X^{-1}(B)$ is a short notation for the preimage $\{\omega : X(\omega) \in B\}$ of B under X, is called the sigma algebra generated by X. (ii) X is called a random variable on $(\Omega, \mathcal{G})$ if it is measurable with respect to $\mathcal{G}$, that is if

$$\sigma(X) \subset \mathcal{G} \quad .$$

Remark 27 *(i) It is useful to know some properties of preimages. We have for any subset B of $\mathbb{R}$, and for any (non necessarily countable) sequence $(B_\alpha)_{\alpha \in \mathcal{A}}$ of subsets of $\mathbb{R}$:*

$$\begin{aligned} X^{-1}(B^c) &= \left(X^{-1}(B)\right)^c \\ X^{-1}\left(\bigcup_{\alpha \in \mathcal{A}} B_\alpha\right) &= \bigcup_{\alpha \in \mathcal{A}} X^{-1}(B_\alpha) \\ X^{-1}\left(\bigcap_{\alpha \in \mathcal{A}} B_\alpha\right) &= \bigcap_{\alpha \in \mathcal{A}} X^{-1}(B_\alpha) \end{aligned}$$

(ii) $\sigma(X)$ is a sigma algebra. Indeed, $\emptyset = X^{-1}(\emptyset) \in \sigma(X)$. Moreover, if $A = X^{-1}(B)$ for some subset of $\mathbb{R}$, then

$$A^c = \left(X^{-1}(B)\right)^c = X^{-1}(B^c) \in \sigma(X) \quad ,$$

because B^c is a subset of $\mathbb{R}$. Similarly, given a sequence $(A_i)_{i \in \mathbb{N}}$ such that $A_i = X^{-1}(B_i)$ for a sequence of subsets $(B_i)_{i \in \mathbb{N}}$ of $\mathbb{R}$ we have:

$$\bigcup_{i \in \mathbb{N}} A_i = \bigcup_{i \in \mathbb{N}} X^{-1}(B_i) = X^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right) \in \sigma(X) \quad ,$$

because $\bigcup_{i \in \mathbb{N}} B_i$ is a subset of $\mathbb{R}$. (iii) $\sigma(X)$ represents the (partial) information set that is available about an outcome $\omega \in \Omega$ by observing the values of X.

Example 28 *In the two period binomial model* S_0*,* S_1 *and* S_2 *are all (trivially) measurable with respect to the finest sigma algebra* $\mathcal{F}$ *over* Ω*. However, since* S_0 *is constant we have*

$$\sigma(S_0) = \{\emptyset, \Omega\} = \mathcal{G}_0 \quad ,$$

and S_0 *is* $\mathcal{G}_0$ *measurable. Further,*

$$\sigma(S_1) = \{\emptyset, \Omega, \{HT, HH\}, \{TT, TH\}\} = \mathcal{G}_1 \quad ,$$

and S_1 *is* $\mathcal{G}_1$ *but not* $\mathcal{G}_0$ *measurable. Finally,*

$$\begin{aligned} \sigma(S_2) &= \{\emptyset, \Omega, \{HH\}, \{TH, HT, TT\}, \{TT\}, \{TH, HT, HH\}, \{TH, HT\}, \{TT, HH\}\} \\ &= \mathcal{G}_3 \quad . \end{aligned}$$

Therefore, S_2 *is* $\mathcal{G}_3$ *but not* $\mathcal{G}_1$ *measurable. On the other hand,* S_1 *is* $\mathcal{G}_1$ *but not* $\mathcal{G}_3$ *measurable (why?).*

1.2.4 Expected Value of Random Variables Defined on Finite Measurable Spaces

For the whole section let $(\Omega, \mathcal{G}, P)$ be a probability space such that the cardinality of Ω is finite ($|\Omega| < \infty$). We will extend the concept of expected value of a random variable to the non finite sample space setting in a later section. Further, let $X : (\Omega, \mathcal{G}) \to \mathbb{R}$ be a random variable.

Definition 29 *(i) The expected value* $E(X)$ *of a random variable* X *defined on a finite sample space is given by*

$$E(X) := \sum_{\omega \in \Omega} X(\omega) P(\omega) \quad .$$

(ii) The variance $Var(X)$ *of* X *is given by*

$$Var(X) := E\left[(X - E(X))^2\right] = E\left(X^2\right) - (E(X))^2 \quad .$$

Example 30 *In the two period binomial model of Example 1 we have:*

$$\begin{aligned} S_2(HH) &= 16 \quad ; \quad P(HH) = p^2 \\ S_2(HT) &= S_2(TH) = 4 \quad ; \quad P(TH) = P(HT) = p(1-p) \\ S_2(TT) &= 1 \quad ; \quad P(HH) = (1-p)^2 \end{aligned}$$

Therefore,

$$E(S_2) = 16 \cdot p^2 + 4 \cdot 2 \cdot p(1-p) + 1 \cdot (1-p)^2$$

1.2.5 Examples of Probability Spaces and Random Variables with Finite Sample Space

Example 31 *The Bernoulli distribution with parameter p is a probability measure P on the measurable space $(\Omega, \mathcal{G})$ given by $\Omega := \{0, 1\}$, $\mathcal{G} := \mathcal{F}$, such that:*

$$P(1) = p \in (0, 1) \quad .$$

Example 32 *The Binomial distribution with parameters n and p is a probability measure P on a measurable space $(\Omega, \mathcal{G})$ given below. The sample space is given by*

$$\Omega := \{n - \textit{dimensional sequences with components } 0 \textit{ or } 1\} \quad .$$

For instance, a possible element of Ω is

$$\omega = \underbrace{0010100...1111}_{n\ components} \quad .$$

Further, we set $\mathcal{G} := \mathcal{F}$. Finally, P is given by

$$P(\omega) = p^{\#\ of\ 1\ in\ \omega}\ (1-p)^{\#\ of\ 0\ in\ \omega} \quad .$$

For instance, using the properties of a probability measure we have:

$$\begin{aligned} P(\textit{at least a 1 over the } n \textit{ components}) &= 1 - P(\textit{no 1 over the } n \textit{ components}) \\ &= 1 - (1-p)^n \quad , \end{aligned}$$

and so forth.

Example 33 *A discrete uniform distribution modelling the toss of a fair die is obtained by setting $\Omega := \{1,2,3,4,5,6\}$, $\mathcal{G} := \mathcal{F}$, and*

$$P(\omega) = \frac{1}{6} \quad , \quad \omega \in \Omega \quad .$$

For instance, using the properties of a probability measure we then have:

$$P(\text{obtaining an even number}) = P(2) + P(4) + P(6) = \frac{1}{2} \quad ,$$

and so forth.

Example 34 *A discrete uniform distribution modelling the toss of two independent fair dies is obtained by setting*

$$\Omega := \{11, 12, 13, 14, 15, 16, 21, 22, ..., 66\}$$

$\mathcal{G} := \mathcal{F}$, and

$$P(\omega) = \frac{1}{36} \quad , \quad \omega \in \Omega \quad .$$

For instance, using the properties of a probability measure we then have:

$$P(\text{the sum of the two numbers is larger than 10}) = P(66) + P(56) + P(65) = \frac{1}{12} \quad ,$$

and so forth. Let $X : \Omega \to \{2,3,4,..,12\}$ be the function giving the sum of the numbers on the two dies. We have:

$$\sigma(X) = \left\{ \emptyset, \Omega, \underbrace{\{11\}}_{X^{-1}(2)}, \underbrace{\{12,21\}}_{X^{-1}(3)}, \underbrace{\{13,31,22\}}_{X^{-1}(4)}, ... \right\} \subset \mathcal{F} \quad ,$$

that is X is a random variable on $(\Omega, \mathcal{F})$.

1.3 General Probability Spaces

Definition 21 of a probability space does not require the assumption $|\Omega| < \infty$.

1.3.1 Some First Examples of Probability Spaces with non finite Sample Spaces

A first simple example of a probability space defined on a non finite sample space is the following.

Example 35 *Let* $\Omega = \mathbb{R}, \mathcal{G} = \mathcal{B}(\mathbb{R})$ *and define*

$$P(A) = \mu_0 (A \cap [0,1]) \quad .$$

P *is a probability measure, the uniform distribution on the interval* $[0,1]$*. Indeed, we have:*

1. $P(\Omega) = \mu_0 (\Omega \cap [0,1]) = \mu_0 ([0,1]) = 1$.

2. *For any disjoint sequence* $(A_i)_{i\in\mathbb{N}} \subset \mathcal{B}(\mathbb{R})$ *it follows*

$$P(\cup_{i\in\mathbb{N}} A_i) = \mu_0 ((\cup_{i\in\mathbb{N}} A_i) \cap [0,1]) = \mu_0 (\cup_{i\in\mathbb{N}} (A_i \cap [0,1])) = \sum_{i\in\mathbb{N}} \mu_0 (A_i \cap [0,1]) = \sum_{i\in\mathbb{N}} P(A_i)$$

More generally, setting

$$P(A) = \frac{\mu_0 (A \cap [a,b])}{\mu_0 ([a,b])} \quad ,$$

defines a uniform distribution on the interval $[a,b]$*.*

A famous example of a probability space with non finite sample space is the one underlying a Poisson distribution on $\mathbb{N}$.

Example 36 *Let* $\Omega := \mathbb{N}$ *and* $\mathcal{G} := \mathcal{F}$ *. Thus in this case* Ω *is an infinite, countable, sample space. We define for any* $\omega \in \Omega$

$$P(\omega) := \frac{\lambda^\omega}{\omega!} e^{-\lambda} \quad , \quad \lambda > 0 \quad .$$

Setting for $A \in \mathcal{F}$

$$P(A) := \sum_{\omega\in A} P(\omega) \quad ,$$

one obtains the Poisson distribution on $(\mathbb{N}, \mathcal{F})$ *with parameter* λ*.* P *is a probability measure on* $(\Omega, \mathcal{F})$*. Indeed, we have*

$$P(\Omega) = \sum_{\omega\in\Omega} P(\omega) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1 \quad ,$$

and, for any disjoint sequence $(A_i)_{i\in\mathbb{N}} \subset \mathcal{F}$,

$$P\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{\omega\in\bigcup_{i\in\mathbb{N}} A_i} P(\omega) = \sum_{i\in\mathbb{N}}\left(\sum_{\omega\in A_i} P(\omega)\right) = \sum_{i\in\mathbb{N}} P(A_i) \quad .$$

The last example of a probability space with non finite sample space that we present is the one underlying a Binomial experiment where $n \to \infty$.

Example 37 *Let* $\Omega := \{T, H\}^{\infty}$ *be the space of infinite sequences with components* T *or* H*. Thus any outcome* $\omega \in \Omega$ *is of the form*

$$\omega = (\omega_i)_{i\in\mathbb{N}} \quad , \quad \omega_i \in \{T, H\} \quad .$$

This is an infinite, uncountable, sample space. Therefore, some caution is needed in constructing a suitable sigma algebra on Ω*, on which we are enabled in a second step to extend the binomial distribution in a consistent way. We define*

$$\mathcal{G}_n := \{\textit{The sigma algebra generated by the first } n \textit{ tosses}\} \quad ,$$

for any $n \in \mathbb{N}$*. For instance, we obtain for* $\mathcal{G}_1$*:*

$$\mathcal{G}_1 = \{\emptyset, \Omega, \{\omega \in \Omega : \omega_1 = T\}, \{\omega \in \Omega : \omega_1 = H\}\} \quad ,$$

and so on for $n > 1$*. We know that there is a sigma algebra* $\mathcal{F}$ *over* Ω *such that* $\mathcal{G}_n \subset \mathcal{F}$ *for all* $n \in \mathbb{N}$*. However, this sigma algebra is too large to assign binomial probabilities on it in a consistent way. Therefore, we work in the sequel with the smallest sigma algebra containing all* $\mathcal{G}_n$*'s. We define*

$$\mathcal{G} := \bigcap_{\substack{\mathcal{H}\supset\cup_{n\in\mathbb{N}}\mathcal{G}_n \\ \mathcal{H} \textit{ is sigma algebra over } \Omega}} \mathcal{H} \quad ,$$

the sigma algebra generated by $\cup_{n\in\mathbb{N}}\mathcal{G}_n$*. Notice that* $\mathcal{G}$ *contains events that can be quite rich and that do not belong to any* $\mathcal{G}_n$*,* $n \in \mathbb{N}$*. An example of such an event is*

$$A := \{H \textit{ on every toss}\} = \{\omega \in \Omega : \omega_i = H \textit{ for all } i \in \mathbb{N}\} = \bigcap_{n\in\mathbb{N}} \underbrace{\{\omega \in \Omega : \omega_i = H \textit{ for } i \leq n\}}_{\in\mathcal{G}_n} \in \mathcal{G} \quad ,$$

where

$$\{\omega \in \Omega : \omega_i = H \text{ for } i \leq n\} = \{H \text{ on the first } n \text{ tosses}\} \quad .$$

We now define a probability measure P *on* $\mathcal{G}$ *whose restriction on any* $\mathcal{G}_n$ *is a binomial distribution with parameters* n *and* p*. Precisely, define for any* $A \in \mathcal{G}_n$ *and some given* $n \in \mathbb{N}$

$$P(A) = p^{\#\text{ of } H \text{ in the first } n \text{ tosses}} (1-p)^{\#\text{ of } T \text{ in the first } n \text{ tosses}} \quad .$$

For instance, for the event

$$\{H \text{ on the first } 2 \text{ tosses}\} = \{\omega \in \Omega : \omega_i = H \text{ for } i \leq 2\} \quad ,$$

we obtain

$$P(H \text{ on the first } 2 \text{ tosses}) = p^2 \quad ,$$

and so forth. Using the properties of a probability measure we can then uniquely extend P *to all of* $\mathcal{G}$*. For instance, we have*

$$P(H \text{ on all tosses}) \leq P(H \text{ on the first } n \text{ tosses}) = p^n \quad ,$$

for all $n \in \mathbb{N}$*. Therefore, for* $p \in (0,1)$ *it follows*

$$P(H \text{ on all tosses}) = 0 \quad .$$

1.3.2 Continuity Properties of Probability Measures

Two further continuity properties of a probability measure - in excess of the properties in Proposition 23 - are useful when working with countable set operations over monotone sequences of events. They are given below.

Proposition 38 *Let* $(A_n)_{n\in\mathbb{N}} \subset \mathcal{G}$ *be a countable sequence of events. It then follows:*

1. *If* $A_1 \subset A_2 \subset \ldots$*, then:*

$$P(A_n) \underset{n\to\infty}{\uparrow} P\left(\bigcup_{n\in\mathbb{N}} A_n\right) \quad ,$$

   (continuity from below).

2. If $A_1 \supset A_2 \supset ...$, then:

$$P(A_n) \underset{n\to\infty}{\downarrow} P\left(\bigcap_{n\in\mathbb{N}} A_n\right)$$

(continuity from above).

Proof. 1. Let $A := \bigcup_{n\in\mathbb{N}} A_n$. We have,

$$A = \bigcup_{n\in\mathbb{N}} (A_n \backslash A_{n-1}) \quad ,$$

where $A_0 := \emptyset$. Thus, under the given assumption the event A is written as a countable, disjoint, union of subsets of $\mathcal{G}$. It then follows using the properties of a probability measure

$$P(A) = \sum_{n\in\mathbb{N}} P(A_n \backslash A_{n-1}) = \sum_{n\in\mathbb{N}} (P(A_n) - P(A_{n-1})) = \lim_{n\to\infty} (P(A_n) - P(A_0)) = \lim_{n\to\infty} P(A_n) \quad .$$

2. We have

$$P(A_n) \underset{n\to\infty}{\downarrow} P\left(\bigcap_{n\in\mathbb{N}} A_n\right) \Leftrightarrow P(A_n^c) \underset{n\to\infty}{\uparrow} P\left(\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c\right) = P\left(\bigcup_{n\in\mathbb{N}} A_n^c\right) \quad ,$$

by de Morgan's law. The proof now follows from 1. ■

1.3.3 Random Variables

For the whole section let $(\Omega, \mathcal{G}, P)$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a Borel measurable space over $\mathbb{R}$.

When working with uncountable sample spaces, the measurability requirement behind Definition 26 of a random variable for finite sample spaces has to be modified. Basically, we are going to require measurability only for preimages of any Borel subset of $\mathbb{R}$, rather than measurability for preimages of any subset of $\mathbb{R}$. This is a necessary step, in order to be able to assign consistently probabilities to Borel events determined by the images of some random variable on $(\Omega, \mathcal{G}, P)$.

Definition 39 *Let $X : \Omega \to \mathbb{R}$ be a real valued function. (i) The sigma algebra*

$$\sigma(X) := \left\{X^{-1}(B) \quad : \quad B \in \mathcal{B}(\mathbb{R})\right\} \quad ,$$

is the sigma algebra generated by X. *(ii)* X *is a random variable on* $(\Omega, \mathcal{G})$ *if*

$$\sigma(X) \subset \mathcal{G} \quad .$$

Example 40 *For a set* $A \subset \Omega$ *let a function* $1_A : \Omega \to \{0, 1\}$ *be defined by*

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & otherwise \end{cases} \quad .$$

1_A *is called the indicator function of the set* A. *We have (please verify)*

$$\sigma(1_A) = \{\emptyset, \Omega, A, A^c\} \quad .$$

Hence, 1_A *is a random variable over* $(\Omega, \mathcal{G})$ *if and only if* $A \in \mathcal{G}$.

The measurability property in Definition 44 allows us to assign in a natural way probabilities also to Borel events that are induced by images of random variables, as is illustrated in the next example.

Example 41 *Let* X *be a random variable on a probability space* $(\Omega, \mathcal{F}, P)$. *For any event* $B \in \mathcal{B}(\mathbb{R})$ *we define*

$$\mathcal{L}_X(B) := P\left(X^{-1}(B)\right) \quad . \tag{4}$$

$\mathcal{L}_X$ *is a probability measure on* $\mathcal{B}(\mathbb{R})$, *the probability distribution of* X *(or the probability induced by* X *on* $\mathcal{B}(\mathbb{R})$*). Remark, that (4) is well defined, precisely because of the measurability of the random variable* X. *Showing that* $\mathcal{L}_X$ *is indeed a probability measure is very simple. In fact, we have:*

$$\mathcal{L}_X(\mathbb{R}) = P\left(X^{-1}(\mathbb{R})\right) = P(X \in \mathbb{R}) = P(\Omega) = 1 \quad .$$

Moreover, for any sequence $(B_i)_{i \in \mathbb{N}}$ *of disjoint events we obtain:*

$$\mathcal{L}_X\left(\bigcup_{i \in \mathbb{N}} B_i\right) = P\left(X^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right)\right) = P\left(\bigcup_{i \in \mathbb{N}} X^{-1}(B_i)\right) = \sum_{i=1}^{\infty} P\left(X^{-1}(B_i)\right) = \sum_{i=1}^{\infty} \mathcal{L}_X(B_i) \quad ,$$

using in the third equality the fact that $(B_i)_{i \in \mathbb{N}}$ *(and thus also* $\left(X^{-1}(B_i)\right)_{i \in \mathbb{N}}$*) is a sequence of disjoint events.*

Checking measurability of a candidate random variable can be by definition a quite hard and lengthy task, since we have to check preimages of any Borel subset of $\mathbb{R}$. Fortunately, the next result offers a much easier criterion by which measurability is easy to verify in many applications.

Proposition 42 *For a function* $X : \Omega \to \mathbb{R}$ *let*

$$\mathcal{E} := \left\{X^{-1}\left((-\infty, t)\right) : t \in \mathbb{R}\right\} = \left\{\{X < t\} : t \in \mathbb{R}\right\} \quad ,$$

be the set of preimages of open intervals of the form $(-\infty, t)$ *under* X*. Then it follows:*

$$\mathcal{E} \subset \mathcal{G} \quad \Leftrightarrow \quad \sigma(X) \subset \mathcal{G} \quad .$$

Proof. Define:

$$\mathcal{H} := \left\{B \in \mathcal{B}(\mathbb{R}) : X^{-1}(B) \in \mathcal{G}\right\} \subset \mathcal{B}(\mathbb{R}) \quad .$$

It is sufficient to show that under the given conditions $\mathcal{B}(\mathbb{R}) \subset \mathcal{H}$, i.e. $\mathcal{B}(\mathbb{R}) = \mathcal{H}$. We start by showing that $\mathcal{H}$ is a sigma algebra. We have first

$$X^{-1}(\emptyset) = \emptyset \in \mathcal{G} \quad ,$$

hence $\emptyset \in \mathcal{H}$. Second, for a set $B \in \mathcal{H}$ it follows

$$X^{-1}(B^c) = \left(\underbrace{X^{-1}(B)}_{\in \mathcal{G}}\right)^c \in \mathcal{G} \quad .$$

Finally, for a sequence $(B_n)_{n \in \mathbb{N}} \subset \mathcal{H}$ we have

$$X^{-1}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \bigcup_{n \in \mathbb{N}} \left(\underbrace{X^{-1}(B_n)}_{\in \mathcal{G}}\right) \in \mathcal{G} \quad ,$$

showing that $\mathcal{H}$ is a sigma algebra as claimed. Since $\mathcal{B}(\mathbb{R})$ is by definition the smallest sigma algebra containing all open intervals on the real line it is sufficient to show that under the given conditions $\mathcal{H}$ contains all open intervals on the real line. To this end, recall that all sets of the form $(-\infty, t)$ are by assumption elements of $\mathcal{H}$. For a general open interval (a, b), $a \leq b$ it then follows:

$$\begin{aligned}
X^{-1}((a,b)) &= X^{-1}\left((-\infty,b)\cap\left(\bigcap_{n\in\mathbb{N}}\left(-\infty,a+\frac{1}{n}\right)\right)^c\right)\\
&= X^{-1}\left((-\infty,b)\cap\left(\bigcup_{n\in\mathbb{N}}\left(-\infty,a+\frac{1}{n}\right)^c\right)\right)\\
&= \underbrace{X^{-1}((-\infty,b))}_{\in\mathcal{G}}\cap\left(\bigcup_{n\in\mathbb{N}}\left(\underbrace{X^{-1}\left(-\infty,a+\frac{1}{n}\right)}_{\in\mathcal{G}}\right)^c\right)\in\mathcal{G}.
\end{aligned}$$

This concludes the proof of the proposition. ■

Example 43 *Let $(X_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of random variables on $(\Omega,\mathcal{G})$. It then follows:*

1. *aX_1+bX_2 is a random variable for any a, $b\in\mathbb{R}$*
2. *$\sup_{n\in\mathbb{N}}X_n$ and $\inf_{n\in\mathbb{N}}X_n$ are random variables*
3. *$\limsup X_n := \lim_{n\to\infty}\sup_{k\geq n}X_k$ and $\liminf X_n := \lim_{n\to\infty}\inf_{k\geq n}X_k$ are random variables.*

Proof. We apply several times Proposition 42. 1. For $a,b\neq 0$ we have

$$\{aX_1+bX_2<t\}=\bigcup_{r\in\mathbb{Q}}\{aX_1<r\}\cap\{bX_2<t-r\}=\bigcup_{r\in\mathbb{Q}}\underbrace{\left\{X_1<\frac{r}{a}\right\}}_{\in\mathcal{G}}\cap\underbrace{\left\{X_2<\frac{t-r}{b}\right\}}_{\in\mathcal{G}}\in\mathcal{G}\quad.$$

For statement 2. we obtain:

$$\left\{\sup_{n\in\mathbb{N}}X_n<t\right\}=\bigcap_{n\in\mathbb{N}}\underbrace{\{X_n<t\}}_{\in\mathcal{G}}\in\mathcal{G}\quad,\quad\left\{\inf_{n\in\mathbb{N}}X_n<t\right\}=\bigcup_{n\in\mathbb{N}}\underbrace{\{X_n<t\}}_{\in\mathcal{G}}\in\mathcal{G}\quad.$$

3. For any $n\in\mathbb{N}$ it follows that $Y_n:=\sup_{k\geq n}X_k$ and $Z_n:=\inf_{k\geq n}X_k$ are random variables, by 2. Moreover, the sequences $(Y_n)_{n\in\mathbb{N}}$ and $(Z_n)_{n\in\mathbb{N}}$ are monotonically decreasing and increasing, respectively. Thefore:

$$\begin{aligned}
\{\limsup X_n<t\} &= \left\{\lim_{n\to\infty}Y_n<t\right\}=\bigcup_{n\in\mathbb{N}}\underbrace{\{Y_n<t\}}_{\in\mathcal{G}}\in\mathcal{G}\quad.\\
\{\liminf X_n<t\} &= \left\{\lim_{n\to\infty}Z_n<t\right\}=\bigcap_{n\in\mathbb{N}}\underbrace{\{Z_n<t\}}_{\in\mathcal{G}}\in\mathcal{G}\quad.
\end{aligned}$$

This concludes the proof. ■

1.3.4 Expected Value and Lebesgue Integral

For the whole section let $(\Omega, \mathcal{G}, P)$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be the Borel measurable space over $\mathbb{R}$.

The expected value of a general random variable is defined as its Lebesgue integral with respect to some probability measure P on $(\Omega, \mathcal{G})$. More generally, Lebesgue integrals of measurable functions can be defined with respect to some measure (as for instance Lebesgue measure μ_0) defined on a corresponding measurable space (as for instance the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

The construction of the Lebesgue integral for a general random variable X starts by defining the value of the Lebesgue integral for linear combinations of indicator functions, goes over to extend the integral to functions that are pointwise monotonic limits of sequences of simple functions, and finally defines the integral for the more general case of an integrable random variable (see below the precise definition.)

Definition 44 *(i) A random variable X is simple if*

$$X = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i} \quad ,$$

where $n \in \mathbb{N}$, $c_1, .., c_n \in \mathbb{R}$, and $A_1, .., A_n \in \mathcal{G}$ are mutually disjoint events. The (vector) space of simple random variables on $(\Omega, \mathcal{G})$ is denoted by $S(\mathcal{G})$.The expected value $E(X)$ of a simple function X is defined by

$$E(X) := \int_{\Omega} X dP := \sum_{i=1}^{n} c_i P(A_i) \quad .$$

(ii) Let $X \geq 0$ be a non negative random variable. The expected value $E(X)$ of X is defined by

$$E(X) := \int_{\Omega} X dP := \sup \left\{ \int_{\Omega} Y dP : Y \leq X \text{ and } Y \in S(\mathcal{G}) \right\} \quad .$$

(iii) A random variable X is integrable, if

$$E\left(X^{+}\right) < \infty \quad , \quad E\left(X^{-}\right) < \infty \quad ,$$

where $X^{+} := \max(X, 0)$ and $X^{-} := \max(-X, 0)$ are the positive and negative part of X, respectively. We denote the (vector) space of integrable random variable by $L_1(P)$. For any $X \in L_1(P)$

the expected value $E(X)$ *of* X *is defined by*

$$E(X) = E\left(X^{+}\right) - E\left(X^{-}\right) \quad .$$

(iv) Finally, for a random variable $X \in L_1(P)$ *and a set* $A \in \mathcal{G}$ *we define*

$$\int_A X dP := \int_\Omega \mathbf{1}_A X dP$$

Remark 45 *(i) The key point in the definition of* $E(X)$ *is (iii). In fact, (iii) is a quite reasonable definition because for any random variable* $X \geq 0$ *there always exists a sequence* $(X_n)_{n\in\mathbb{N}}$ *of simple random variables converging monotonically pointwise to* X *from below. Such a sequence is obtained for instance by setting for any* $\omega \in \Omega$

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{1}_{\left\{\frac{k-1}{2^n} < X \leq \frac{k}{2^n}\right\}}(\omega) + n\mathbf{1}_{\{X>n\}}(\omega) \quad .$$

Moreover, it can be shown that the limit of the sequence of integrals $E(X_n)$ *does not depend on the choice of the specific approximating sequence. Therefore, (iii) in Definition 44 could be also equivalently written as*

$$E(X) := \lim_{n\to\infty} E(X_n) := \lim_{n\to\infty} \int_\Omega X_n dP \quad ,$$

for a given approximating sequence $(X_n)_{n\in\mathbb{N}}$*. (ii) As mentioned, expected values are by definition just integrals of measurable functions with respect to some probability measure. In fact, the definition of the Lebesgue integral of a measurable function with respect to some measure* μ*, say, follows exactly the same steps as above, readily by replacing everywhere the probability measure* P *with the measure* μ *in (i), (ii), (iii) and (iv).*

Let us discuss some first (very) simple examples of expected values computed using the above definitions.

Example 46 *Let* $\Omega := \mathbb{R}$, $\mathcal{G} := \mathcal{B}(\mathbb{R})$ *and set for any* $A \in \mathcal{G}$

$$P(A) = \mu_0(A \cap [0,1]) \quad .$$

The expected value of $X := \mathbf{1}_{\mathbb{Q}}$ *is*

$$E(X) = 1 \cdot \mu_0 (\mathbb{Q} \cap [0,1]) = 0 \quad ,$$

because $\mathbb{Q}$ *is a countable set. Notice, that this function is not Riemann integrable in the usual sense. The expected value of*

$$Y(\omega) := \begin{cases} \infty & \omega = 0 \\ 0 & otherwise \end{cases} \quad ,$$

can be computed as the limit of the expected values in an approximating sequence $(X_n)_{n\in\mathbb{N}}$ *of simple functions given by*

$$X_n(\omega) := \begin{cases} n & \omega = 0 \\ 0 & otherwise \end{cases} \quad .$$

Hence:

$$E(Y) = \lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} n \cdot \mu_0 (\{0\} \cap [0,1]) = \lim_{n\to\infty} n \cdot \mu_0 (\{0\}) = 0 \quad .$$

Notice, that also Y *is not Riemann integrable in the usual sense.*

The basic properties of the above integral definition are collected in the next proposition.

Proposition 47 *Let* X,$Y \in L_1(P)$ *and* a,$b \in \mathbb{R}$*; it then follows:*

1. $E(aX + bY) = aE(X) + bE(Y)$

2. *If* $X \leq Y$ *pointwise, then*

$$E(X) \leq E(Y)$$

3. *For two sets* $A, B \in \mathcal{G}$ *such that* $A \cap B = \emptyset$ *it follows*

$$\int_{A\cup B} X dP = \int_\Omega \mathbf{1}_{A\cup B} X dP = \int_\Omega (\mathbf{1}_A + \mathbf{1}_B) X dP =_{1.} \int_A X dP + \int_B X dP \quad .$$

Proof. 1. For brevity we show this property only for indicator functions $X = \mathbf{1}_A$, $Y = \mathbf{1}_B$, where A, $B \in \mathcal{G}$ are disjoint events. We have

$$E(aX + bY) = E(a\mathbf{1}_A + b\mathbf{1}_B) =_{\text{Def (ii)}} aP(A) + bP(B) = aE(\mathbf{1}_A) + bE(\mathbf{1}_B) = aE(X) + bE(Y) \quad .$$

2. If $Y - X \geq 0$, then there exists a sequence of simple approximating functions $X_n \geq 0$ converging monotonically to $Y - X$. This implies:

$$E(Y) - E(X) =_{1.} E(Y - X) = \lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} \sum_{i=1}^{k_n} c_{in} P(A_{in}) \geq 0 \quad ,$$

say, because for any $n \in \mathbb{N}$ we have $c_{1n}, .., c_{k_n n} \geq 0$. ■

1.3.5 Some Further Examples of Probability Spaces with uncountable Sample Spaces

For the whole section let $(\Omega, \mathcal{G}, P)$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be the Borel measurable space over $\mathbb{R}$.

Using Lebesgue integrals we are also able to construct probability measures by integrating a suitable (density) function over events $A \in \mathcal{G}$. A well-known example in this respect arises by integrating the density function of a standard normal distribution.

Example 48 $(\Omega, \mathcal{G}) := (\mathbb{R}, \mathcal{B}(\mathbb{R}))$; $\phi : \mathbb{R} \to \mathbb{R}^+$ *is defined by*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad , \quad x \in \mathbb{R} \quad .$$

ϕ *is the density function of a standard normally distributed random variable and is such that*

$$\int_{\mathbb{R}} \phi(x)\, d\mu_0(x) = \int_{-\infty}^{\infty} \phi(x)\, dx = 1 \quad ,$$

i.e. $\phi \in L_1(\mu_0)$. *A standard normal probability distribution* P *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is obtained by setting for any* $A \in \mathcal{G}$:

$$P(A) := \int_A \phi(x)\, d\mu_0(x) \quad .$$

It is straightforward to verify, using the basic properties of Lebesgue integrals together with some monotone convergence property, that P *is indeed a probability measure.*

More generally, densities can be also defined on abstract probability spaces, as is demonstrated in the next final example.

Example 49 *Let $X \geq 0$ be a random variable on $(\Omega, \mathcal{G})$ such that $X \in L_1(P)$, and define*

$$Q(A) := \frac{E(\mathbf{1}_A X)}{E(X)} = E\left(\mathbf{1}_A \frac{X}{E(X)}\right) \quad .$$

It is easy to verify, using the basic properties of Lebesgue integrals together with some monotone convergence property, that Q is a further probability measure on $(\Omega, \mathcal{G})$. Moreover, the absolute continuity property

$$P(A) = 0 \Rightarrow Q(A) = 0 \quad ,$$

follows from the definition. If, moreover,

$$P(A) = 0 \Longleftrightarrow Q(A) = 0$$

the probabilities Q and P are called equivalent. This property holds when $X > 0$. The random variable $Z := \frac{X}{E(X)}$ is called the Radon Nykodin derivative of Q with respect to P, denoted by $\frac{dQ}{dP}$. By construction $\frac{dQ}{dP}$ is a density function on $(\Omega, \mathcal{G})$ because $\frac{dQ}{dP} \geq 0$ and

$$E\left(\frac{dQ}{dP}\right) = E\left(\frac{X}{E(X)}\right) = 1 \quad .$$

1.4 Stochastic Independence

For the whole section let $(\Omega, \mathcal{G}, P)$ be a probability space

Definition 50 *Two events $A, B \in \mathcal{G}$ are (stochastically) independent if*

$$P(A \cap B) = P(A)P(B) \quad . \tag{5}$$

We use the notation $A \perp B$ to denote two independent events.

Remark 51 *Condition (5) states that two events are independent if and only if their conditional and unconditional probabilities are the same, i.e.:*

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \underset{A \perp B}{=} \frac{P(A)P(B)}{P(B)} = P(A) \quad ,$$

(provided of course $P(B) > 0$). This property is symmetric in A, B.

Example 52 *Stochastic independence is a feature determined by the structure of the underlying probability P. As an illustration of this fact consider again the two period binomial model of Example 1. We have there:*

$$P(HH, HT) P(HT, TH) = \left(p^2 + p(1-p)\right)(2p(1-p)) = 2p^2(1-p) \quad , \tag{6}$$

and

$$P(\{HH, HT\} \cap \{HT, TH\}) = P(HT) = p(1-p) \quad . \tag{7}$$

Therefore, (6) and (7) are equal if and only if $p = \frac{1}{2}$, that is the only binomial probability under which the above events are independent is the one implied by $p = \frac{1}{2}$.

The concept of stochastic independence between events can be naturally extended to stochastic independence between information sets, i.e. sigma algebras.

Definition 53 *Two sigma algebras $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{G}$ are stochastically independent if for all $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$ one has $A \perp B$. We use the notation $\mathcal{G}_1 \perp \mathcal{G}_2$ to denote independent sigma algebras.*

Example 54 *In the two period binomial model of Example 1 we define the two following sigma algebras:*

$$\mathcal{G}_1 := \{\emptyset, \Omega, \{HT, HH\}, \{TT, TH\}\} \quad ,$$

the sigma algebra generated by the first price increment, and

$$\mathcal{G}_2 := \{\emptyset, \Omega, \{HH, TH\}, \{TT, HT\}\} \quad ,$$

the sigma algebra generated by the second price movement. We then have, for any $p \in [0,1]$*:*

$$\mathcal{G}_1 \perp \mathcal{G}_2 \quad .$$

For instance, for the sets $\{HT, HH\}$ *and* $\{HH, TH\}$ *one obtains*

$$P(HT, HH)\, P(HH, TH) = \left(p^2 + p(1-p)\right)\left(p^2 + p(1-p)\right) = p^2 \quad ,$$

and

$$P(\{HT, HH\} \cap \{HH, TH\}) = P(HH) = p^2 \quad .$$

These features derive directly from the way how probabilities are assigned by a binomial distribution where

$$P(\omega) = p^{\#\ of\ H\ in\ \omega} (1-p)^{\#\ of\ T\ in\ \omega} \quad .$$

Finally, we can also define independence between random variables as independence of the information sets they generate.

Definition 55 *Two random variables* X,Y *on* $(\Omega, \mathcal{G}, P)$ *are independent if*

$$\sigma(X) \perp \sigma(Y) \quad .$$

We use the notation $X \perp Y$ *to denote independence between random variables.*

Example 56 *We already discussed that the two sigma algebras* $\mathcal{G}_1$*,* $\mathcal{G}_2$ *of Example 54 are independent in the binomial model. Notice that we have (please verify!)*

$$\mathcal{G}_1 = \{\emptyset, \Omega, \{HT, HH\}, \{TT, TH\}\} = \sigma(S_1/S_0) \quad ,$$

and

$$\mathcal{G}_2 := \{\emptyset, \Omega, \{HH, TH\}, \{TT, HT\}\} = \sigma(S_2/S_1) \quad .$$

Therefore, the stock price returns S_1/S_0 *and* S_2/S_1 *in a binomial model are stochastically independent.*

Example 57 *Let A, $B \in \mathcal{G}$ be two independent events and let the functions*

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ \\ 0 & otherwise \end{cases} \quad , \quad 1_B(\omega) = \begin{cases} 1 & \omega \in B \\ \\ 0 & otherwise \end{cases} \quad ,$$

be the indicator functions of the sets A and B, respectively. We then have (please verify):

$$\sigma(1_A) = \{\emptyset, \Omega, A, A^c\} \quad , \quad \sigma(1_B) = \{\emptyset, \Omega, B, B^c\} \quad .$$

Therefore, $1_A \perp 1_B$ if and only if $A \perp B$ (please verify).

Some properties related to independence are important. The first one says that independence is maintained under (measurable) transformations.

Proposition 58 *Let X,Y be independent random variables on $(\Omega, \mathcal{G}, P)$ and $h, g : \mathbb{R} \to \mathbb{R}$ be two (measurable) functions. It then follows:*

$$h(X) \perp g(Y) \quad .$$

Proof. We give a graphical proof of this statement, which makes use of the fact that preimages of composite mappings are contained in the preimage of the first function in the composition:

$$\begin{array}{ccc} \sigma(X) & \perp_{\text{By assumption}} & \sigma(Y) \\ \cup & & \cup \\ \sigma(h(X)) & & \sigma(g(Y)) \end{array} \quad .$$

■

The second important property of stochastic independence is related to the expectation of a product of random variables.

Proposition 59 *Let X,Y be independent random variables on $(\Omega, \mathcal{G}, P)$. It then follows*

$$E(XY) = E(X)E(Y) \quad .$$

Proof. For the sake of brevity we give the proof for the simplest case where $X = 1_A$, $Y = 1_B$, for events $A, B \in \mathcal{G}$ such that $A \perp B$. As usual, the extension of this result for more general setting requires considering linear combinations of indicator functions, i.e. simple functions, and pointwise limits of simple functions. For the given simplified setting we have:

$$\begin{aligned} E(XY) &= E(1_A 1_B) = E(1_{A\cap B}) = 1 \cdot P(A\cap B) + 0 \cdot P\left((A\cap B)^c\right) \\ &= P(A\cap B) \underset{A\perp B}{=} P(A)\,P(B) = E(1_A)\,E(1_B) = E(X)\,E(Y) \quad . \end{aligned}$$

This concludes the proof. ■

2 Conditional Expectations and Martingales

For the whole section let $(\Omega, \mathcal{G}, P)$ be a probability space

2.1 The Binomial Model Once More

For later reference, we summarize the structure of a general $n-$period binomial model, since it will be used to illustrate some of the concepts introduced below.

- $I := \{0, 1, 2, .., n\}$ is a discrete time index representing the available transaction dates in the model
- The sample space is given by $\Omega := \{\text{Sequences of } n \text{ coordinates } H \text{ or } T\}$ with single outcomes ω of the form

$$\omega = \underbrace{(TTTH..HT)}_{n \text{ coordinates}} \quad ,$$

  for instance.
- $\mathcal{G} := \mathcal{F}$, the sigma algebra of all subsets of Ω
- Dynamics of the stock price and money account:

$$S_t = \begin{cases} uS_{t-1} & \text{with probability } p \\ dS_{t-1} & \text{with probability } 1-p \end{cases} \quad , \quad B_t = (1+r)\,B_{t-1}$$

for given $B_0 = 1$, S_0 and where

$$u = 1/d \quad , \quad u > 1 + r > d \quad .$$

The sequence $(S_t)_{t=0,..,n}$ is a sequence of random variables defined on a *single* probability space $(\Omega, \mathcal{G}, P)$. This is an example of a so called stochastic process on $(\Omega, \mathcal{G}, P)$. Associated with stochastic processes are flows of information sets (i.e. sigma algebras) generated by the process history up to a given time. For instance, for any $t \in I$ we can define

$$\mathcal{G}_t := \sigma\left(\sigma\left(S_0\right), \sigma\left(S_1\right), .., \sigma\left(S_t\right)\right) := \sigma\left(\bigcup_{k=0}^{t} \sigma\left(S_k\right)\right) \quad ,$$

the smallest sigma algebra containing all sigma algebras generated by S_0, S_1,...,S_t. $\mathcal{G}_t$ represents the information about a single outcome $\omega \in \Omega$ which can be obtained exclusively by observing the price process up to time t. Clearly,

$$\mathcal{G}_t \subset \mathcal{G}_s \Longleftrightarrow t \leq s \quad .$$

Therefore, the sequence $(\mathcal{G}_t)_{t=0,..,n}$ constitutes a filtration, the filtration generated by the process $(S_t)_{t=0,..,n}$.

2.2 Sub Sigma Algebras and (Partial) Information

We model partial information about single outcomes $\omega \in \Omega$ or about single events $A \in \mathcal{G}$ using sub sigma algebras of $\mathcal{G}$.

Example 60 *Let X be a random variable on $(\Omega, \mathcal{G})$. Then $\sigma(X)$ is (by definition) a sub sigma algebra of $\mathcal{G}$. $\sigma(X)$ represent the partial information about an outcome $\omega \in \Omega$ which can be obtained by observing $X(\omega)$. For instance, set $n = 3$ in the above binomial model and consider the outcome $\omega = (TTT)$. By observing S_1, i.e. using $\sigma(S_1)$ as the available information set we can only conclude*

$$\omega \in \{TTT, THH, THT, TTH\} \quad (\Leftrightarrow S_1(\omega) = S_0 d) \quad .$$

However, when observing all price movements from $t = 0$ to $t = 3$ we can make use of the sigma algebra

$$\mathcal{G}_3 := \sigma\left(\bigcup_{t=0}^{3} \sigma(S_t)\right) \quad ,$$

to fully identify $\omega \in \Omega$. Both $\sigma(S_1)$ and $\mathcal{G}_3$ are sub sigma algebras of $\mathcal{G}$, which however represent different pieces of information about $\omega \in \Omega$

Based on the above simple considerations we can now formally define what it means for an event to be "realized".

Definition 61 *(i) An event $A \in \mathcal{G}$ is realized by means of a sub sigma algebra $\mathcal{G}' \subset \mathcal{G}$ if $A \in \mathcal{G}'$.*
(ii) Let $\mathcal{G}_t$ be a sigma algebra generated by some price process[1] up to time t. We say that A is realized by means of the price information up to time t if $A \in \mathcal{G}_t$.

Remark 62 *By definition, realization of an event $A \in \mathcal{G}$ by means of $\mathcal{G}'$ is precisely measurability of that event with respect to the sub sigma algebra $\mathcal{G}'$. Precisely, given an event $A \in \mathcal{G}$ we can determine it uniquely using $\mathcal{G}'$, i.e. we can say that A has been realized, if and only if $A \in \mathcal{G}'$. For instance, in the above 3−period binomial model we can consider the event*

$$A = \{TTT\} \quad .$$

Clearly, $A \notin \sigma(S_1)$ since we do not know using $\sigma(S_1)$ the value of the second and the third coin tosses. Therefore, A is not realized by means of $\sigma(S_1)$, i.e. it is not realized by means of the price information up to time 1. However,

$$A \in \mathcal{G}_3 := \sigma\left(\bigcup_{t=0}^{3} \sigma(S_t)\right) \quad ,$$

i.e. A is realized by means of the whole price information available up to time 3.

Example 63 *The event {The first two price returns are both positive} is realized by means of the price information up to time 2, while the event {The total number of positive price returns is 2} is not.*

[1] *See for instance the above examples.*

2.3 Conditional Expectations

For the whole section let X be a random variable on $(\Omega, \mathcal{G})$.

2.3.1 Motivation

Given an event $A = X^{-1}(a) \in \mathcal{G}$, for some $a \in \mathbb{R}$, we are always able to identify for any $\omega \in A$ the corresponding value $X(\omega)$ of the random variable X using the information set $\mathcal{G}$. Indeed, we then have by definition

$$\omega \in X^{-1}(a) \underset{\sigma(X) \subset \mathcal{G}}{\in} \mathcal{G} \quad , \quad \text{i.e. } X(\omega) = a \quad ,$$

for all $\omega \in A$. However, using a coarser information set $\mathcal{G}' \subset \sigma(X)$ it may happen that we are not able to fully determine the value $X(\omega)$ that a random variable X associates to a given single outcome $\omega \in A$. Specifically, it may happen that based on the information available in $\mathcal{G}'$ we can only state for some *non singleton* set $B \in \mathcal{B}(\mathbb{R})$

$$\omega \in X^{-1}(B) \quad , \quad \text{i.e. } X(\omega) \in B \quad . \tag{8}$$

In this case, the information set $\mathcal{G}'$ is not sufficiently fine to fully determine the precise value of $X(\omega)$ associated with a specific $\omega \in A$. Thus, the goal in such a situation is to define a suitable candidate prediction $E(X|\mathcal{G}')(\omega)$ for the unknown value $X(\omega)$ based on the information $\mathcal{G}'$. We will call $E(X|\mathcal{G}')$ the conditional expectation of X conditionally on $\mathcal{G}'$. Notice, that a first necessary requirement on $E(X|\mathcal{G}')$ is that it can be fully determined using the information $\mathcal{G}'$, that is it has to be $\mathcal{G}'-$measurable. Further, a natural idea to compute the prediction $E(X|\mathcal{G}')$ as an unbiased forecast such that the expectation of $E(X|\mathcal{G}')$ and X agree on all sets $A \in \mathcal{G}'$:

$$\int_A E(X|\mathcal{G}')\, dP = \int_A X dP$$

(see below the precise definition).

2.3.2 Definition and Properties

Definition 64 *Let $\mathcal{G}' \subset \mathcal{G}$ be a sub sigma algebra. The conditional expectation $E(X|\mathcal{G}')$ of X conditioned on the sigma algebra $\mathcal{G}'$ is a random variable satisfying:*

1. *$E(X|\mathcal{G}')$ is $\mathcal{G}'-$measurable*

2. *For any $A \in \mathcal{G}'$:*

$$\int_A E(X|\mathcal{G}')\,dP = \int_A XdP \quad ,$$

(partial averaging property).

In the sequel, we write for any further random variable Y on $(\Omega, \mathcal{G})$:

$$E(X|Y) := E(X|\sigma(Y))$$

Remark 65 *(i) $E(X|\mathcal{G}')$ exists, provided $X \in L_1(P)$; this is a consequence of the so called Radon Nykodin Theorem. (ii) The random variable $E(X|\mathcal{G}')$ is unique, up to events of zero probability. Precisely, if Y and Z are two candidate $\mathcal{G}'-$measurable random variables satisfying 2. of the above definition, then:*

$$P(Y = Z) = 1$$

Example 66 *(i) If $\mathcal{G}' = \{\emptyset, \Omega\}$ then $E(X|\mathcal{G}') = E(X)\mathbf{1}_\Omega$, that is conditional expectations conditioned on trivial information sets are unconditional expectations. Indeed, $E(X)\mathbf{1}_\Omega$ is $\mathcal{G}'$ measurable and*

$$\int_\Omega E(X)\,\mathbf{1}_\Omega dP = E(X)\,P(\Omega) = E(X) = \int_\Omega XdP$$

(ii) If X is $\mathcal{G}'-$measurable then $E(X|\mathcal{G}') = X$, that is if the conditioning information set is sufficiently fine to determine X completely then conditional expectations of a random variable are the random variable itself. Indeed, in this case we trivially have:

$$\int_A E(X|\mathcal{G}')\,dP = \int_A XdP \quad ,$$

for any set $A \in \mathcal{G}'$.

Proposition 67 *Let $\mathcal{G}' \subset \mathcal{G}$ be a sub sigma algebra and $X, Y \in L_1(P)$. It then follows:*

1. $E(E(X|\mathcal{G}')) = E(X)$ *(Law of Iterated Expectations).*

2. *For any $a, b \in \mathbb{R}$:*

$$E(aX + bY|\mathcal{G}') = aE(X|\mathcal{G}') + bE(Y|\mathcal{G}') \quad ,$$

   (Linearity).

3. *If $X \geq 0$ then $E(X|\mathcal{G}') \geq 0$ with probability 1 (Monotonicity).*

4. *For any sub sigma algebra $\mathcal{H} \subset \mathcal{G}'$:*

$$E(E(X|\mathcal{G}')|\mathcal{H}) = E(X|\mathcal{H}) \quad ,$$

   (Tower Property).

5. *If $\sigma(X) \perp \mathcal{G}'$ then*

$$E(X|\mathcal{G}') = E(X)\mathbf{1}_\Omega \quad ,$$

   (Independence).

6. *If V is a $\mathcal{G}'-$measurable random variable such that $VX \in L_1(P)$ then*

$$E(VX|\mathcal{G}') = VE(X|\mathcal{G}')$$

Proof. 1. Set $A = \Omega \in \mathcal{G}'$; by definition in then follows

$$E(X) = \int_\Omega X dP = \int_\Omega E(X|\mathcal{G}')\, dP = E(E(X|\mathcal{G}')) \quad .$$

2. By construction $aE(X|\mathcal{G}') + bE(Y|\mathcal{G}')$ is $\mathcal{G}'$ measurable. Moreover, for any $A \in \mathcal{G}'$:

$$\begin{aligned}
\int_A (aE(X|\mathcal{G}') + bE(Y|\mathcal{G}'))\, dP &= a\int_A E(X|\mathcal{G}')\, dP + b\int_A E(Y|\mathcal{G}')\, dP \\
&= a\int_A X dP + b\int_A Y dP \\
&= \int_A (aX + bY)\, dP \quad ,
\end{aligned}$$

using in the first and the third equality the linearity of Lebesgue integrals and in the second equality the definition of conditional expectations.

3. Let

$$A := \{E(X|\mathcal{G}') < 0\} \in \mathcal{G}' \quad .$$

Then,

$$\int_A E(X|\mathcal{G}')\, dP = \int_A X dP \geq 0 \quad ,$$

since $X \geq 0$ and by the monotonicity of Lebesgue integrals. Further, the monotonicity of Lebesgue integrals also implies

$$\int_A E(X|\mathcal{G}')\, dP \leq 0 \quad ,$$

since $\mathbf{1}_A E(X|\mathcal{G}') < 0$. Therefore,

$$\int_A E(X|\mathcal{G}')\, dP = 0 \quad ,$$

implying $P(A) = 0$.

4. $E(X|\mathcal{H})$ is by definition $\mathcal{H}$ measurable. Further, for any $A \in \mathcal{H}$:

$$\int_A E(X|\mathcal{H})\, dP = \int_A X dP = \int_A E\left(X|\mathcal{G}'\right) dP =: \int_A Y dP \quad ,$$

since $A \in \mathcal{G}'$ because $\mathcal{H} \subset \mathcal{G}'$. By definition, this implies that $E(X|\mathcal{H})$ is the conditional expectation of the random variable $Y := E\left(X|\mathcal{G}'\right)$ conditioned on the sigma algebra $\mathcal{H}$.

5. $E(X)\mathbf{1}_\Omega$ is trivially $\mathcal{G}'-$measurable. We show the statement for the case $X = \mathbf{1}_B$, where $B \in \mathcal{G}$. The extension to the general case follows by standard arguments. We have for any $A \in \mathcal{G}'$:

$$\begin{aligned}\int_A E(X)\mathbf{1}_\Omega dP &= E(X)P(A) = E(\mathbf{1}_B)P(A) = P(B)P(A) \\ &= P(A \cap B) = E(\mathbf{1}_A\mathbf{1}_B) = \int_A X dP \quad ,\end{aligned}$$

using in the fourth equality the independence assumption, in the fifth the properties of indicator functions and in the sixth the definition of X. 6. $VE(X|\mathcal{G}')$ is $\mathcal{G}'-$measurable. Again, we show the statement for the simpler case $V = \mathbf{1}_B$, where $B \in \mathcal{G}'$. We have for any $A \in \mathcal{G}'$,

$$\begin{aligned}\int_A V E\left(X \middle| \mathcal{G}'\right) dP &= \int_A \mathbf{1}_B E\left(X \middle| \mathcal{G}'\right) dP = \int_{A\cap B} E\left(X \middle| \mathcal{G}'\right) dP \\ &= \int_{A\cap B} X dP = \int_A \mathbf{1}_B X dP = \int_A V X dP \quad ,\end{aligned}$$

using in the third equality the definition of conditional expectations, and otherwise the properties of indicator functions. ■

Example 68 *In the $n-$period Binomial model we have*

$$E\left(S_1 | \sigma\left(S_1\right)\right) = S_1 \quad ,$$

by the $\sigma\left(S_1\right)-$measurability of S_1. S_2 is not $\sigma\left(S_1\right)-$measurable. However, we know that

$$\sigma\left(S_2/S_1\right) \perp \sigma\left(S_1\right) \quad .$$

Therefore,

$$E\left(S_2 | \sigma\left(S_1\right)\right) = E\left(\frac{S_2}{S_1} S_1 \middle| \sigma\left(S_1\right)\right) = S_1 E\left(\frac{S_2}{S_1} \middle| \sigma\left(S_1\right)\right) = S_1 E\left(\frac{S_2}{S_1}\right) = S_1\left(pu + (1-p)d\right) \quad .$$

More generally, we have

$$\sigma\left(S_t/S_{t-1}\right) \perp \mathcal{G}_{t-1} \quad ,$$

where

$$\mathcal{G}_{t-1} := \sigma\left(\bigcup_{k=0}^{t-1} \sigma\left(S_k\right)\right) \quad ,$$

$t = 1, \ldots, n$. *Therefore, by the same arguments:*

$$E\left(S_t | \mathcal{G}_{t-1}\right) = S_{t-1}\left(pu + (1-p)d\right) \quad .$$

Finally, the tower property gives after some iterations:

$$\begin{aligned}
E\left(S_{t+k}|\mathcal{G}_{t-1}\right) &= E\left(E\left(S_{t+k}|\mathcal{G}_{t+k-1}\right)|\mathcal{G}_{t-1}\right) \\
&= E\left(S_{t+k-1}\left(pu+(1-p)\,d\right)|\mathcal{G}_{t-1}\right) \\
&= \left(pu+(1-p)\,d\right)E\left(S_{t+k-1}|\mathcal{G}_{t-1}\right) \\
&= \ldots \\
&= \left(pu+(1-p)\,d\right)^{k}E\left(S_{t}|\mathcal{G}_{t-1}\right) \\
&= \left(pu+(1-p)\,d\right)^{k+1}S_{t-1} \quad .
\end{aligned}$$

2.4 Martingale Processes

We now introduce a class of stochastic processes that are particularly important in finance: the class of martingale processes. Indeed, it will turn out in a later chapter that the price processes of many financial instruments are martingale processes after a suitable change of probability. In this section we give the necessary definitions and present some first examples of martingale processes.

Definition 69 *(i) Let* $\mathbb{G} := (\mathcal{G}_t)_{t=0,..,n}$ *be a filtration over* $(\Omega, \mathcal{G}, P)$*. The quadruplet* $(\Omega, \mathcal{G}, \mathbb{G}, P)$ *is called a filtered probability space. (ii) A stochastic process* $\mathcal{X} := (X_t)_{t=0,..,n}$ *on a filtered probability space* $(\Omega, \mathcal{G}, \mathbb{G}, P)$ *is adapted (is* $\mathbb{G}-$*adapted) if for any* $t = 0, .., n$ *the random variable* X_t *is* $\mathcal{G}_t-$*measurable. (iii) A* $\mathbb{G}-$*adapted process is a martingale if for any* $t = 0, .., n-1$ *one has*

$$X_t = E\left(X_{t+1}|\,\mathcal{G}_t\right) \quad , \tag{9}$$

(martingale condition). The process is a submartingale (a supermartingale) if in (9) the "$\leq$*" sign (the "*$\geq$*" sign) holds.*

Remark 70 *Notice, that in Definition 69 both the filtration* $\mathbb{G}$ *and the relevant probability* P *are crucial in determining the validity of the martingale condition (9) for an adapted process. Indeed,* *different probabilities and filtrations can imply (9) to be satisfied or not. For instance, in the $n-$period binomial model we obtained, using the filtration generated by the stock price process,*

$$E\left(S_t|\mathcal{G}_{t-1}\right) = S_{t-1}\left(pu + (1-p)\,d\right) \quad .$$

Therefore, the only binomial probability measure under which the stock price process is a martingale is the one satisfying

$$pu + (1-p)\,d = 1 \ , \ \textit{i.e.} \ p = \frac{1-d}{u-d} \quad . \tag{10}$$

The binomial probabilities such that $p > (1-d)/(u-d)$ ($p < (1-d)/(u-d)$) imply a stock price process that is a submartingale (a supermartingale).

Being a martingale is a quite strong condition on a stochastic process, which strongly relates future process coordinates with current ones. This is made more explicit below.

Proposition 71 *Let $(X_t)_{t=0,..,n}$ be a martingale on the filtered probability space $(\Omega, \mathcal{G}, \mathbb{G}, P)$.*

1. *It then follows for any $t, s \in \{0, 1, \ldots, n\}$ such that $s \geq t$:*

$$X_t = E\left(\left. X_s \right| \mathcal{G}_t\right) \quad .$$

2. *If $(Y_t)_{t=0,..,n}$ is a further martingale on the filtered probability space $(\Omega, \mathcal{G}, \mathbb{G}, P)$ and such that $Y_n = X_n$ then $Y_t = X_t$ almost surely for all $t \in \{0, 1, \ldots, n\}$.*

Proof. 1. The tower property combined with the martingale property implies

$$X_t = E\left(\left. X_{t+1} \right| \mathcal{G}_t\right) = E\left(E\left(\left. X_{t+2} \right| \mathcal{G}_{t+1}\right) \right| \mathcal{G}_t\right) = E\left(\left. X_{t+2} \right| \mathcal{G}_t\right) = ... = E\left(\left. X_{t+k} \right| \mathcal{G}_t\right) \quad ,$$

for any $k = s - t$.

2. From 1. we have

$$X_t = E\left(\left. X_n \right| \mathcal{G}_t\right) = E\left(\left. Y_n \right| \mathcal{G}_t\right) = Y_t \quad .$$

This concludes the proof. ■

Example 72 *AR(1) process: Let* $(\varepsilon_t)_{t=1,\ldots,n}$ *be an identically distributed, zero mean, adapted process on a filtered probability space* $(\Omega, \mathcal{G}, \mathbb{G}, P)$ *and such that for any* t *the random variable* ε_t *is independent from the process history up to time* $t-1$*, i.e.:*

$$\sigma(\varepsilon_t) \perp \sigma\left(\bigcup_{i=1}^{t-1} \sigma(\varepsilon_i)\right) \quad , \quad t = 1, \ldots, n \quad . \tag{11}$$

An Autoregressive Process of Order 1 (AR(1)) is defined by

$$X_t = \begin{cases} 0 & t = 0 \\ \rho X_{t-1} + \varepsilon_t & t > 0 \end{cases} ,$$

where $\rho \in \mathbb{R}$*. It is easily seen that* $(X_t)_{t=0,..,n}$ *is* $\mathbb{G}-$*adapted. Furthermore, for any* $t = 1, \ldots, n$*,*

$$E(X_t | \mathcal{G}_{t-1}) = E(\rho X_{t-1} + \varepsilon_t | \mathcal{G}_{t-1}) = \rho E(X_{t-1} | \mathcal{G}_{t-1}) + E(\varepsilon_t | \mathcal{G}_{t-1}) = \rho X_{t-1} + E(\varepsilon_t) = \rho X_{t-1} \quad ,$$

using in the second equality the linearity of conditional expectations, in the third the $\mathcal{G}_{t-1}-$*measurability of* X_{t-1} *and the independence assumption (11), and in the fourth the zero mean property of* ε_t *(*$E(\varepsilon_t) = 0$*). Therefore, an AR(1) process is a martingale if and only if* $\rho = 1$*. The process resulting for* $\rho = 1$ *is called a "Random Walk" process.*

Example 73 *MA(1) process: Let* $(\varepsilon_t)_{t=0,..,n}$ *be the same process as in Example 72. A Moving Average Process of Order 1 (MA(1)) is defined by*

$$X_t = \begin{cases} 0 & t = 0 \\ \varepsilon_1 & t = 1 \\ \varepsilon_t + \rho\varepsilon_{t-1} & t > 1 \end{cases} ,$$

where $\rho \in \mathbb{R}$*. It is easily seen that* $(X_t)_{t=0,..,n}$ *is* $\mathbb{G}-$*adapted. Furthermore, for any* $t = 2, .., n$ *we have, similarly to above,*

$$E(X_t | \mathcal{G}_{t-1}) = E(\varepsilon_t + \rho\varepsilon_{t-1} | \mathcal{G}_{t-1}) = \rho E(\varepsilon_{t-1} | \mathcal{G}_{t-1}) + E(\varepsilon_t | \mathcal{G}_{t-1}) = \rho\varepsilon_{t-1} + E(\varepsilon_t) = \rho\varepsilon_{t-1} \quad .$$

Therefore,

$$X_{t-1} = E(X_t | \mathcal{G}_{t-1}) \quad \Longleftrightarrow \quad \varepsilon_{t-1} + \rho\varepsilon_{t-2} = \rho\varepsilon_{t-1} \quad ,$$

implying that in order to satisfy the martingale condition one must have for all $t = 2, .., n$

$$\begin{array}{ll} \varepsilon_{t-1} = 0 & \textit{if } \rho = 0 \\ \varepsilon_{t-1} = \frac{\rho}{\rho-1}\varepsilon_{t-2} & \textit{if } \rho \neq 0 \end{array} \quad .$$

However, this is in evident contradiction with the independence assumption on the process $(\varepsilon_t)_{t=0,..,n}$.

We thus conclude that MA(1) processes can never be martingales.

3 Pricing Principles in the Absence of Arbitrage

This section considers the pricing problem of a general European derivative in the context of an $n-$period Binomial pricing model. The model structure is:

- $I := \{0, 1, 2, .., n\}$ is a discrete time index representing the available transaction dates in the model
- The sample space is given by $\Omega :=$ {Sequences of n coordinates H or T} with single outcomes ω of the form

  $$\omega = \underbrace{(TTTH..HT)}_{n \text{ coordinates}} \quad ,$$

  for instance.
- $\mathcal{G} := \mathcal{F}$, the sigma algebra of all subsets of Ω
- Dynamics of the stock price and money account:

  $$S_t = \begin{cases} uS_{t-1} & \text{with probability } p \\ dS_{t-1} & \text{with probability } 1-p \end{cases} \quad , \quad B_t = (1+r)\,B_{t-1}$$

  for given $B_0 = 1$, S_0 and where

  $$u = 1/d \quad , \quad u > 1 + r > d \quad . \tag{12}$$

- $(\mathcal{G}_t)_{t=0,...,n}$ is the filtration generated by the stock price process $(S_t)_{t=0,...,n}$

- A binomial probability P on $(\Omega, \mathcal{F})$ is obtained by defining for some $p \in (0,1)$,

$$P(\omega) := p^{\#\text{ of H in }\omega}(1-p)^{\#\text{ of T in }\omega} \ .$$

- A binomial risk adjusted probability measure $\widetilde{P}$ on $(\Omega, \mathcal{F})$ is obtained by defining

$$\widetilde{P}(\omega) := \widetilde{p}^{\#\text{ of H in }\omega}(1-\widetilde{p})^{\#\text{ of T in }\omega} \ ,$$

  where

$$\widetilde{p} = \frac{1+r-d}{u-d} \ , \tag{13}$$

  is under condition (12) a risk adjusted probability in the one period binomial model.

The characterizing property of a risk adjusted probability measure is to make discounted stock prices under such measure martingales. Therefore, this measure is also called a risk adjusted (or risk neutral) martingale measure. Existence of a risk adjusted martingale measure is equivalent to the absence of arbitrage opportunities. We now show formally all these properties in the setting of a binomial pricing model.

3.1 Stock Prices, Risk Neutral Probability Measures and Martingales

Discounted stock prices are defined next for completeness.

Definition 74 *The stochastic process* $\left(\frac{S_t}{B_t}\right)_{t=0,..,n}$ *is called the discounted stock price process.*

The terminology of Definition 74 is obvious, since one has for any $t \in 0,..,n$ (recall the normalization $B_0 = 1$):

$$\frac{S_t}{B_t} = \frac{1}{(1+r)^t} S_t \ .$$

The next proposition shows that under the risk adjusted measure $\widetilde{P}$ discounted stock prices are martingales.

Proposition 75 *The discounted stock price process*

$$\left(\frac{S_t}{B_t}, \mathcal{G}_t\right)_{t=0,\ldots,n} ,$$

is a martingale under the probability $\widetilde{P}$.

Proof. Since B_t is deterministic, S_t/B_t is $\mathcal{G}_t-$measurable if and only if S_t is $\mathcal{G}_t-$measurable. Therefore $(S_t/B_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is an adapted process. Moreover,

$$\widetilde{E}\left(\left.\frac{S_{t+1}}{B_{t+1}}\right| \mathcal{G}_t\right) = \widetilde{E}\left(\left.\frac{S_t}{B_{t+1}} \cdot \frac{S_{t+1}}{S_t}\right| \mathcal{G}_t\right) = \frac{S_t}{B_t}\frac{1}{1+r}\widetilde{E}\left(\left.\frac{S_{t+1}}{S_t}\right| \mathcal{G}_t\right) ,$$

using the $\mathcal{G}_t-$measurability of S_t/B_{t+1}. Therefore, $(S_t/B_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is a martingale under $\widetilde{P}$ if and only if

$$\widetilde{E}\left(\left.\frac{S_{t+1}}{S_t}\right| \mathcal{G}_t\right) = 1 + r .$$

Indeed, we have

$$\widetilde{E}\left(\left.\frac{S_{t+1}}{S_t}\right| \mathcal{G}_t\right) = \widetilde{E}\left(\frac{S_{t+1}}{S_t}\right) = \widetilde{p}u + (1 - \widetilde{p})\, d = 1 + r ,$$

using in the first equality the independence of the binomial increments and in the second equality definition (13). This concludes the proof. ■

3.2 Self Financing Strategies, Risk Neutral Probability Measures and Martingales

The martingale property for discounted stock prices under $\widetilde{P}$ is valid more generally for dynamic portfolios that are self-financed, i.e. portfolios that are rebalanced at any time using only past capital gains. This fact is very important for the pricing of a derivative, because hedging portfolios are a particular case of a self-financing portfolio.

Definition 76 *(i) An adapted process* $(\Delta_t, \mathcal{G}_t)_{t=0,\ldots,n}$ *with value process* $(X_t)_{t=0,\ldots,n}$ *defines a portfolio process if* Δ_t *is the number of stock at time* t *in the portfolio and* $(X_t - \Delta_t S_t)/B_t$ *is the* *number of units of the money account. (ii) A portfolio process* $(\Delta_t, \mathcal{G}_t)_{t=0,...,n}$ *is self-financed if*

$$X_{t+1} = \Delta_t S_{t+1} + (X_t - \Delta_t S_t)(1+r) \quad .$$

(iii) A self-financed portfolio is an arbitrage opportunity if $X_0 = 0$ *and*

$$X_n \geq 0 \quad , \quad P(X_n > 0) > 0 \quad .$$

Let us remark a few important aspects on the above definition of a self financed portfolio.

Remark 77 *(i) The adaptedness condition on a portfolio process ensures that at any time* t *the number of stocks in the portfolio is determined using only information available at that time. (ii) The self-financing condition of a self-financing portfolio says that the portfolio value* X_{t+1} *at any time* $t+1$ *must be obtained as the sum of the stock and money account positions at time* t *evaluated at* $t+1$ *prices:*

$$X_{t+1} = \Delta_t \cdot S_{t+1} + (X_t - \Delta_t S_t)(1+r) = \underbrace{\Delta_t}_{\substack{\#\text{ of stocks} \\ \text{at time } t}} \cdot \underbrace{S_{t+1}}_{\substack{\text{Stock price} \\ \text{at time } t+1}} + \underbrace{\frac{X_t - \Delta_t S_t}{B_t}}_{\substack{\#\text{ of bonds} \\ \text{at time } t}} \underbrace{B_{t+1}}_{\substack{\text{Bond price} \\ \text{at time } t+1}} \quad .$$

(iii) The self-financing condition of a self-financed portfolio already implies that the value process $(X_t, \mathcal{G}_t)_{t=0,...,n}$ *of a self-financed portfolio is adapted. Indeed, for any* $t = 0, .., n-1$,

$$X_{t+1} = \Delta_t S_{t+1} + (X_t - \Delta_t S_t)(1+r) \quad ,$$

i.e. X_{t+1} *is a linear combination of random variables that are* $\mathcal{G}_{t+1}-$*measurable, and is therefore* $\mathcal{G}_{t+1}-$*measurable. (iv) An arbitrage portfolio is simply a self-financed strategy of zero initial cost and with non negative and non zero final value.*

The martingale property under $\widetilde{P}$ of discounted value processes of a self-financed portfolio is proved next.

Proposition 78 *The discounted portfolio value process*

$$\left(\frac{X_t}{B_t}, \mathcal{G}_t\right)_{t=0,...,n} \quad ,$$

of any self-financed portfolio process $(\Delta_t, \mathcal{G}_t)_{t=0,...,n}$ *is a martingale under the probability* $\widetilde{P}$.

Proof. We already showed that $(X_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is an adapted process. Therefore, it remains to show that the martingale condition for the discounted value process $(X_t/B_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is satisfied. We have:

$$\begin{aligned} \widetilde{E}\left(\left. \frac{X_{t+1}}{B_{t+1}} \right| \mathcal{G}_t \right) &= \widetilde{E}\left(\left. \frac{\Delta_t \left(S_{t+1} - S_t \left(1+r\right)\right) + X_t \left(1+r\right)}{B_{t+1}} \right| \mathcal{G}_t \right) \\ &= \frac{\Delta_t}{B_{t+1}} \widetilde{E}\left(\left. S_{t+1} - S_t \left(1+r\right) \right| \mathcal{G}_t \right) + \widetilde{E}\left(\left. \frac{X_t}{B_t} \right| \mathcal{G}_t \right) \\ &= \frac{\Delta_t}{B_{t+1}} \widetilde{E}\left(\left. S_{t+1} - S_t \left(1+r\right) \right| \mathcal{G}_t \right) + \frac{X_t}{B_t} \quad , \end{aligned}$$

using in the first equality the self-financing definition, in the second the $\mathcal{G}_t-$measurability of Δ_t/B_{t+1} and the linearity of conditional expectations, and in the third the $\mathcal{G}_t-$measurability of X_t/B_t. Thus, $(X_t/B_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is a martingale if and only if

$$\widetilde{E}\left(\left. S_{t+1} - S_t \left(1+r\right) \right| \mathcal{G}_t \right) = 0 \quad ,$$

i.e. if and only if

$$\widetilde{E}\left(\left. \frac{S_{t+1}}{S_t} \right| \mathcal{G}_t \right) = 1 + r \quad .$$

This is precisely what we have shown in the proof of Proposition 75. Therefore, the proof is completed. ■

3.3 Existence of Risk Neutral Probability Measures and Derivatives Pricing

Self financed portfolios are precisely the type of dynamic portfolios that can be used to hedge derivatives. Indeed, the self-financing condition implies that if we are able to fully replicate a contingent claim by means of a self-financed portfolio then we are also able to fully eliminate the risk deriving from the random pay-off of the contingent claim.

Definition 79 *(i) An European derivative V_T with maturity $T \in I$ is a $\mathcal{G}_T-$measurable random variable. (ii) A European derivative V_T is hedgeable if there exists a self-financed portfolio process*

$(\Delta_t, \mathcal{G}_t)_{t=0,\ldots,n}$ *with value process* $(X_t, \mathcal{G}_t)_{t=0,\ldots,n}$ *such that*

$$X_T = V_T \quad .$$

Notice that if an European contingent claim V_T is hedgeable, then absence of arbitrage opportunities immediately implies that its price is the value of the corresponding hedging portfolio. We state this important fact in the next Proposition under point (i) for completeness.

Proposition 80 *(i) If a European contingent claim* V_T *is hedgeable, then in the absence of arbitrage opportunities we have for any* $t \in I$*:*

$$V_t = X_t \quad .$$

(ii) The risk neutral valuation formula is obtained:

$$V_t = \frac{B_t}{B_T} \widetilde{E} \left(V_T | \, \mathcal{G}_t \right) = \frac{1}{(1+r)^{T-t}} \widetilde{E} \left(V_T | \, \mathcal{G}_t \right) \quad .$$

Proof. (i) Without loss of generality assume that $V_0 > X_0$, in order to imply a contradiction with the no arbitrage assumption. Then, a portfolio short in one unit of the hedging portfolio and long one unit of the derivative at time 0 costs $X_0 - V_0 < 0$. Holding the derivative until maturity and rebalancing the short position in the portfolio according to its self financing dynamics yields a pay off $X_T - V_T = 0$ at maturity T. Investing in the money account the amount $V_0 - X_0$ yields a final pay-off $(V_0 - X_0)(1+r)^T > 0$ at maturity, i.e. an arbitrage opportunity. Therefore, one must have $V_0 = X_0$. (ii) We have, by (i) and the martingale property of $(X_t, \mathcal{G}_t)_{t=0,\ldots,n}$ under $\widetilde{P}$ (see also Proposition 78):

$$\frac{V_t}{B_t} = \frac{X_t}{B_t} = \widetilde{E} \left(\left. \frac{X_T}{B_T} \right| \mathcal{G}_t \right) = \widetilde{E} \left(\left. \frac{V_T}{B_T} \right| \mathcal{G}_t \right) \quad ,$$

where the last equality arises because $(\Delta_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is an hedging portfolio for V_T. This concludes the proof. ■

The implication of the results in this section is that any hedgeable derivative has a price given by a risk neutral valuation formula. But, when is a derivative hedgeable? This is discussed in the next secion.

3.4 Uniqueness of Risk Neutral Probability Measures and Derivatives Hedging

Can a simple European derivative always be hedged? The answer depends on the pricing model used. For instance, in the standard binomial model this is the case. On the other hand, in the discrete time/continuous state space model in the next chapter this is not the case.

Basically, the answer depends on the relation between the number of basic instruments available to construct a hedging portfolio and the number of independent risk factors in the model. Roughly speaking, if the number of available instruments is sufficiently large then every contingent claim is perfectly hedgeable and thus obtains a unique price. Models that satisfy this property are called complete.

Definition 81 *In the absence of arbitrage opportunities a pricing model is complete if for any European derivative V_T there exists a hedging portfolio strategy $(\Delta_t, \mathcal{G}_t)_{t=0,\ldots,n}$ for V_T with value process $(X_t, \mathcal{G}_t)_{t=0,\ldots,n}$ and such that*

$$X_T = V_T \quad .$$

As mentioned, the standard binomial model is complete. This statement is made precise in the next result.

Theorem 82 *The binomial model is complete. Precisely, for any European derivative V_T there exists a hedging portfolio strategy $(\Delta_t, \mathcal{G}_t)_{t=0,..,T-1}$ for V_T with value process $(X_t, \mathcal{G}_t)_{t=0,..,T-1}$. For any $t = 0, .., T$ the value X_t at time t is given by*

$$X_t = B_t \widetilde{E}\left(\frac{V_T}{B_T} \middle| \mathcal{G}_t \right)$$

and for any $t = 0, .., T-1$ *the stock position* Δ_t *is given by*

$$\Delta_t = \frac{V_{t+1}(\omega_1, .., \omega_t, H) - V_{t+1}(\omega_1, .., \omega_t, T)}{S_{t+1}(\omega_1, .., \omega_t, H) - S_{t+1}(\omega_1, .., \omega_t, T)} \quad . \tag{14}$$

Proof. Define a self-financed portfolio process with stock position at time t given by Δ_t in (14) and with value process dynamics given recursively by

$$X_{t+1} = \Delta_t S_{t+1} + (X_t - \Delta_t S_t)(1+r) \quad ,$$

where

$$X_0 = \widetilde{E}\left(\left. \frac{V_T}{B_T} \right| \mathcal{G}_0 \right) \quad .$$

We show that for any $t = 0, .., T$ one has

$$X_t = V_t := \widetilde{E}\left(\left. \frac{V_T}{B_T} \right| \mathcal{G}_t \right) \quad .$$

This statement is correct by construction for $t = 0$. Thus, assume it is correct for some $t < T$, that is

$$X_t = V_t = B_t \widetilde{E}\left(\left. \frac{V_T}{B_T} \right| \mathcal{G}_t \right) \quad .$$

We show by induction that then it is correct also for $t+1$, i.e. that

$$\begin{aligned} X_{t+1}(\omega_1, .., \omega_t, H) &= V_{t+1}(\omega_1, .., \omega_t, H) \\ X_{t+1}(\omega_1, .., \omega_t, T) &= V_{t+1}(\omega_1, .., \omega_t, T) \quad . \end{aligned}$$

For brevity, we show the first of these two equalities. The second follows in a similar way. The self-financing condition gives:

$$X_{t+1}(H) = \frac{V_{t+1}(H) - V_{t+1}(T)}{S_{t+1}(H) - S_{t+1}(T)} (S_{t+1} - S_t(1+r)) + V_t(1+r) \,,$$

using the definition (14) of Δ_t. Moreover, we know that $(V_t/B_t, \mathcal{G}_t)_{t=0,\dots,n}$ is a martingale under $\widetilde{P}$, implying

$$V_t(1+r) = V_t \frac{B_{t+1}}{B_t} = \widetilde{E}\left(V_{t+1} | \, \mathcal{G}_t \right) = \widetilde{p} V_{t+1}(H) + (1-\widetilde{p}) V_{t+1}(T) \quad .$$

We thus obtain

$$\begin{aligned}
X_{t+1}(H) &= \frac{V_{t+1}(H) - V_{t+1}(T)}{S_{t+1}(H) - S_{t+1}(T)} \left(S_{t+1}(H) - S_t(1+r)\right) + \widetilde{E}\left(V_{t+1} \middle| \mathcal{G}_t\right) \\
&= \frac{V_{t+1}(H) - V_{t+1}(T)}{(u-d)\, S_t} \left(u - (1+r)\right) S_t + \widetilde{E}\left(V_{t+1} \middle| \mathcal{G}_t\right) \\
&= \frac{u-(1+r)}{(u-d)} \left(V_{t+1}(H) - V_{t+1}(T)\right) + \left(\widetilde{p} V_{t+1}(H) + (1-\widetilde{p})\, V_{t+1}(T)\right) \\
&= (1-\widetilde{p}) \left(V_{t+1}(H) - V_{t+1}(T)\right) + \left(\widetilde{p} V_{t+1}(H) + (1-\widetilde{p})\, V_{t+1}(T)\right) \\
&= V_{t+1}(H) \quad ,
\end{aligned}$$

using in the last equality the definition

$$\widetilde{p} = \frac{1+r-d}{u-d} \quad .$$

Since, for $t = T$ we obtain

$$X_T = V_T = \widetilde{E}\left(V_T \middle| \mathcal{G}_T\right) = V_T \quad ,$$

almost surely, i.e. $(\Delta_t, \mathcal{G}_t)_{t=0,..T}$ is an hedge portfolio, as claimed. This concludes the proof. ■

3.5 Existence of Risk Neutral Probability Measures and Absence of Arbitrage

4 Introduction to Stochastic Processes

For the whole section let $(\Omega, \mathcal{G}, P)$ be a probability space.

4.1 Basic Definitions

A stochastic process is a mathematical model to describe the realizations of a random experiment at some different dates defined on a time index set I, as for instance $I = \{0, 1, 2, .., n\}$, $I = \mathbb{N}$, $I = [0, \infty)$.

Definition 83 *Let I be a time index set. (i) A family $\mathcal{X} := (X_t)_{t \in I}$ of random variables on $(\Omega, \mathcal{G}, P)$ is called a stochastic process. (ii) If I is countable the process is called a discrete-time stochastic process. If it is uncountable the process is called a continuous-time stochastic process. (iii) For any $\omega \in \Omega$ the real valued function $t \longmapsto X_t(\omega)$ is called a trajectory of the process.*

Remark 84 *(i) The important thing to note in the above definition is that all random variables in the family $\mathcal{X}$ are defined on a single probability space $(\Omega, \mathcal{G}, P)$. In fact, when constructing a stochastic process satisfying a set of a priori desirable properties one will have to construct a family $\mathcal{X}$ of random variables defined on a single probability space $(\Omega, \mathcal{G}, P)$. This puts quite strong restrictions on the way how stochastic processes can be obtained. (ii) We can also think of a stochastic process $\mathcal{X}$ as a (measurable) function*

$$\mathcal{X} : \Omega \times I \to \mathbb{R} \quad ; \quad (\omega, t) \mapsto X_t(\omega) \quad ,$$

i.e. as a random variable defined on a measurable space with sample space $\Omega \times I$.

The concept of a filtration of an adapted process and of a martingale extend in a natural way to continuous time stochastic processes.

Definition 85 *(i) A family $\mathbb{G} := (\mathcal{G}_t)_{t \in I}$ of sub sigma algebras of $\mathcal{G}$ is a filtration if for any $t, s \in I$*

$$\mathcal{G}_t \subset \mathcal{G}_s \Leftrightarrow t < s \ .$$

(ii) A stochastic process $\mathcal{X} := (X_t)_{t \in I}$ is $\mathbb{G}-$adapted if for any $t \in I$ the random variable X_t is $\mathcal{G}_t-$measurable. (iii) An adapted process $(X_t, \mathcal{G}_t)_{t \in I}$ is a martingale if for any $t, s \in I$ such that $s > t$ the martingale condition

$$X_t = E\left(X_s | \mathcal{G}_t \right) \ ,$$

is satisfied.

4.2 Discrete Time Brownian Motion

A first example of a discrete time process with continuous state space is a random walk process with normally distributed innovations. This is the discrete time analogue of the continuous time Brownian motion process.

Example 86 *(Discrete Time Brownian Motion). Let* $\mathcal{Y} :=(Y_t)_{t=0,..,n}$ *be a sequence of iid* $\mathcal{N}(0,1)$ *random variables*[2] *on* $(\Omega, \mathcal{G}, P)$*. We define* $Z_0 = 0$ *and*

$$Z_t = \sum_{i=1}^{t} Y_i \quad .$$

$(Z_t)_{t=0,..,n}$ *is a random walk with normally distributed innovations and is the discrete time analogue of the (continuous time) Brownian motion process. We immediately have the following properties of discrete time Brownian motion:*

$$Z_t \backsim \mathcal{N}(0,t) \ , \tag{15}$$

i.e. Z_t *is normally distributed with mean zero and a variance increasing proportionally with time, and for* $s > t$

$$Z_s - Z_t \perp \sigma(Z_k; k \leq t) = \sigma(Y_k; k \leq t) \ , \tag{16}$$

i.e. increments of Brownian motions are independent of the past and current history of the process. Finally, we also have for any two time points $s > t$*:*

$$\begin{aligned} Cov(Z_t, Z_s) &= E(Z_t Z_s) - E(Z_t) E(Z_s) \\ &= E((Z_s - Z_t + Z_t) Z_t) \\ &= E((Z_s - Z_t) Z_t) + E\left(Z_t^2\right) \\ &= E(Z_s - Z_t) E(Z_t) + Var(Z_t) \\ &= \min(t, s) \ , \end{aligned}$$

[2] *This defines already a discrete time stochastic process.*

using in the second equality the zero mean property of Brownian motion, in the third the independence of its increments and in the fourth again the zero mean property. The same result arises for the case $t < s$. Therefore:

$$Cov\left(Z_t, Z_s\right) = \min\left(t, s\right) \ .$$

Further, define

$$\mathcal{G}_t := \sigma\left(Y_1, .., Y_t\right) \ ,$$

the sigma algebra generated by the process $\mathcal{Y}$ up to time t. Notice that we have:

$$Y_1 = Z_1 \ , \ Y_2 = Z_2 - Z_1 \ , \ Y_3 = Z_3 - Z_2, ..., Y_n = Z_n - Z_{n-1}$$

Therefore, the sigma algebras generated by $\mathcal{Y}$ and by $(Z_t)_{t=0,..,n}$ up to a given time are the same. This implies that Z_t is $\mathcal{G}_t-$measurable and that $(Z_t)_{t=0,..,n}$ is $(\mathcal{G}_t)_{t=0,..,n}$ adapted. Further, for any time indices $s > t$ we have:

$$E\left(Z_s | \mathcal{G}_t\right) = E\left(Z_s - Z_t + Z_t | \mathcal{G}_t\right) = E\left(Z_s - Z_t | \mathcal{G}_t\right) + Z_t = Z_t \ ,$$

i.e. discrete time Brownian motion is a martingale process.

Some further examples of a martingale process are obtained by looking at some simple functionals of (discrete time) Brownian motion. The first one arises simply by recentering squared (discrete time) Brownian motion by its variance.

Example 87 *Let $(Z_t, \mathcal{G}_t)_{t=0,..,n}$ be a discrete time Brownian motion. For the adapted process*

$(X_t, \mathcal{G}_t)_{t=0,..,n} := \left(Z_t^2, \mathcal{G}_t\right)_{t=0,..,n}$ *it follows for* $s > t$*:*

$$\begin{aligned} E\left(X_s \middle| \mathcal{G}_t\right) &= E\left(Z_s^2 \middle| \mathcal{G}_t\right) \\ &= E\left(\left(Z_s - Z_t + Z_t\right)^2 \middle| \mathcal{G}_t\right) \\ &= E\left(\left(Z_s - Z_t\right)^2 \middle| \mathcal{G}_t\right) + E\left(Z_t^2 \middle| \mathcal{G}_t\right) + E\left(2\left(Z_s - Z_t\right) Z_t \middle| \mathcal{G}_t\right) \\ &= E\left(\left(Z_s - Z_t\right)^2\right) + Z_t^2 + 2Z_t E\left(\left(Z_s - Z_t\right) \middle| \mathcal{G}_t\right) \\ &= (s-t) + X_t + 2Z_t E\left(\left(Z_s - Z_t\right)\right) \\ &= (s-t) + X_t \geq X_t \quad , \end{aligned}$$

using in the third equality the linearity of conditional expectations, in the fourth the independence of Brownian increments and the $\mathcal{G}_t-$*measurabiltiy of* Z_t*, and in the fifth the zero mean property of Brownian motion. This implies that* $(X_t, \mathcal{G}_t)_{t=0,..,n}$ *is a submartingale. However, by similar arguments as those listed above we see that the process* $\left(Z_t^2 - t, \mathcal{G}_t\right)_{t=0,..,n}$ *is a martingale.*

The last example of a Brownian functional that gives a martingale process is exponential (discrete time) Brownian motion.

Example 88 *(Exponential Brownian Motion) Let* $(Z_t, \mathcal{G}_t)_{t=0,..,n}$ *be a discrete time Brownian motion. For the adapted process* $(X_t, \mathcal{G}_t)_{t=0,..,n}$ *defined by*

$$X_t = \exp\left(\sigma Z_t - \frac{\sigma^2 t}{2}\right)$$

it follows for $s > t$*:*

$$\begin{aligned} E\left(X_s \middle| \mathcal{G}_t\right) &= E\left(\exp\left(\sigma Z_s - \frac{\sigma^2 s}{2}\right) \middle| \mathcal{G}_t\right) \\ &= E\left(\exp\left(\sigma\left(Z_s - Z_t\right) - \frac{\sigma^2\left(s-t\right)}{2}\right) \exp\left(\sigma Z_t - \frac{\sigma^2 t}{2}\right) \middle| \mathcal{G}_t\right) \\ &= \exp\left(\sigma Z_t - \frac{\sigma^2 t}{2}\right) E\left(\exp\left(\sigma\left(Z_s - Z_t\right) - \frac{\sigma^2\left(s-t\right)}{2}\right) \middle| \mathcal{G}_t\right) \\ &= X_t \exp\left(-\frac{\sigma^2\left(s-t\right)}{2}\right) E\left(\exp\sigma\left(Z_s - Z_t\right)\right) \quad , \end{aligned} \tag{17}$$

using in the last equality the independence of Brownian increments. Now, since $Z_s - Z_t \backsim \mathcal{N}(0, s-t)$*, the expression* $E\left(\exp\sigma\left(Z_s - Z_t\right)\right)$ *is the moment generating function of a* $\mathcal{N}(0, s-t)$

distributed random variable, evaluated at the point σ. Thus,

$$E(\exp\sigma(Z_s - Z_t)) = M_{Z_s - Z_t}(\sigma) = \exp\left(\frac{\sigma^2(s-t)}{2}\right) \quad .$$

With this result, we obtain in (17) the martingale property for the process $(X_t, \mathcal{G}_t)_{t=0,..,n}$.

4.3 Girsanov Theorem: Application to a Semicontinuous Pricing Model

This section considers the pricing problem of a general European derivative in the context of an $n-$period semicontinuous pricing model.

4.3.1 A Semicontinuous Pricing Model

The model structure is:

- $I := \{0, 1, 2, .., n\}$ is a discrete time index representing the available transaction dates in the model
- The sample space is given by $\Omega := \mathbb{R}^n$ with single outcomes ω of the form

  $$\omega = (\omega_1, \omega_2, .., \omega_n) \quad ,$$

  where $\omega_i \in \mathbb{R}$, $i = 1, \ldots, n$.
- $\mathcal{G} := \mathcal{B}(\mathbb{R}^n)$ the Borel sigma algebra on $\mathbb{R}^n$
- Dynamics of the stock price and money account:

  $$\begin{aligned} S_t &= S_{t-1}\exp\left(\sigma Y_t - \frac{\sigma^2}{2}\right)\exp(\mu) \quad , \qquad (18) \\ B_t &= \exp(r) B_{t-1} \quad , \end{aligned}$$

  for some μ, r, $\sigma > 0$, for given $B_0 = 1$, S_0 and where $(Y_t)_{t=1,\ldots,n}$ is an iid $\mathcal{N}(0,1)$ sequence of random variables on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
- $(\mathcal{G}_t)_{t=0,\ldots,n}$ is the filtration generated by $(Y_t)_{t=1,\ldots,n}$, which coincides with the filtration generated by the stock price process $(S_t)_{t=0,\ldots,n}$.

- A probability P on $(\Omega, \mathcal{G})$ such that $(Y_t)_{t=1,\dots,n}$ is an iid $\mathcal{N}(0,1)$ sequence.

Some simple properties of the above asset price dynamics can be immediately deduced from the above definitions. Firstly, the above money account dynamics gives:

$$B_t = \exp(rt) \quad ,$$

implying a continuous interest rate compounding. Secondly, for the risky asset price dynamics we get:

$$\begin{aligned} E\left(\left.\frac{S_s}{S_t}\right| \mathcal{G}_t\right) &= E\left(\exp\left(\sigma \sum_{i=t+1}^{s} Y_i - \frac{\sigma^2(s-t)}{2}\right)\exp(\mu(s-t))\right| \mathcal{G}_t\Bigg) \\ &= E\left(\exp\left(\sigma(Z_s - Z_t) - \frac{\sigma^2(s-t)}{2}\right)\right| \mathcal{G}_t\Bigg)\exp(\mu(s-t)) \\ &= \exp(\mu(s-t)) \quad , \end{aligned}$$

because exponential Brownian motion is a martingale process. Therefore, $\exp(\mu(s-t))$ is the expected rate of return on the stock, or alternatively

$$\mu = \frac{\log E(S_s|\mathcal{G}_t) - \log S_t}{s-t} \quad ,$$

is the continuous expected rate of returns on the stock. Similarly, for the variance of logarithmic stock returns one gets

$$Var\left(\log\left(\frac{S_s}{S_t}\right)\middle| \mathcal{G}_t\right) = Var\left(\left(\sigma(Z_s - Z_t) - \frac{\sigma^2(s-t)}{2}\right)\middle| \mathcal{G}_t\right) = \sigma^2 Var(Z_s - Z_t) = \sigma^2(s-t) \ ,$$

i.e.

$$\sigma^2 = \frac{Var\left(\log\left(\frac{S_s}{S_t}\right)\middle| \mathcal{G}_t\right)}{s-t} \quad ,$$

is the continuous variance rate on the stock.

4.3.2 Risk Neutral Valuation in the Semicontinuous Model

In the semicontinuous model it is not possible to hedge perfectly any European derivative using a suitable self-financed hedge portfolio, i.e. this model is not complete. Therefore, the derivation/computation of a suitable risk neutral probability measure for pricing derivatives must follow other arguments than those adopted in the binomial model setting. Fortunately, a powerful theorem from the theory of stochastic process, namely Girsanov Theorem, can assist us in constructing such probability measures by using pure probabilistic arguments. We first give for completeness some basic definitions which are the pendant of Definition 76 for the semicontinuous model setting

Definition 89 *(i) An adapted process $(\Delta_t, \mathcal{G}_t)_{t=0,\ldots,n}$ with value process $(X_t)_{t=0,\ldots,n}$ defines a portfolio process if Δ_t is the number of stock at time t in the portfolio and $(X_t - \Delta_t S_t)/B_t$ is the number of units of the money account. (ii) A portfolio process $(\Delta_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is self-financed if*

$$X_{t+1} = \Delta_t S_{t+1} + (X_t - \Delta_t S_t) \exp(r) \quad .$$

(iii) A self-financed portfolio is an arbitrage opportunity if $X_0 = 0$ and

$$X_n \geq 0 \quad , \quad P(X_n > 0) > 0 \quad .$$

(iv) A probability $\widetilde{P}$ on the measurable space $(\Omega, \mathcal{G})$ which is equivalent to P and such that the discounted price process $(S_t/B_t, \mathcal{G}_t)_{t=0,\ldots,n}$ is a martingale, is called a risk adjusted (risk neutral) probability measure.

Remark 90 *Notice that self-financed portfolios and arbitrage strategies in the semicontinuous model are defined exactly as in the earlier binomial setting, with the only difference that now interest rates are continuously compounded. By contrast, a risk neutral probability measure $\widetilde{P}$ in the semicontinuous model is explicitly required to be equivalent to the physical probability P. This ensures that the null sets of these two probabilities coincide. This property was by construction satisfied in the binomial model, where no null probability events - apart from the trivial empty set - could arise.*

To highlight the properties that a risk neutral measure in the semicontinuous model should have we consider again the discounted price process $(S_t/B_t, \mathcal{G}_t)_{t=0,\ldots,n}$, which is given by

$$\frac{S_{t+1}}{B_{t+1}} = \frac{S_t \exp\left(\sigma Y_{t+1} - \frac{\sigma^2}{2} + \mu\right)}{B_t \exp(r)} = \frac{S_t}{B_t} \exp\left(\sigma Y_{t+1} - \frac{\sigma^2}{2} + \mu - r\right) \quad .$$

Therefore,

$$\begin{aligned}\widetilde{E}\left(\left.\frac{S_{t+1}}{B_{t+1}}\right|\mathcal{G}_t\right) &= \frac{S_t}{B_t}\widetilde{E}\left(\exp\left(\sigma Y_{t+1}-\frac{\sigma^2}{2}+\mu-r\right)\Big|\,\mathcal{G}_t\right)\\ &= \frac{S_t}{B_t}\widetilde{E}\left(\exp\left(\sigma\left(Y_{t+1}+\frac{\mu-r}{\sigma}\right)\right)\Big|\,\mathcal{G}_t\right)\exp\left(-\frac{\sigma^2}{2}\right)\\ &= \frac{S_t}{B_t}\widetilde{E}\left(\exp\left(\sigma\left(Y_{t+1}+\theta\right)\right)|\,\mathcal{G}_t\right)\exp\left(-\frac{\sigma^2}{2}\right)\ ,\end{aligned} \tag{19}$$

where

$$\theta=\frac{\mu-r}{\sigma}\ ,$$

is the so called market price of risk. Recall that if under $\widetilde{P}$ the random variable $\widetilde{Y}_{t+1}:=Y_{t+1}+\theta$ is both standard normally distributed and independent of $\mathcal{G}_t$ then:

$$\widetilde{E}\left(\exp\left(\sigma\left(Y_{t+1}+\theta\right)\right)|\,\mathcal{G}_t\right)=\widetilde{E}\left(\exp\left(\sigma\widetilde{Y}_{t+1}\right)\right)=\exp\left(\frac{\sigma^2}{2}\right)\ ,$$

and the martingale property follows for time t. Thus, if under a probability $\widetilde{P}$ the process $\left(\widetilde{Y}_t\right)_{t=1,\ldots,n}=\left(Y_t+\theta\right)_{t=1,\ldots,n}$ is an iid $\mathcal{N}(0,1)$ random sequence, then under $\widetilde{P}$ the discounted stock price process is a martingale and $\widetilde{P}$ is a risk neutral measure. This is equivalent to stating that under $\widetilde{P}$ the process $\left(\widetilde{Z}_t\right)_{t=0,..,n}$ given by

$$\widetilde{Z}_t=Z_t+\theta t=\sum_{i=1}^{t}\left(Y_i+\theta\right)=\sum_{i=1}^{t}\widetilde{Y}_i\ ,$$

is a discrete time Brownian motion. Notice, that under the physical probability P the process $\left(\widetilde{Z}_t\right)_{t=0,..,n}$ is not a Brownian motion but a so called Brownian motion with drift. Therefore, the probability $\widetilde{P}$ "reconverts " a process which is a Brownian motion with drift under P to a standard Brownian motion.

How can we construct such a probability measure $\widetilde{P}$? The answer is provided by Girsanov's Theorem.

4.3.3 A Discrete Time Formulation of Girsanov Theorem

A discrete time formulation of the famous Girsanov Theorem is proved in the sequel.

Theorem 91 *Let $(\Omega, \mathcal{G}, P)$ be a probability space such that the process $(Y_t)_{t=1,\ldots,n}$ is an iid $\mathcal{N}(0,1)$ random sequence (or equivalently the process $(Z_t)_{0=1,\ldots,n}$ is a discrete time Brownian motion). Define a further measure $\widetilde{P}$ on $(\Omega, \mathcal{G})$ by*

$$\widetilde{P}(A) = \int_A \frac{d\widetilde{P}}{dP} dP \quad , \quad A \in \mathcal{G} \quad , \tag{20}$$

where

$$\frac{d\widetilde{P}}{dP} = \exp\left(-\theta \sum_{t=1}^{n} Y_i - \frac{\theta^2 n}{2}\right) = \exp\left(-\theta Z_n - \frac{\theta^2 n}{2}\right) \quad . \tag{21}$$

It then follows:

1. *$\widetilde{P}$ is a probability measure equivalent to P,*
2. *The process $\left(\widetilde{Y}_t\right)_{t=0,..,n} = (Y_t + \theta)_{t=1,\ldots,n}$ is an iid $\mathcal{N}(0,1)$ random sequence under $\widetilde{P}$,*
3. *The process $\left(\widetilde{Z}_t\right)_{t=0,..,n} = (Z_t + \theta t)_{t=0,\ldots,n}$ is a discrete time Brownian motion under $\widetilde{P}$.*

Proof. 1. By the properties of Lebesgue integrals $\widetilde{P}$ is a measure on $(\Omega, \mathcal{G})$. Moreover, we have

$$\int_\Omega \frac{d\widetilde{P}}{dP} dP = E\left(\frac{d\widetilde{P}}{dP}\right) = E\left(\exp\left(-\theta Z_n - \frac{\theta^2 n}{2}\right)\right) = E\left(\exp\left(-\theta Z_n\right)\right)\exp\left(-\frac{\theta^2 n}{2}\right) \quad .$$

Recall that $Z_n \backsim \mathcal{N}(0,n)$ under P. Therefore,

$$E\left(\exp\left(-\theta Z_n\right)\right) = \exp\left(\frac{\theta^2 n}{2}\right) \quad ,$$

by the properties of moment generating functions of normally distributed random variables, implying $\int_\Omega \frac{d\widetilde{P}}{dP} dP = 1$. Thus, $\frac{d\widetilde{P}}{dP}$ is a strictly positive proper density and $\widetilde{P}$ is a probability measure equivalent to P. 2. To show that under $\widetilde{P}$ the random sequence $\left(\widetilde{Y}_t\right)_{t=1,\ldots,n} := (Y_t + \theta)_{t=1,\ldots,n}$ is iid $\mathcal{N}(0,1)$ let us denote by $\widetilde{\mathcal{L}}_{\widetilde{Y}_1,..,\widetilde{Y}_n}$ and $\widetilde{\mathcal{L}}_{Y_1,..,Y_n}$ ($\mathcal{L}_{\widetilde{Y}_1,..,\widetilde{Y}_n}$ and $\mathcal{L}_{Y_1,..,Y_n}$) the distribution induced by $\widetilde{Y}_1,..,\widetilde{Y}_n$ and $Y_1,..,Y_n$, respectively, on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ under $\widetilde{P}$ (under P), that is

$$\begin{aligned}
\widetilde{\mathcal{L}}_{\widetilde{Y}_1,..,\widetilde{Y}_n}(B) &= \widetilde{P}\left(\left(\widetilde{Y}_1,..,\widetilde{Y}_n\right) \in B\right) \quad , \quad \widetilde{\mathcal{L}}_{Y_1,..,Y_n}(B) = \widetilde{P}\left((Y_1,..,Y_n) \in B\right) \quad , \\
\mathcal{L}_{\widetilde{Y}_1,..,\widetilde{Y}_n}(B) &= P\left(\left(\widetilde{Y}_1,..,\widetilde{Y}_n\right) \in B\right) \quad , \quad \mathcal{L}_{Y_1,..,Y_n}(B) = P\left((Y_1,..,Y_n) \in B\right) \quad ,
\end{aligned}$$

for any $B \in \mathcal{B}(\mathbb{R}^n)$. We have, for any $B \in \mathcal{B}(\mathbb{R}^n)$:

$$\widetilde{\mathcal{L}}_{\widetilde{Y}_1,..,\widetilde{Y}_n}(B) = \widetilde{P}\left(\left(\widetilde{Y}_1,..,\widetilde{Y}_n\right) \in B\right) = \widetilde{P}\left((Y_1,..,Y_n) \in B - \theta\right) = \widetilde{\mathcal{L}}_{Y_1,..,Y_n}(B - \theta) \quad ,$$

where $B - \theta := \{(x_1 - \theta, .., x_n - \theta) : x \in B\}$, because for any $i = 1, \ldots, n$, it follows $Y_i = \widetilde{Y}_i - \theta$. Further, recall that the joint distribution of $(Y_t)_{t=1,\ldots,n}$ under P is iid $\mathcal{N}(0,1)$, i.e.

$$\mathcal{L}_{Y_1,..,Y_n}(B) = \int_B (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^n y_t^2\right) dy_1 \cdots dy_n \quad .$$

Therefore, we obtain

$$\begin{aligned}
\widetilde{\mathcal{L}}_{Y_1,..,Y_n}(B-\theta) &= \int_{B-\theta} \exp\left(-\theta\sum_{t=1}^n y_t - \frac{\theta^2 n}{2}\right) d\mathcal{L}_{Y_1,..,Y_n}(y_1,..,y_n) \\
&= \int_{B-\theta} \exp\left(-\theta\sum_{t=1}^n y_t - \frac{\theta^2 n}{2}\right) (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^n y_t^2\right) dy_1 \cdots dy_n \\
&= \int_{B} \exp\left(-\theta\sum_{t=1}^n (\widetilde{y}_t - \theta) - \frac{\theta^2 n}{2}\right) (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^n (\widetilde{y}_t - \theta)^2\right) d\widetilde{y}_1 \cdots d\widetilde{y}_n \\
&= \int_{B} (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^n \widetilde{y}_t^2\right) d\widetilde{y}_1 \cdots d\widetilde{y}_n \quad .
\end{aligned}$$

Since the function

$$f(\widetilde{y}) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^n \widetilde{y}_t^2\right) \quad ,$$

is the joint density function of an iid sequence of $\mathcal{N}(0,1)$ random variables we have shown that $\widetilde{\mathcal{L}}_{\widetilde{Y}_1,..,\widetilde{Y}_n}$ is such a normal distribution on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, as claimed. Statement 3. follows directly from statement 2. ■

Using Girsanov theorem, we are now able to give a risk neutral probability measure for the above semicontinuous pricing model. We summarize this finding in the next corollary.

Corollary 92 *In the semicontinous pricing model with stock price dynamics (18) a risk adjusted martingale measure* $\widetilde{P}$ *on* $(\Omega, \mathcal{G})$ *is obtained by setting for any* $A \in \mathcal{G}$,

$$\widetilde{P}(A) := \int_A \exp\left(-\theta Z_n - \frac{\theta^2 n}{2}\right) dP \quad ,$$

where

$$\theta = \frac{\mu - r}{\sigma} \quad ,$$

is the market price of risk in the model.

Inspired by the previous results in the Binomial model, we are now tempted to price European derivatives also in the semicontinuous model by means of a risk neutral valuation formula under $\widetilde{P}$. Notice, that since the model is incomplete there could exist more than one risk neutral probability measure for this setting, in excess to the just identified probability measure $\widetilde{P}$. This causes the problem of finding adequate criteria for selecting one of these probabilities to price contingent claims in incomplete markets. Nevertheless, using $\widetilde{P}$ we can still compute, at least formally, the corresponding risk neutral pricing formula as a specific expectation. Moreover, we can *define* the $t-$time price of an European derivative V_T in the semicontinuous model as

$$V_t := \frac{B_t}{B_T}\widetilde{E}\left(V_T|\,\mathcal{G}_t\right) = \exp\left(-r\left(T-t\right)\right)\widetilde{E}\left(V_T|\,\mathcal{G}_t\right) \quad .$$

In the case of a call pay-off $V_T = (S_T - K)^+$ this yields the famous Black-Scholes pricing formula using pure probabilistic arguments. In order to motivate this pricing formula completely, we will have to work out - in a later section - a model where trading can evolve in continuous time, the Black and Scholes model. In this setting we will also be able to construct hedging strategies for any European derivative in the model and to show that the above pricing approach is the only one consistent with the absence of arbitrage opportunities in the Black and Scholes model. To this end we will have to introduce some continuous time stochastic processes more explicitly and to develop a stochastic integral calculus, Itô's calculus, where integrals are defined with respect to increments of a continuous time Brownian motion (see below).

Before doing that, we conclude this section by computing Black-Scholes formula in the semi-continuous model by means of a risk neutral valuation formula under $\widetilde{P}$.

4.3.4 A Discrete Time Derivation of Black and Scholes Formula

The famous Black and Scholes pricing formula for an European call option arises in our semicontinuous setting as the discouted risk neutral expectation of the call pay-off at maturity. This is shown in the next result.

Proposition 93 *(Black and Scholes Call Price Formula) The time* 0 *discounted risk neutral expectation of the call pay-off* $(S_T - K)^+$ *of a call option with maturity* T *and strike price* K *is given by:*

$$\frac{B_0}{B_T}\widetilde{E}\left((S_T - K)^+\right) = S_0\mathcal{N}(d_1) - K\exp(-rT)\mathcal{N}(d_2) \quad ,$$

where

$$d_1 = \frac{\log\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \quad , \quad d_2 = d_1 - \sigma\sqrt{T} \quad .$$

Proof. Writing

$$(S_T - K)^+ = (S_T - K)\,\mathbf{1}_{\{S_T > K\}} = S_T\mathbf{1}_{\{S_T > K\}} - K\mathbf{1}_{\{S_T > K\}} \quad ,$$

we have

$$\widetilde{E}\left((S_T - K)^+\right) = \widetilde{E}\left((S_T - K)\,\mathbf{1}_{\{S_T > K\}}\right) = \widetilde{E}\left(S_T\mathbf{1}_{\{S_T > K\}}\right) - K\widetilde{E}\left(\mathbf{1}_{\{S_T > K\}}\right) \quad . \tag{22}$$

Moreover,

$$\widetilde{E}\left(\mathbf{1}_{\{S_T > K\}}\right) = \widetilde{P}(S_T > K) \quad .$$

To compute this probability notice that we have:

$$\{S_T > K\} = \left\{\log\frac{S_T}{S_0} > \log\frac{K}{S_0}\right\} \quad .$$

Moreover, using the explicit form for the stock price dynamics in the semicontinuous model,

$$\log\frac{S_T}{S_0} = \sigma Z_T - \left(\frac{\sigma^2}{2} - \mu\right)T = \sigma\left(Z_T + \theta T\right) - \left(\frac{\sigma^2}{2} - r\right)T \quad ,$$

where $\theta = (\mu - r)/\sigma$, it follows

$$\begin{aligned}
\{S_T > K\} &= \left\{\sigma\left(Z_T + \theta T\right) > \log\frac{K}{S_0} + \left(\frac{\sigma^2}{2} - r\right)T\right\} \\
&= \left\{\frac{Z_T + \theta T}{\sqrt{T}} > \frac{\log\frac{K}{S_0} + \left(\frac{\sigma^2}{2} - r\right)T}{\sigma\sqrt{T}}\right\} \\
&= \left\{-\frac{Z_T + \theta T}{\sqrt{T}} < \frac{\log\frac{S_0}{K} - \left(\frac{\sigma^2}{2} - r\right)T}{\sigma\sqrt{T}}\right\} \\
&= \left\{-\frac{Z_T + \theta T}{\sqrt{T}} < d_1 - \sigma\sqrt{T}\right\} \\
&= \left\{-\frac{Z_T + \theta T}{\sqrt{T}} < d_2\right\} \quad .
\end{aligned}$$

Now, notice that by Girsanov theorem $Z_T + \theta T$ is a standard Brownian motion under the probability measure $\widetilde{P}$, so that

$$-\frac{Z_T + \theta T}{\sqrt{T}} \backsim \mathcal{N}(0,1) \quad .$$

Therefore,

$$\widetilde{P}\left(S_T > K\right) = \widetilde{P}\left(-\frac{Z_T + \theta T}{\sqrt{T}} < d_2\right) = \mathcal{N}\left(d_2\right) \quad . \tag{23}$$

We now compute the first term in the difference on the RHS of (22), discounted by $B_T = \exp(rT)$. We have:

$$\begin{aligned}
\widetilde{E}\left(\frac{S_T}{B_T}\mathbf{1}_{\{S_T > K\}}\right) &= \int_{\{S_T > K\}} \frac{S_T}{B_T} d\widetilde{P} \\
&= S_0 \int_{\{S_T > K\}} \frac{S_T}{S_0 B_T} d\widetilde{P} \\
&= S_0 \int_{\{S_T > K\}} \exp\left(\sigma\left(Z_T + \theta T\right) - \frac{\sigma^2 T}{2}\right) d\widetilde{P}
\end{aligned}$$

Now, recall that (see above)

$$\{S_T > K\} = \left\{Z_T + \theta T > -d_2\sqrt{T}\right\} \quad ,$$

and that under $\widetilde{P}$ we have

$$Z_T + \theta T \backsim \mathcal{N}(0, T) \quad .$$

This gives

$$\begin{aligned}
\widetilde{E}\left(\frac{S_T}{B_T}\mathbf{1}_{\{S_T>K\}}\right) &= S_0\int_{\{S_T>K\}}\exp\left(\sigma\left(Z_T+\theta T\right)-\frac{\sigma^2 T}{2}\right)d\widetilde{P} \\
&= S_0\int_{-d_2\sqrt{T}}^{\infty}\exp\left(\sigma z-\frac{\sigma^2 T}{2}\right)\frac{1}{\sqrt{2\pi T}}\exp\left(-\frac{1}{2}\left(\frac{z}{\sqrt{T}}\right)^2\right)dz \\
&= S_0\int_{-d_2\sqrt{T}}^{\infty}\frac{1}{\sqrt{2\pi T}}\exp\left(-\frac{1}{2}\left(\frac{z-\sigma T}{\sqrt{T}}\right)^2\right)dz \quad . \qquad (24)
\end{aligned}$$

Finally, notice that

$$\int_{-d_2\sqrt{T}}^{\infty}\frac{1}{\sqrt{2\pi T}}\exp\left(-\frac{1}{2}\left(\frac{z-\sigma T}{\sqrt{T}}\right)^2\right)dz$$

is the probability of the interval $\left[-d_2\sqrt{T},\infty\right)$ under a $\mathcal{N}\left(\sigma T, T\right)$ distribution, which is the same as the probability of the interval $\left[-d_2-\sigma\sqrt{T},\infty\right)$ under a $\mathcal{N}\left(0,1\right)$ distribution. By symmetry, this is also equal to the probability of the interval $\left(-\infty, d_2+\sigma\sqrt{T}\right]$ under a $\mathcal{N}\left(0,1\right)$ distribution, i.e.:

$$S_0\int_{-d_2\sqrt{T}}^{\infty}\frac{1}{\sqrt{2\pi T}}\exp\left(-\frac{1}{2}\left(\frac{z-\sigma T}{\sqrt{T}}\right)^2\right)dz = S_0\mathcal{N}\left(d_2+\sigma\sqrt{T}\right) = S_0\mathcal{N}\left(d_1\right) \quad . \qquad (25)$$

Putting terms together we finally obtain:

$$\frac{B_0}{B_T}\widetilde{E}\left(\left(S_T-K\right)^+\right) = \widetilde{E}\left(\frac{S_T}{B_T}\mathbf{1}_{\{S_T>K\}}\right) - \frac{K}{B_T}\widetilde{E}\left(\mathbf{1}_{\{S_T>K\}}\right) = S_0\mathcal{N}\left(d_1\right) - K\exp\left(-rT\right)\mathcal{N}\left(d_2\right) \quad ,$$

from (23) and (25). ■

4.4 Continuous Time Brownian Motion

The starting point to develop a continuous time model for the stock price is the Brownian motion process.

Definition 94 *A continuous time adapted process $\left(Z_t, \mathcal{G}_t\right)_{t\geq 0}$ on a probability space $\left(\Omega, \mathcal{G}, P\right)$ is a (standard) Brownian motion if*

1. $Z_0 = 0$

2. For any $s > t$ if follows

$$Z_s - Z_t \backsim \mathcal{N}(0, s-t) \quad , \quad Z_s - Z_t \perp \mathcal{G}_t \quad .$$

3. For any $\omega \in \Omega$ the mapping $t \longmapsto Z_t(\omega)$ is continuous.

We shall speak sometimes of Brownian motion $(Z_t, \mathcal{G}_t)_{0 \leq t \leq T}$ on $[0, T]$ for some $T > 0$ and the meaning of this is apparent.

Remark 95 *(i) The filtration $(\mathcal{G}_t)_{t \geq 0}$ is part of the definition. A natural choice of a filtration is the one generated by the process coordinates, defined by setting*

$$\mathcal{G}_t = \mathcal{G}_t^Z := \sigma(\sigma(Z_u); 0 \leq u \leq t) \ .$$

In some cases[3] , it is important to work with a larger filtration that the one generated by the process. In the sequel we will assume $\mathcal{G}_t$ to be at least augmented, i.e., for any $t \geq 0$:

$$\mathcal{G}_t = \sigma\left(\mathcal{G}_t^Z \cup \mathcal{N}\right) \quad , \tag{26}$$

where $\mathcal{N}$ is the family of all subsest of $\mathcal{G}$ having probability 0. Notice, that $\mathcal{G}_t^Z$ does not contain $\mathcal{N}$, so that $\left(\mathcal{G}_t^Z\right)_{t \geq 0}$ is not augmented. The augmented filtration implied by (26) is sometimes called the natural filtration of a Brownian motion process. (ii) The fact that a probability space $(\Omega, \mathcal{G}, P)$ and an adapted stochastic process $(Z_t, \mathcal{G}_t)_{t \geq 0}$ with the Brownian motion properties indeed can be constructed is a fundamental result in probability theory. (iii) It can be shown that Brownian motion is the only process with continuous paths and with independent stationary increments, i.e satisfying

$$Z_s - Z_t \perp \mathcal{G}_t \quad \text{and} \quad \mathcal{L}_{Z_s - Z_t} = \mathcal{L}_{Z_{s-t} - Z_0} \quad ,$$

[3] *As for instance when constructing solutions to some particular stochastic differential equations.*

for any $s \geq t$. (iv) It can also be shown that Brownian motion is the only martingale with continuous paths such that for any $0 \leq t \leq s$:

$$E\left((Z_s - Z_t)^2 \middle| \mathcal{G}_t \right) = s - t \quad .$$

The finite dimensional distributions of a Brownian motion process are easily obtained from the definition.

Proposition 96 *For any finite index set $0 \leq t_1 < t_2 < ... < t_n$ the finite dimensional distribution of the random vector $(Z_{t_1}, Z_{t_2}, .., Z_{t_n})'$ is Gaussian,*

$$\mathcal{L}_{Z_{t_1}, Z_{t_2}, .., Z_{t_n}} = \mathcal{N}(0, \Sigma) \ ,$$

where

$$\Sigma = \begin{bmatrix} t_1 & \cdots & t_1 & t_1 \\ \vdots & t_2 & \cdots & t_2 \\ t_1 & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_n \end{bmatrix} .$$

Proof. We have

$$\begin{pmatrix} Z_{t_1} \\ Z_{t_2} \\ \vdots \\ Z_{t_n} \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{\Lambda} \begin{pmatrix} Z_{t_1} - Z_0 \\ Z_{t_2} - Z_{t_1} \\ \vdots \\ Z_{t_n} - Z_{t_{n-1}} \end{pmatrix} ,$$

i.e. any finite vector of coordinates of a Brownian motion can be written as a linear function of a vector of Gaussian Brownian increments, and is thus also Gaussian with expectation 0 and covariance matrix

$$\Sigma = \Lambda D \Lambda' \ ,$$

where

$$D = \begin{bmatrix} t_1 & 0 & 0 & \cdots & 0 \\ 0 & t_2 - t_1 & 0 & \cdots & 0 \\ 0 & 0 & t_3 - t_2 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & t_n - t_{n-1} \end{bmatrix}$$

Explicit computation of Σ concludes the proof. ■

In some cases, transformations of a Brownian motion give again a Brownian motion process. Here are some well-known examples.

Example 97 *Let $(Z_t, \mathcal{G}_t)_{t\geq 0}$ be a standard Brownian motion on a probability space $(\Omega, \mathcal{G}, P)$. The following transformations give again a Brownian motion:*

1. *Symmetry: $(-Z_t, \mathcal{G}_t)_{t\geq 0}$*
2. *Scaling: $\left(c^{-\frac{1}{2}} Z_{ct}, \mathcal{G}_{ct}\right)_{t\geq 0}$*
3. *Time reversal: For given $T > 0$, the process $\left(V_t, \mathcal{G}_t^V\right)_{0\leq t\leq T}$ defined by*

$$V_t = Z_T - Z_{T-t} \quad , \quad \mathcal{G}_t^V := \sigma\left(V_u; 0 \leq u \leq t\right) \quad .$$

For instance, to show 3. notice first that for any $\omega \in \Omega$ the map $t \longmapsto V_t(\omega) = Z_T(\omega) - Z_{T-u}(\omega)|_{u=t}$ is continuous, because it consists of sums and compositions of continuous functions of t. Further, by definition $V_0 = Z_T - Z_{T-0} = 0$ and any finite dimensional distribution of $(V_t)_{0\leq t\leq T}$ is Gaussian since coordinates of $(V_t)_{0\leq t\leq T}$ arise as simple linear transformations of coordinates of $(Z_t)_{0\leq t\leq T}$. Finally, for any $t \leq T$

$$E(V_t) = E(Z_T - Z_{T-t}) = 0 \quad ,$$

and for any $0 \leq u \leq t \leq s \leq T$

$$\begin{aligned} Cov\left(V_s, V_u\right) &= Cov\left(Z_T - Z_{T-s}, Z_T - Z_{T-u}\right) \\ &= T + (T-s) \wedge (T-u) - T \wedge (T-u) - (T-s) \wedge T \\ &= T + (T-s) - (T-u) - (T-s) = u \quad . \end{aligned}$$

In particular, then

$$Cov\left(V_s - V_u, V_u\right) = 0 \quad ,$$

for any $0 \leq u \leq t$. *Since any finite dimensional distribution of* $(V_t)_{0 \leq t \leq T}$ *is Gaussian this implies* $V_s - V_t \perp \sigma\left(V_u; 0 \leq u \leq t\right) = \mathcal{G}_t^V$. *Thus,* $\left(V_t, \mathcal{G}_t^V\right)_{0 \leq t \leq T}$ *satisfies the definition of a Brownian motion process.*

As in the discrete time setting, some simple examples of continuous martingales are obtained by considering some specific functionals of Brownian motion. For completeness, we give in the next results two examples that are the most relevant to our exposition.

Example 98 *Let* $(Z_t, \mathcal{G}_t)_{t \geq 0}$ *be a standard Brownian motion on a probability space* $(\Omega, \mathcal{G}, P)$. *Then the processes*

1. $\left(Z_t^2 - t, \mathcal{G}_t\right)_{t \geq 0}$
2. $\left(\exp\left(\sigma Z_t - \frac{\sigma^2 t}{2}\right), \mathcal{G}_t\right)_{t \geq 0}$

are both martingales.

Proof. It is obvious that both processes are $(\mathcal{G}_t)_{t \geq 0}$ −adapted, since they are both simple measurable functions of Brownian motion. Moreover, the proof of the martingale property is obtained readily with the same arguments as for the proof of the discrete time case in Example 87 and 88 ■

5 Introduction to Stochastic Calculus

For the whole chapter let $(Z_t, \mathcal{G}_t)_{t\geq 0}$ be a Brownian motion on a probability space $(\Omega, \mathcal{G}, P)$.

5.1 Starting Point, Motivation

To motivate the introduction of a stochastic integral consider again the discrete time dynamics for the discounted value process $(X_t/B_t)_{t=0,..,n}$ of a self-financed portfolio $(\Delta_t)_{t=0,..,n-1}$:

$$\begin{aligned}
\frac{X_{t+1}}{B_{t+1}} &= \Delta_t \frac{S_{t+1}}{B_{t+1}} + \frac{(X_t - \Delta_t S_t)}{B_{t+1}} \cdot \frac{B_{t+1}}{B_t} \\
&= \Delta_t \left(\frac{S_{t+1}}{B_{t+1}} - \frac{S_t}{B_t} \right) + \frac{X_t}{B_t} \\
&= \Delta_t \left(\frac{S_{t+1}}{B_{t+1}} - \frac{S_t}{B_t} \right) + \Delta_{t-1} \left(\frac{S_t}{B_t} - \frac{S_{t-1}}{B_{t-1}} \right) + \frac{X_{t-1}}{B_{t-1}} \\
&= ... = \sum_{i=0}^{t} \Delta_i \left(\frac{S_{i+1}}{B_{i+1}} - \frac{S_i}{B_i} \right) + \frac{X_0}{B_0}
\end{aligned}$$

Recall, that under the risk neutral measure $\widetilde{P}$ the discounted price process $(S_t/B_t)_{t=0,..,n}$ is a martingale. Thus, we have represented the time variation of the discounted values of a self-financed portfolio as a sum of portfolio exposures $(\Delta_t)_{t=0,..,n-1}$ weighted by the increments of a martingale process under $\widetilde{P}$:

$$\underbrace{\frac{X_t}{B_t} - \frac{X_0}{B_0}}_{\substack{\text{Change in discounted} \\ \text{portfolio value}}} = \sum_{i=0}^{t-1} \underbrace{\Delta_i}_{\substack{\text{Portfolio} \\ \text{exposure}}} \underbrace{\left(\frac{S_{i+1}}{B_{i+1}} - \frac{S_i}{B_i} \right)}_{\substack{\text{Martingale} \\ \text{increment}}} \quad . \tag{27}$$

Expression (27) is an example of a so called martingale transform. Martingale transforms are the discrete analogues of stochastic integrals in which the process $(\Delta_t)_{t=0,..,n-1}$ is used as the integrand and the process $(S_t/B_t)_{t=0,..,n}$ is used as an integrator. Informally, we could thus introduce the suggestive notation:

$$\frac{X_t}{B_t} - \frac{X_0}{B_0} = \int_0^t \Delta \cdot d\left(\frac{S}{B}\right) \quad ,$$

to denote such martingale transforms. In a later section this will denote a stochastic integral of an adapted process Δ with respect to the martingale process S/B over the time interval $[0, t]$.

Definition 99 *Let* $M = (M_t, \mathcal{G}_t)_{t=0,..,n}$ *be a martingale and* $H = (H_t, \mathcal{G}_t)_{t=0,..,n-1}$ *an adapted process on a probability space* $(\Omega, \mathcal{G}, P)$*. The process* $X = H \bullet M$ *defined dy*

$$X_t = \sum_{i=0}^{t-1} H_i \left(M_{i+1} - M_i \right) \quad ,$$

for $t > 0$ *and by* $X_0 = 0$ *is called the martingale transform of* M *by* H*.*

Example 100 *(i) Equation (27) defines* $(X_t/B_t)_{t=0,..,n}$ *as the martingale transform of the* $\widetilde{P}-$*martingale* $(S_t/B_t, \mathcal{G}_t)_{t=0,..,n}$ *by* $(\Delta_t, \mathcal{G}_t)_{t=0,..,n-1}$*. Therefore, after an appropriate change of probability measure the discounted value processes of self financed portfolios are martingale transforms. (ii) Let* $(Z_t, \mathcal{G}_t)_{t\geq 0}$ *be a Brownian motion on a probability space* $(\Omega, \mathcal{G}, P)$ *and* $H = (H_t, \mathcal{G}_t)_{t=0,..,n-1}$ *be an adapted process. Then the process* $(X_t)_{t=0,..,n}$ *defined by* $X_0 = 0$ *and*

$$X_t = \sum_{i=0}^{t-1} H_i \left(Z_{i+1} - Z_i \right) \ ,$$

is the martingale transform of the the Brownian motion process $(Z_t, \mathcal{G}_t)_{t\geq 0}$ *by* H*.*

Modulo some integrability conditions (which are for example satisfied if H is a bounded process), martingale transforms are martingales. Indeed, for any $s > t$:

$$\begin{aligned} E\left((H \bullet M)_s \middle| \mathcal{G}_t \right) &= \sum_{i=0}^{t-1} H_i \left(M_{i+1} - M_i \right) + E\left(\sum_{i=t}^{s-1} H_i \left(M_{i+1} - M_i \right) \middle| \mathcal{G}_t \right) \\ &= (H \bullet M)_t + \sum_{i=t}^{s-1} E\left[H_i \underbrace{E\left(M_{i+1} - M_i \middle| \mathcal{G}_i \right)}_{=0} \middle| \mathcal{G}_t \right] \quad , \end{aligned}$$

by the tower property. In fact, we will see in a later section that under some integrability conditions on the integrand H stochastic integrals are martingale processes also in the more general continuous time setting.

Finally, notice that

$$\begin{aligned} E\left((H \bullet M)_t^2 \right) &= E\left[\sum_{i=0}^{t-1} H_i \left(M_{i+1} - M_i \right) \sum_{j=0}^{t-1} H_j \left(M_{j+1} - M_j \right) \right] \\ &= \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} E\left[H_i H_j \left(M_{i+1} - M_i \right) \left(M_{j+1} - M_j \right) \right] \quad . \end{aligned}$$

Now, for any $i < j$ one has

$$E\left[H_i H_j \left(M_{i+1} - M_i\right)\left(M_{j+1} - M_j\right)\right] = E\left[H_i \left(M_{i+1} - M_i\right) H_j \underbrace{E\left(\left(M_{j+1} - M_j\right) \middle| \mathcal{G}_j\right)}_{=0}\right] = 0 \quad ,$$

again by the tower property, and similarly for $j < i$. Therefore, we get

$$E\left(\left(H \bullet M\right)_t^2\right) = E\left(\sum_{i=0}^{t-1} H_i^2 \left(M_{i+1} - M_i\right)^2\right) \quad , \tag{28}$$

i.e. a discrete time version of the so called Itô isometry (see below). Specifically, for the case where M is a Brownian motion process, equation (28) has the simpler form

$$E\left(\left(H \bullet Z\right)_t^2\right) = E\left(\sum_{i=0}^{t-1} H_i^2 E\left(\left(Z_{i+1} - Z_i\right)^2 \middle| \mathcal{G}_i\right)\right) = E\left(\sum_{i=0}^{t-1} H_i^2 \left(i + 1 - i\right)\right) = E\left(\int_0^t H_s^2 \cdot ds\right) \ ,$$

writing for any $\omega \in \Omega$ the expression $\sum_{i=0}^{t-1} H_i^2(\omega)$ as a standard Lebesgue integral $\int_0^t H_s^2(\omega) \cdot ds$.

There are several important theoretical and applied settings that ask for an extension of the martingale transform (of the discrete time stochastic integral) concept to a more general class of integrands defined on a continuous index set and with possibly not piecewise constant paths. From a more financially oriented perspective, extending the class of integrands suitable for a stochastic integration with respect to a martingale will allow us to enlarge the set of self-financed portfolios that can act as an hedging portfolio. Eventually, this will allow us to perfectly replicate some contingent claims which in the semicontinuous model setting could not be perfectly hedged.

5.2 The Stochastic Integral

The definition and the construction of the stochastic integral for continuous time integrands with respect to a Brownian motion process is performed in this section. Since with probability one the trajectories of a Brownian motion process are of unbounded variation, this construction cannot happen simply by integrating pathwise the Brownian trajectories via a standard Lebesgue Stjielties integral.

5.2.1 Some Basic Preliminaries

The basic idea in constructing the stochastic integral of an adapted integrand $H := (H_t, \mathcal{G}_t)_{0 \le t \le T}$ is to interpret such processes as random variables $H : [0,T] \times \Omega \to \mathbb{R}$ on the product measurable space

$$([0,T] \times \Omega, \mathcal{B}([0,T)) \otimes \mathcal{G}) \quad .$$

This space becomes a measure space $([0,T] \times \Omega, \mathcal{B}([0,T)) \otimes \mathcal{G}, \mu_T)$ when equipped with the product measure $\mu_T : \mathcal{B}([0,T]) \otimes \mathcal{G} \to [0,T]$ defined by

$$\mu(A) = E\left[\int_0^T \mathbf{1}_A(t,\omega)\, dt\right] = \int_\Omega \left(\int_0^T \mathbf{1}_A(t,\omega)\, dt\right) dP(\omega) \quad ,$$

for any $A \in \mathcal{B}([0,T)) \otimes \mathcal{G}$. In particular, for any adapted process $H := (H_t, \mathcal{G}_t)_{0 \le t \le T}$ one can define the L_2-norm of H as

$$\|H\|_{2,T} := \left(\int_{[0,T]\times\Omega} H^2 d\mu_T\right)^{\frac{1}{2}} = \left(E\left(\int_0^T H^2 dt\right)\right)^{\frac{1}{2}} \quad ,$$

provided of course $\|H\|_{2,T} < \infty$. We call the space of measurable processes $(H_t, \mathcal{G}_t)_{0 \le t \le T}$ such that $\|H\|_{2,T} < \infty$ the space of squared integrable random variables on $([0,T] \times \Omega, \mathcal{B}([0,T]) \otimes \mathcal{G}, \mu_T)$ and denote it by $\mathcal{H}_T$. Precisely:

$$\mathcal{H}_T := \left\{\mathcal{B}([0,T]) \otimes \mathcal{G} - \text{measurable processes } H \text{ such that } E\left(\int_0^T H^2 dt\right) < \infty\right\} \quad . \tag{29}$$

This space equipped with the norm $\|\cdot\|_{2,T}$ is a normed vector space.

Example 101 *(i) The Brownian motion process $Z := (Z_t, \mathcal{G}_t)_{0 \le t \le T}$ is an element of $\mathcal{H}_T$. Indeed,*

$$E\left(\int_0^T Z_t^2 dt\right) = \int_0^T E\left(Z_t^2\right) dt = \int_0^T t dt = \frac{T^2}{2} < \infty \quad .$$

(ii) The squared Brownian motion process $Z^2 := \left(Z_t^2, \mathcal{G}_t\right)_{0 \le t \le T}$ is also an element of $\mathcal{H}_T$. Indeed,

$$E\left(\int_0^T \left(Z_t^2\right)^2 dt\right) = \int_0^T E\left(Z_t^4\right) dt = \int_0^T 3\left(E\left(Z_t^2\right)\right)^2 dt = 3\int_0^T t^2 dt = T^3 < \infty \quad ,$$

using the fact that for a normally $\mathcal{N}\left(0,\sigma^2\right)-$distributed random variable X one has $E\left(X^4\right) = 3\sigma^4$.
(iii) In fact, by similar arguments as the ones above any power Z^k, $k \in \mathbb{N}$, of a Brownian motion can be shown to be an element of $\mathcal{H}_T$.

For simplicity, we fix in the sequel $T > 0$ and use the notation $\mathcal{H} := \mathcal{H}_T$. Convergence in the space $\mathcal{H}$ means convergence in the $\|\cdot\|_{2,T}$ −norm.

Definition 102 *(i) A sequence $(H^n)_{n\in\mathbb{N}} \subset \mathcal{H}$ is said to converge to some process $H \in \mathcal{H}$ if and only if $\|H^n - H\|_{2,T} \to 0$, as $n \to \infty$, i.e. if and only if*

$$\left(E\left(\int_0^T (H^n - H)^2\, dt\right)\right)^{\frac{1}{2}} \underset{n\to\infty}{\to} 0 \quad ,$$

or, equivalently, if and only if

$$E\left(\int_0^T (H^n - H)^2\, dt\right) \underset{n\to\infty}{\to} 0 \quad .$$

(ii) In that case we call H the limit[4] of the sequence $(H^n)_{n\in\mathbb{N}}$ and denote it by $H = \lim_{n\to\infty} H^n$.

As usual, as for instance when defining standard Lebesgue integrals, one starts from integrands that are piecewise constant with respect to some underlying variable, i.e. simple integrands, and then extends the integral definition to more general integrands by means of a suitable limit argument.

5.2.2 Simple Integrands

Simple integrands in the space $\mathcal{H}$ and stochastic integrals for simple processes are defined as follows.

Definition 103 *(i) An adapted process $H = (H_t, \mathcal{G}_t)_{t\geq 0}$ is simple if there exists a partition $0 = t_0 < t_1 < .. < t_n = T$ of $[0,\infty)$ such that*

$$H_t(\omega) = \xi_0(\omega)\,\mathbf{1}_{\{0\}}(t) + \sum_{i=0}^{n-1} \xi_i(\omega)\,\mathbf{1}_{(t_i,t_{i+1}]}(t) \quad , \quad \text{for all } (t,\omega) \in [0,T]\times\Omega \quad ,$$

and for some $\mathcal{G}_{t_i}$−measurable, bounded, random variables $\xi_i(\omega)$, $i = 0,..,n$. The vector space of simple processes is denoted by $\mathcal{S}$. (ii) For a simple process $H = (H_t, \mathcal{G}_t)_{0\leq t\leq T} \in \mathcal{S}$, the stochastic

[4] *Notice, that if $H = \lim_{n\to\infty} H^n$, then H can be modified on a set of μ_T measure 0 without affecting the value of $\|\cdot\|_{2,T}$. Therefore, every limit in $\mathcal{H}$ is in fact a class of processes that can differ on a set of μ_T measure 0.*

integral of H with respect to the Brownian motion $Z := (Z_t, \mathcal{G}_t)_{t\geq 0}$ is the martingale transform of Z by H. That is, for any $t \in (t_k, t_{k+1}]$ and $k = 0, .., n-1$, we define:

$$\int_0^t H_s dZ_s := \sum_{i=0}^{k-1} \xi_i \left(Z_{t_{i+1}} - Z_{t_i}\right) + \xi_k \left(Z_t - Z_{t_k}\right) \quad .$$

Sometimes we will write for brevity $\int_0^t HdZ := \int_0^t H_s dZ_s$.

Remark 104 *(i) Notice that by construction $\mathcal{S} \subset \mathcal{H}$ since*

$$\begin{aligned} E\left(\int_0^T H^2 dt\right) &= E\left(\int_0^T \left(\xi_0^2 \mathbf{1}_{\{0\}}(s) + \sum_{i=0}^{n-1} \xi_i^2 \mathbf{1}_{(t_i, t_{i+1}]}(s)\right) ds\right) \\ &= E\left(\sum_{i=0}^{n-1} \xi_i^2 (t_{i+1} - t_i)\right) = \sum_{i=0}^{n-1} (t_{i+1} - t_i) E\left(\xi_i^2\right) < \infty \quad , \end{aligned} \tag{30}$$

by the boundedness of $\xi_1, .., \xi_n$. (ii) For any given $t \geq 0$ one can define the Brownian motion process stopped at time t by

$$Z^t := (Z_{s\wedge t})_{s\geq 0} \quad .$$

With this definition we have for any $t \in (t_k, t_k + 1]$:

$$Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t} = \begin{cases} Z_{t_i+1} - Z_{t_i} & i < k \\ Z_t - Z_{t_k} & i = k \\ Z_t - Z_t = 0 & i > k \end{cases} ,$$

implying

$$\int_0^t H_s dZ_s = \sum_{i=0}^{k-1} \xi_i \left(Z_{t_{i+1}} - Z_{t_i}\right) + \xi_k \left(Z_t - Z_{t_k}\right) = \sum_{i=0}^{n-1} \xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) \quad .$$

Moreover, we have

$$E\left(Z_{t_i+1}^t \middle| \mathcal{G}_{t_i}\right) = E\left(Z_{(t_i+1)\wedge t} \middle| \mathcal{G}_{t_i}\right) = \begin{cases} Z_{t_i} = Z_{t_i\wedge t} = Z_{t_i}^t & i \leq k \\ Z_t = Z_{t_i\wedge t} = Z_{t_i}^t & i > k \end{cases} ,$$

that is $(Z_s^t, \mathcal{G}_s)_{s=0,t_1,t_2,..,t_n}$ is a discrete time martingale and $\int_0^t H_s dZ_s$ is the martingale transform of Z^t by H. Finally, since for any $\omega \in \Omega$ and $t_i \in \{0, t_1, t_2, .., t_n\}$ the mapping $t \longmapsto Z_{t_i}^t(\omega)$ is a *continuous one, it follows that the mapping*

$$t \longmapsto \int_0^t H_s dZ_s = \sum_{i=0}^{n-1} \xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i \wedge t} \right) \quad ,$$

is, as a linear combination of continuous functions, also continuous. Therefore, the stochastic integral process $\left(\int_0^t HdZ \right)_{t\geq 0}$ *is a continuous time process with continuous trajectories. Notice, that the integrand* H *is a process with possibly discontinuous trajectories. Therefore, the stochastic integral is a "regularizing" operator that maps possibly discontinuous processes into process with continuous trajectories.*

It is not suprising that the stochastic integral process $\left(\int_0^t HdZ \right)_{t\geq 0}$ is a martingale process, since it can be written as a martingale transform. Moreover, second moments of stochastic integrals can be often easily computed by means of the so called Itô isometry. We summarize these properties in the next result for the case of a simple integrand $H \in \mathcal{S}$. In the next section, such properties will hold also for stochastic integrals of more general integrands $H \in \mathcal{H}$.

Proposition 105 *Let* $H, H' \in \mathcal{S}$. *Then it follows:*

1. *The stochastic integral is a linear operator, that is:*

$$\int_0^t (\alpha H + \beta H')\, dZ = \alpha \int_0^t HdZ + \beta \int_0^t H' dZ,$$

   for any α,$\beta \in \mathbb{R}$.

2. $\left(\int_0^t HdZ, \mathcal{G}_t \right)_{t\geq 0}$ *is a martingale with continuous trajectories.*

3. *The Itô isometry holds:*

$$E\left[\left(\int_0^t HdZ \right)^2 \right] = E\left(\int_0^t H^2 ds \right) \quad .$$

Proof. 1. This is immediate from the definition of the stochastic integral $\int_0^t HdZ$ as a (stochastic) linear combination of Brownian increments. To prove 2. it remains to show that

$\left(\int_0^t HdZ, \mathcal{G}_t\right)_{t\geq 0}$ is a martingale (continuity was already established in Remark 104). Thus, let $t \leq s$. We then have

$$E\left(\int_0^s HdZ \middle| \mathcal{G}_t\right) = \sum_{i=0}^{n-1} E\left(\xi_i \left(Z_{(t_i+1)\wedge s} - Z_{t_i\wedge s}\right) \middle| \mathcal{G}_t\right) \quad ,$$

and

$$E\left(\xi_i \left(Z_{(t_i+1)\wedge s} - Z_{t_i\wedge s}\right) \middle| \mathcal{G}_t\right) = \begin{cases} \left.\begin{array}{c} \xi_i E\left(Z_{(t_i+1)\wedge s} - Z_{t_i\wedge s} \middle| \mathcal{G}_t\right) \\ = \xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge s}\right) \\ = \xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) \end{array}\right\} & t \geq t_i \\ \left.\begin{array}{c} E\left(\xi_i E\left(Z_{(t_i+1)\wedge s} - Z_{t_i\wedge s} \middle| \mathcal{G}_{t_i}\right) \middle| \mathcal{G}_t\right) \\ = 0 = \xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) \end{array}\right\} & t < t_i \end{cases} ,$$

because for $t \leq t_i$ it follows $Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t} = 0$. Therefore.

$$E\left(\int_0^s HdZ \middle| \mathcal{G}_t\right) = \sum_{i=0}^{n-1} E\left(\xi_i \left(Z_{(t_i+1)\wedge s} - Z_{t_i\wedge s}\right) \middle| \mathcal{G}_t\right) = \sum_{i=0}^{n-1} \xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) = \int_0^t HdZ \quad ,$$

To prove 3. note first that we have

$$\begin{aligned} E\left[\left(\int_0^t HdZ\right)^2\right] &= E\left(\sum_{i=0}^{n-1} \xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) \sum_{j=0}^{n-1} \xi_j \left(Z_{(t_j+1)\wedge t} - Z_{t_j\wedge t}\right)\right) \\ &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} E\left(\xi_i \xi_j \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) \left(Z_{(t_j+1)\wedge t} - Z_{t_j\wedge t}\right)\right) \quad . \end{aligned}$$

Further, if $i < j$ (i.e if $t_{i+1} \leq t_j$) it follows

$$\begin{aligned} & E\left(\xi_i \xi_j \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) \left(Z_{(t_j+1)\wedge t} - Z_{t_j\wedge t}\right)\right) \\ = \; & E\left(\xi_i \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right) \xi_j E\left(Z_{(t_j+1)\wedge t} - Z_{t_j\wedge t} \middle| \mathcal{G}_{t_j}\right)\right) \\ = \; & 0 \quad , \end{aligned}$$

by the independence of Brownian increments. A similar argument applies for the case $j < i$. Moreover, for $i = j$ we get

$$\begin{aligned} E\left(\xi_i^2 \left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right)^2\right) &= E\left(\xi_i^2 E\left(\left(Z_{(t_i+1)\wedge t} - Z_{t_i\wedge t}\right)^2 \middle| \mathcal{G}_{t_i}\right)\right) \\ &= \begin{cases} 0 = E\left(\xi_i^2 \left((t_i+1)\wedge t - t_i \wedge t\right)\right) & t \leq t_i \\ E\left(\xi_i^2 \left((t_i+1)\wedge t - t_i \wedge t\right)\right) & t > t_i \end{cases} . \end{aligned}$$

Consequently,

$$\begin{aligned} E\left[\left(\int_0^t HdZ\right)^2\right] &= \sum_{i=0}^{n-1} E\left(\xi_i^2\left((t_i+1)\wedge t - t_i \wedge t\right)\right) = E\left(\sum_{i=0}^{n-1} \xi_i^2\left((t_i+1)\wedge t - t_i \wedge t\right)\right) \\ &= E\left(\int_0^t H^2 ds\right), \end{aligned}$$

as claimed. This concludes the proof. ■

Basically, one can think about the problem of defining a stochastic integral for more general integrands than simple processes, as the problem of extending smoothly the integral definition of a process $H \in \mathcal{S}$ to a larger space of adapted integrands. Smoothness of the extension procedure is desirable in order to maintain the integral properties in Proposition 105 - which are valid for integrands $H \in \mathcal{S}$ - also for the resulting stochastic integral of a more general integrand.

It turns out that the adequate space on which stochastic integrals can be smoothly extended is the space $\mathcal{H}$ defined in (29). This fact relies on an approximation result which states that any $H \in \mathcal{H}$ can be approximated by a sequence of simple processes $(H^n)_{n\in\mathbb{N}} \subset \mathcal{S}$ converging to H, i.e. approximating H in the norm defined on $\mathcal{H}$.

For completeness, we state without proof this crucial approximation finding precisely in the next proposition.

Proposition 106 *For any process $H \in \mathcal{H}$ there exists a sequence $(H^n)_{n\in\mathbb{N}} \subset \mathcal{S}$ such that*

$$E\left(\int_0^T (H - H^n)^2 ds\right) \underset{n\to\infty}{\to} 0 .$$

Example 107 *We illustrate the above approximation result for the case where $H = Z$. We know (see Example 101) that the Brownian motion process $Z := (Z_t)_{0\le t\le T}$ is an element of $\mathcal{H}$. Z can be approximated by means of a sequence $(H^n)_{n\in\mathbb{N}} \subset \mathcal{H}$ given by*[5]

$$H_t^n(\omega) := \sum_{i=0}^{2^n-1} Z_{iT/2^n}(\omega)\, \mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t) \quad .$$

[5] *Notice, that for any fixed $\omega \in \Omega$ this is the same type of approximation procedure we would use to define a standard Lebesgue integral of the function $t \longmapsto Z_t(\omega)$ (see again Remark 45).*

Indeed, we have

$$
\begin{aligned}
E\int_0^T (Z-H^n)^2\,dt &= \int_0^T E\,(Z_t-H_t^n)^2\,dt \\
&= \int_0^T E\left(Z_t-\sum_{i=0}^{2^n-1} Z_{iT/2^n}\mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t)\right)^2 dt \\
&= \int_0^T E\left(\sum_{i=0}^{2^n-1}\left(Z_t-Z_{iT/2^n}\right)\mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t)\right)^2 dt \\
&= \int_0^T \sum_{i=0}^{2^n-1} E\left(Z_t-Z_{iT/2^n}\right)^2 \mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t)\,dt \\
&= \sum_{i=0}^{2^n-1}\int_0^T (t-iT/2^n)\,\mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t)\,dt \\
&= \sum_{i=0}^{2^n-1}\frac{(T/2^n)^2}{2} = \frac{T}{2^{n+1}} \underset{n\to\infty}{\to} 0 \quad .
\end{aligned}
$$

Therefore, $H^n \underset{n\to\infty}{\to} Z$ *in the space* $\mathcal{H}$*. Remark that strictly speaking* $H^n \notin \mathcal{S}$ *because* Z_t *is unbounded for any* t*. However, any* H^n *can be approximated by a sequence* $\left(K^{nk}\right)_{k\in\mathbb{N}} \subset \mathcal{S}$ *defined by*

$$
\begin{aligned}
K_t^{nk}(\omega) &= \sum_{i=0}^{2^n-1} Z_{iT/2^n}(\omega)\,\mathbf{1}_{(iT/2^n,iT/2^n+1]}(t)\,\mathbf{1}_{\left\{Z_{iT/2^n}\le k\right\}}(\omega) \\
&\quad + \sum_{i=0}^{2^n-1} k\mathbf{1}_{(iT/2^n,iT/2^n+1]}(t)\,\mathbf{1}_{\left\{\left|Z_{iT/2^n}\right|>k\right\}}(\omega) \quad .
\end{aligned}
$$

Indeed,

$$
\begin{aligned}
E\left(\int_0^T \left(H_t^n-K_t^{nk}\right)^2 dt\right) &= \int_0^T E\left(\left(H_t^n-K_t^{nk}\right)^2\right) dt \\
&= \int_0^T E\left(\left(\sum_{i=0}^{2^n-1}\left(Z_{iT/2^n}-k\right)\mathbf{1}_{\left\{\left|Z_{iT/2^n}\right|>k\right\}}\mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t)\right)^2\right) dt \\
&= \int_0^T \left(\sum_{i=0}^{2^n-1} E\left(\left(Z_{iT/2^n}-k\right)^2\mathbf{1}_{\left\{\left|Z_{iT/2^n}\right|>k\right\}}\right)\mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t)\right) dt
\end{aligned}
$$

This last integral is finite and non negative, since it is the Lebesgue integral of a simple (deterministic) function of t*:*

$$
t \longmapsto \sum_{i=0}^{2^n-1} E\left[\left(Z_{iT/2^n}-k\right)^2\mathbf{1}_{\left\{\left|Z_{iT/2^n}\right|>k\right\}}\right]\mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t) \tag{31}
$$

Moreover, by the properties of normal distributions one has for any $i = 0, .., 2^n - 1$ and any $k > k'$, the strict inequality

$$E\left[\left(Z_{iT/2^n} - k\right)^2 \mathbf{1}_{\left\{\left|Z_{iT/2^n}\right| > k\right\}}\right] < E\left[\left(Z_{iT/2^n} - k'\right)^2 \mathbf{1}_{\left\{\left|Z_{iT/2^n}\right| > k'\right\}}\right] \quad ,$$

implying that as a function of k the corresponding Lebesgue integrals in (31) build a monotonically strictly decreasing non negative sequence. Therefore, this sequence of integrals can only converge to a limit 0, i.e.:

$$E\left(\int_0^T \left(H_t^n - K_t^{nk}\right)^2 dt\right) \underset{k\to\infty}{\downarrow} 0 \quad .$$

Summarizing, Z can be approximated by the sequence $(H^n)_{n\in\mathbb{N}} \subset \mathcal{H}$ and any H^n can be approximated by a sequence $\left(K^{nk}\right)_{k\in\mathbb{N}} \subset \mathcal{S}$. In turn, Z can be approximated by a sequence $(W^n)_{n\in\mathbb{N}} \subset \mathcal{S}$ where $W^n := K^{nn}$.

5.2.3 Squared Integrable Integrands

We can now define the stochastic integral of any process $H \in \mathcal{H}$ by using the approximation result in Proposition 106.

We notice first that for any sequence $(H^n)_{n\in\mathbb{N}} \subset \mathcal{S}$ converging to a process $H \in \mathcal{H}$, i.e. such that

$$E\int_0^T (H_s - H_s^n)^2 ds \underset{n\to\infty}{\to} 0 \quad ,$$

it follows, for any $0 \le t \le T$:

$$\begin{aligned} E\left[\left(\int_0^t H_s^n dZ - \int_0^t H_s^m dZ\right)^2\right] &= E\left[\left(\int_0^t (H_s^n - H_s^m)\, dZ\right)^2\right] \\ &= E\int_0^t (H_s^n - H_s^m)^2 ds \\ &\le 2E\int_0^T (H_s - H_s^n)^2 ds + 2E\int_0^T (H_s - H_s^m)^2 ds \underset{n,m\to\infty}{\to} 0 \end{aligned} \tag{32}$$

using in the first equality the linearity of stochastic integrals for simple processes, in the second the Itô isometry applied to the simple process $H_s^n - H_s^m$, and in the last inequality the bound $(a+b)^2 \leq 2\left(a^2+b^2\right)$, where $a, b \in \mathbb{R}$.

Therefore, for any $0 \leq t \leq T$ the sequence $\left(\int_0^t H_s^n dZ\right)_{n\in\mathbb{N}}$ is a Cauchy sequence of random variables with mean 0 and variance $E\left(\int_0^t (H^n)^2 ds\right) < \infty$, i.e. a Cauchy sequence in the space of squared integrable random variables on $(\Omega, \mathcal{G}, P)$. It is well known that in this space any Cauchy sequence converges to a well defined element of the space. Therefore, it is a natural idea to define the limit of the sequence $\left(\int_0^t H_s^n dZ\right)_{n\in\mathbb{N}}$ as the stochastic integral $\int_0^t H dZ$ of H with respect to Z. We state this precisely in the next definition.

Definition 108 *Let $H \in \mathcal{H}$ be a squared integrable process and let $(H^n)_{n\in\mathbb{N}} \subset \mathcal{S}$ be any sequence of simple processes approximating H, i.e. such that:*

$$E\int_0^T (H-H^n)^2 ds \underset{n\to\infty}{\to} 0 \quad .$$

For any $0 \leq t \leq T$ the stochastic integral of H with respect to the Brownian motion $Z := (Z_t, \mathcal{G}_t)_{0\leq t\leq T}$ is defined by

$$\int_0^t H_s dZ_s := \lim_{n\to\infty} \int_0^t H_s^n dZ_s \quad .$$

Example 109 *We compute the stochastic integral as the corresponding limit in an example where explicit computations are possible. Recall from Example 107 that the sequence $(H^n) \subset \mathcal{S}$ defined by*

$$H_t^n(\omega) := \sum_{i=0}^{2^n-1} Z_{iT/2^n}(\omega)\,\mathbf{1}_{(iT/2^n,(i+1)T/2^n]}(t) \ ,$$

converges in the space $\mathcal{H}$ to the Brownian motion process Z. We have:

$$\begin{aligned}\int_0^T H_s^n dZ_s &= \sum_{i=0}^{2^n-1} Z_{iT/2^n}\left(Z_{(i+1)T/2^n} - Z_{iT/2^n}\right)\\ &= \frac{1}{2}\left(Z_T^2 - Z_0^2\right) - \frac{1}{2}\sum_{i=0}^{2^n-1}\left(Z_{(i+1)T/2^n} - Z_{iT/2^n}\right)^2 \quad .\end{aligned}$$

Moreover, we have

$$E\left(\sum_{i=0}^{2^n-1}\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2\right)=\sum_{i=0}^{2^n-1}E\left(\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2\right)=\sum_{i=0}^{2^n-1}\frac{T}{2^n}=T \quad ,$$

and

$$\begin{aligned}
E\left[\left(\sum_{i=0}^{2^n-1}\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2-T\right)^2\right] &= E\left[\left(\sum_{i=0}^{2^n-1}\left(\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2-\frac{T}{2^n}\right)\right)^2\right]\\
&= \sum_{i=0}^{2^n-1}E\left[\left(\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2-\frac{T}{2^n}\right)^2\right]\\
&= \sum_{i=0}^{2^n-1}E\left(\left(\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2-\frac{T}{2^n}\right)^2\right) \quad ,
\end{aligned}$$

using in the second equality the independence of Brownian, which implies

$$E\left[\left(\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2-\frac{T}{2^n}\right)\left(\left(Z_{(j+1)T/2^n}-Z_{jT/2^n}\right)^2-\frac{T}{2^n}\right)\right]=0 \quad ,$$

for $i \neq j$. Further we obtain

$$\begin{aligned}
E\left(\left(\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2-\frac{T}{2^n}\right)^2\right) &= E\left[\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^4\right]\\
&\quad +\left(\frac{T}{2^n}\right)^2-2\cdot\frac{T}{2^n}E\left(\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2\right)\\
&= 3\left(\frac{T}{2^n}\right)^2+\left(\frac{T}{2^n}\right)^2-2\left(\frac{T}{2^n}\right)^2=2\left(\frac{T}{2^n}\right)^2 \quad .
\end{aligned}$$

Therefore,

$$E\left[\left(\sum_{i=0}^{2^n-1}\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2-T\right)^2\right]=\frac{T}{2^{n-1}}\underset{n\to\infty}{\to}0 \quad ,$$

i.e.

$$\sum_{i=0}^{2^n-1}\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2\underset{n\to\infty}{\to}T \quad ,$$

in the space of squared integrable random variables on $(\Omega, \mathcal{G}, P)$. Summarizing this gives

$$\begin{aligned}
\int_0^T H_s dZ_s &= \lim_{n\to\infty}\int_0^T H_s^n dZ_s\\
&= \lim_{n\to\infty}\left(\frac{1}{2}Z_T^2-\frac{1}{2}\sum_{i=0}^{2^n-1}\left(Z_{(i+1)T/2^n}-Z_{iT/2^n}\right)^2\right)\\
&= \frac{1}{2}Z_T^2-\frac{1}{2}T \quad .
\end{aligned}$$

5.2.4 Properties of Stochastic Integrals

It is immediate from the definition of the stochastic integral of a process $H \in \mathcal{H}$ defined as the limit of a sequence of stochastic integrals of processes $H^n \in \mathcal{S}$, that linearity is preserved in the limit, i.e.

$$\int_0^t (\alpha H_s + \beta H'_s)\, dZ_s = \alpha \int_0^t H_s dZ_s + \beta \int_0^t H'_s dZ_s \quad ,$$

for any H, $H' \in \mathcal{H}$ and α, $\beta \in \mathbb{R}$.

Furthermore, the key property of the space $\mathcal{H}$ from the perspective of stochastic integration is that convergence of a sequence $(H^n)_{n\in\mathbb{N}} \subset \mathcal{H}$ to some limit $H \in \mathcal{H}$ defines the corresponding stochastic integral as the limit of the sequence $\left(\int H^n dZ\right)_{n\in\mathbb{N}}$ *in the space of squared integrable random variables on* $(\Omega, \mathcal{G}, P)$ (see again equation (32) and the following discussion). In fact, this implies convergence of the first two moments and the conditional expectations of the sequence $\left(\int H^n dZ\right)_{n\in\mathbb{N}}$ to the first two moments and the conditional expectations of the limit

$$\int H dZ = \lim_{n\to\infty} \int H^n dZ \quad .$$

For instance, this gives

$$E\left(\int_0^t H_u dZ_u \middle| \mathcal{G}_s\right) = \lim_{n\to\infty} E\left(\int_0^t H^n_u dZ_u \middle| \mathcal{G}_s\right) = \lim_{n\to\infty} \int_0^s H^n_u dZ_u = \int_0^s H^n_u dZ_u \quad ,$$

where $t \geq s$, i.e. the martingale property. Therefore, the martingale property and the Itô isometry of stochastic integrals, which have been shown to hold for integrals of simple processes, are maintained for stochastic integrals of integrands $H \in \mathcal{H}$.

Finally, it can also be shown at the cost of some more technical details that convergence in the space of squared integrable random variables together with the martingale property and the continuity of stochastic integrals of simple processes implies the continuity of the stochastic integral of a process $H \in \mathcal{H}$. For completeness, we summarize the above discussion in the next proposition.

Proposition 110 *Let $H, H' \in \mathcal{H}$. It then follows:*

1. *The stochastic integral is a linear operator, that is:*

$$\int_0^t (\alpha H + \beta H')\, dZ = \alpha \int_0^t H dZ + \beta \int_0^t H' dZ,$$

for any $\alpha,\beta \in \mathbb{R}$.

2. $\left(\int_0^t H dZ, \mathcal{G}_t\right)_{0\leq t\leq T}$ *is a martingale with continuous trajectories.*

3. *The Itô isometry holds:*

$$E\left[\left(\int_0^t H dZ\right)^2\right] = E\left(\int_0^t H^2 ds\right) \quad .$$

5.3 Itô's Lemma

This section introduces a differential calculus for differentiable functions of stochastic integrals. This calculus is called Itô's calculus and its primary tool is Itô's formula, a version of the fundamental theorem of calculus for stochastic differentials.

5.3.1 Starting Point, Motivation and Some First Examples

To introduce the basic ideas behind stochastic differentials we start with a simple illustrative example. Consider the squared Brownian motion process $Z^2 := \left(Z_t^2, \mathcal{G}_t\right)_{t\geq 0}$. The goal is to express Z_t^2 by means of an integral form of the type

$$Z_t^2 - Z_0^2 = \int_0^t K_s ds + \int_0^t H_s dZ_s \quad ,$$

for some suitable adapted integrands $(K_t, \mathcal{G}_t)_{t\geq 0}$ and $(H_t, \mathcal{G}_t)_{t\geq 0}$. Such a representation would motivate the suggestive stochastic differential notation

$$d\left(Z_t^2\right) = K_t dt + H_t dZ_t \quad .$$

Under the naive assumption that for any $\omega \in \Omega$ the Brownian path $t \longmapsto Z_t(\omega)$ is a differentiable function (we know it is not!) one would be tempted to apply the standard fundamental theorem of calculus to write

$$
\begin{aligned}
Z_t^2(\omega) - Z_0^2(\omega) &= \int_0^t \frac{d}{dt} Z_s^2(\omega)\, ds = 2\int_0^t Z_s(\omega) \cdot \underbrace{\frac{d}{dt} Z_s(\omega)\, ds}_{\substack{= dZ_s(\omega) \text{ under} \\ \text{the differentiability} \\ \text{assumption}}} \\
&= 2\int_0^t Z_s(\omega) \cdot dZ_s(\omega) \quad .
\end{aligned} \tag{33}
$$

Unfortunately, this approach cannot work, as we know. In fact, we already developed a particular stochastic integral construction in order to avoid the fact that Brownian trajectories are of unbounded variation and thus not differentiable.

Indeed, the naive approach (33) leads immediately to an internal inconsistency which can be highlighted as follows. First, notice that the stochastic integral process $\left(\int_0^t ZdZ\right)_{0\leq t\leq T}$ on the RHS of (33) is a martingale, since $(Z_t, \mathcal{G}_t)_{0\leq t\leq T} \in \mathcal{H}$ (cf. again Example 101). At the same time, the process $\left(Z_t^2, \mathcal{G}_t\right)_{0\leq t\leq T}$ is a submartingale, since $\left(Z_t^2 - t, \mathcal{G}_t\right)_{0\leq t\leq T}$ is a martingale (cf. again Example (98)):

$$
E\left(Z_s^2 - s \middle| \mathcal{G}_t\right) = Z_t^2 - t \quad \Longleftrightarrow \quad E\left(Z_s^2 \middle| \mathcal{G}_t\right) = Z_t^2 + (s-t) > Z_t^2 \quad , \tag{34}
$$

for any $s > t$. Thus, (34) shows that a fundamental theorem of calculus for stochastic integrals has to be of a different form that the standard one. In particular, it appears that the standard Theorem neglects a deterministic term in Z_t^2 which is the expected value $t = E\left(Z_t^2\right)$ of Z_t^2 in the LHS of (33). Therefore, one could be tempted to write

$$
Z_t^2 - t = 2\int_0^t ZdZ \quad ,
$$

i.e.

$$
Z_t^2 - Z_0^2 = 2\left(\int_0^t ZdZ + \frac{1}{2}t\right) \quad , \tag{35}
$$

in order to avoid, at least superficially, the inconsistency behind (33). It turns out that this is the correct guess. Indeed, the structure behind (35) can be highlighted as a special case of Itô's formula by setting $g(Z_t) = Z_t^2$, to get

$$g(Z_t) - g(Z_0) = \int_0^t g'(Z_s)\, dZ_s + \frac{1}{2}\int_0^t g''(Z_s)\, ds \quad . \tag{36}$$

Notice that in (36) Z_t^2 has been decomposed as the sum of a stochastic integral giving the martingale component of Z_t^2, and a standard pathwise Lebesgue integral describing the deterministic trend of Z_t^2. Expression (36) gives the mathematical foundation for the stochastic differential form

$$dg(Z_t) = g'(Z_t)\, dZ_t + \frac{1}{2} g''(Z_t)\, dt \quad ,$$

of Itô's formula.

An immediate extension of Itô's formula (36) arises when g is a function of both t and Z_t. In this case, the rule is to apply standard differentials to deterministic variables and stochastic differentials - by means of Itô's formula (36) - to stochastic variables, that is:

$$g(t, Z_t) - g(0, Z_0) = \int_0^t \frac{\partial g}{\partial Z}(s, Z_s)\, dZ_s + \int_0^t \left(\frac{\partial g}{\partial t}(s, Z_s) + \frac{1}{2}\frac{\partial^2 g}{\partial^2 Z}(s, Z_s) \right) ds \quad , \tag{37}$$

or, in differential form,

$$dg(t, Z_t) = \frac{\partial g}{\partial Z}(t, Z_t)\, dZ_t + \left(\frac{\partial g}{\partial t}(t, Z_t) + \frac{1}{2}\frac{\partial^2 g}{\partial^2 Z}(t, Z_t) \right) dt$$

Example 111 *(Geometric Brownian Motion) A geometric Brownian motion process* $(S_t, \mathcal{G}_t)_{t \geq 0}$ *is defined by*

$$S_t = S_0 \exp\left(\sigma Z_t + \left(\mu - \frac{\sigma^2}{2} \right) t \right) =: g(t, Z_t) \quad , \tag{38}$$

for given S_0, $\sigma > 0$ *and* $\mu \in \mathbb{R}$. *It then follows*

$$\begin{aligned} \frac{\partial g}{\partial t}(t, Z_t) &= S_0 \exp\left(\sigma Z_t + \left(\mu - \frac{\sigma^2}{2} \right) t \right) \left(\mu - \frac{\sigma^2}{2} \right) = S_t \left(\mu - \frac{\sigma^2}{2} \right) \\ \frac{\partial g}{\partial Z}(t, Z_t) &= S_0 \exp\left(\sigma Z_t + \left(\mu - \frac{\sigma^2}{2} \right) t \right) \sigma = S_t \sigma \\ \frac{\partial^2 g}{\partial^2 Z}(t, Z_t) &= S_t \sigma^2 \quad . \end{aligned}$$

Itô's formula (37) then gives

$$\begin{aligned} S_t - S_0 &= g(t, Z_t) - g(0, Z_0) \\ &= \int_0^t \frac{\partial g}{\partial Z}(s, Z_s)\, dZ_s + \int_0^t \left(\frac{\partial g}{\partial t}(s, Z_s) + \frac{1}{2}\frac{\partial^2 g}{\partial^2 Z}(s, Z_s) \right) ds \quad , \\ &= \int_0^t S_s \sigma dZ_s + \int_0^t \left(S_s \left(\mu - \frac{\sigma^2}{2} \right) + \frac{1}{2} S_s \sigma^2 \right) ds \\ &= \int_0^t S_s \mu ds + \int_0^t S_s \sigma dZ_s \quad , \end{aligned}$$

or, in differential form,

$$dS_t = \mu S_t dt + \sigma S_t dZ_t \quad .$$

Remark 112 *Example 111 provides a first example of a solution to a stochastic differential equation. Indeed, if we consider the problem of determining an adapted process $(S_t, \mathcal{G}_t)_{t \geq 0}$ that solves the stochastic differential equation*

$$dS_t = \mu S_t dt + \sigma S_t dZ_t \quad ,$$

for a given S_0, Example 111 already gives the answer: $(S_t, \mathcal{G}_t)_{t \geq 0}$ is the Geometric Brownian motion process defined in (38).

5.3.2 A Simplified Derivation of Itô's Formula

We illustrate some of the main ideas behind the proof of Itô's formula (37) by providing a proof for the simpler case where

$$g(t, Z) = \frac{1}{2}(t + Z)^2 \quad . \tag{39}$$

Thus, we are going to show that

$$g(t, Z) = \int_0^t \frac{\partial g}{\partial Z}(s, Z_s)\, dZ_s + \int_0^t \left(\frac{\partial g}{\partial t}(s, Z_s) + \frac{1}{2}\frac{\partial^2 g}{\partial^2 Z}(s, Z_s) \right) ds \quad ,$$

i.e., for our specific case (apply Itô's formula (37) to (39)),

$$\frac{1}{2}(t + Z_t)^2 = \int_0^t \left(\frac{1}{2} + s + Z_s \right) ds + \int_0^t (s + Z_s)\, dZ_s \quad . \tag{40}$$

Proof. Let us fix first a partition $\{t_0, .., t_{2^n}\}$ of the interval $[0, t]$, given by

$$t_i = \frac{it}{2^n} \quad , \quad i = 0, .., 2^n \quad .$$

We then have,

$$\frac{1}{2}(t + Z_t)^2 = \frac{1}{2}\sum_{i=0}^{2^n-1}\left[\left(t_{i+1} + Z_{t_{i+1}}\right)^2 - \left(t_i + Z_{t_i}\right)^2\right] \quad . \tag{41}$$

We can now apply an *exact* second order Taylor approximation to any term in the sum (41):

$$\begin{aligned}
\frac{1}{2}\left[\left(t_{i+1} + Z_{t_{i+1}}\right)^2 - \left(t_i + Z_{t_i}\right)^2\right] &= g\left(t_{i+1}, Z_{t_{i+1}}\right) - g\left(t_i, Z_{t_i}\right) \\
&= \frac{\partial g\left(t_i, Z_{t_i}\right)}{\partial t}\left(t_{i+1} - t_i\right) + \frac{\partial g\left(t_i, Z_{t_i}\right)}{\partial Z}\left(Z_{t_{i+1}} - Z_{t_i}\right) \\
&\quad + \frac{1}{2}\left(\frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 t}\left(t_{i+1} - t_i\right)^2 + \frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 Z}\left(Z_{t_{i+1}} - Z_{t_i}\right)^2\right) \\
&\quad + \frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial t \partial Z}\left(t_{i+1} - t_i\right)\left(Z_{t_{i+1}} - Z_{t_i}\right) \quad .
\end{aligned}$$

Explicit computations then give

$$\frac{\partial g\left(t_i, Z_{t_i}\right)}{\partial t} = t_i + Z_{t_i} = \frac{\partial g\left(t_i, Z_{t_i}\right)}{\partial Z} \quad ,$$

and

$$\frac{\partial g\left(t_i, Z_{t_i}\right)}{\partial t \partial Z} = \frac{\partial g\left(t_i, Z_{t_i}\right)}{\partial Z \partial t} = \frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 t} = \frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 Z} = 1 \quad .$$

Therefore,

$$\begin{aligned}
\left(\frac{t_{i+1} + Z_{t_{i+1}}}{2}\right)^2 - \left(\frac{t_i + Z_{t_i}}{2}\right)^2 &= \left(t_i + Z_{t_i}\right)\left(t_{i+1} - t_i + Z_{t_{i+1}} - Z_{t_i}\right) \\
&\quad + \frac{1}{2}\left(t_{i+1} - t_i\right)^2 + \frac{1}{2}\left(Z_{t_{i+1}} - Z_{t_i}\right)^2 \\
&\quad + \frac{1}{2} \cdot 2\left(t_{i+1} - t_i\right)\left(Z_{t_{i+1}} - Z_{t_i}\right) \quad .
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{2}(t + Z_t)^2 &= \sum_{i=0}^{2^n-1}\left(t_i + Z_{t_i}\right)\left(t_{i+1} - t_i + Z_{t_{i+1}} - Z_{t_i}\right) \\
&\quad + \frac{1}{2}\sum_{i=0}^{2^n-1}\left[\left(t_{i+1} - t_i\right)^2 + \left(Z_{t_{i+1}} - Z_{t_i}\right)^2 + 2\left(t_{i+1} - t_i\right)\left(Z_{t_{i+1}} - Z_{t_i}\right)\right] \quad .
\end{aligned} \tag{42}$$

We now compute the limit as $n \to \infty$ of each term in this expression. We first have

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}(t_i+Z_{t_i})(t_{i+1}-t_i)=\int_0^t(s+Z_s)\,ds \quad ,$$

a pathwise Lebesgue integral. Further,

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}(t_i+Z_{t_i})\left(Z_{t_{i+1}}-Z_{t_i}\right)=\int_0^t(s+Z_s)\,dZ_s \quad ,$$

a stochastic integral (see again Example 109). Moreover,

$$\lim_{n\to\infty}\frac{1}{2}\sum_{i=0}^{2^n-1}\left(Z_{t_{i+1}}-Z_{t_i}\right)^2=\frac{t}{2}=\frac{1}{2}\int_0^t ds \quad , \tag{43}$$

again from the computations in Example 109.

In order to prove (40), we thus have to show that all remaining terms in (42) converge to zero. Indeed, we have:

$$\sum_{i=0}^{2^n-1}(t_{i+1}-t_i)^2=\sum_{i=0}^{2^n-1}\left(\frac{1}{2^n}\right)^2=\frac{1}{2^n}\underset{n\to\infty}{\to}0 \quad .$$

Moreover,

$$\sum_{i=0}^{2^n-1}(t_{i+1}-t_i)\left(Z_{t_{i+1}}-Z_{t_i}\right)=\frac{1}{2^n}\sum_{i=0}^{2^n-1}\left(Z_{t_{i+1}}-Z_{t_i}\right)=\frac{1}{2^n}(Z_t-Z_0) \quad .$$

Hence, for any $\omega\in\Omega$:

$$\sum_{i=0}^{2^n-1}(t_{i+1}-t_i)\left(Z_{t_{i+1}}-Z_{t_i}\right)(\omega)\underset{n\to\infty}{\to}0 \quad .$$

This concludes the proof. ■

Remark 113 *The above proof has been considerably simplified by the fact that the function*

$$g(t,Z)=\frac{1}{2}(t+Z)^2$$

can be locally exactly approximated by a sequence of second order Taylor expansions. In the more general case one will have to work with exact Taylor approximations of the form

$$\begin{aligned}
g\left(t_{i+1},Z_{t_{i+1}}\right)-g\left(t_i,Z_{t_i}\right) &= \frac{\partial g\left(t_i,Z_{t_i}\right)}{\partial t}(t_{i+1}-t_i)+\frac{\partial g\left(t_i,Z_{t_i}\right)}{\partial Z}\left(Z_{t_{i+1}}-Z_{t_i}\right)\\
&\quad \frac{1}{2}\left(\frac{\partial^2 g\left(t_i^*,Z_{t_i}^*\right)}{\partial^2 t}(t_{i+1}-t_i)^2+\frac{\partial^2 g\left(t_i^*,Z_{t_i}^*\right)}{\partial^2 Z}\left(Z_{t_{i+1}}-Z_{t_i}\right)^2\right)\\
&\quad +\frac{\partial^2 g\left(t_i^*,Z_{t_i}^*\right)}{\partial t\partial Z}(t_{i+1}-t_i)\left(Z_{t_{i+1}}-Z_{t_i}\right) \quad ,
\end{aligned}$$

for some t_i^*, Z_i^* such that $t_i^* \in [t_i, t_{i+1}]$ and $Z_i^* \in \left[Z_{t_i}, Z_{t_{i+1}}\right]$, and show that the residual approximation error goes to 0 as $n \to \infty$.

The above simplified proof of Itô's formula shows why the non standard second derivative term

$$\frac{1}{2}\int_0^t \frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 Z} ds$$

appears. In fact, we have shown for our specific example that

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}\frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 Z}\left(Z_{t_{i+1}}-Z_{t_i}\right)^2 = \lim_{n\to\infty}\sum_{i=0}^{2^n-1}\left(Z_{t_{i+1}}-Z_{t_i}\right)^2 = \int_0^t ds = \int_0^t \frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 Z} ds \quad .$$

Notice that

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}\left(Z_{t_{i+1}}-Z_{t_i}\right)^2 = t \quad , \tag{44}$$

is the quadratic variation of the Brownian motion process on the interval $[0,t]$, which is of order t and thus not[6] zero. Therefore, the further term in Itô's formula derives precisely from the non zero quadratic variation of Brownian motion.

By contrast to that, we have shown that

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}\frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial^2 t}\left(t_{i+1}-t_i\right)^2 = \lim_{n\to\infty}\sum_{i=0}^{2^n-1}\left(t_{i+1}-t_i\right)^2 = 0 \quad ,$$

i.e. that the second order derivative terms arising from the dependence on the deterministic argument t have zero quadratic variation, and thus do not contribute to Itô's formula.

Similarly, for the mixed second order derivative terms we have shown

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}\frac{\partial^2 g\left(t_i, Z_{t_i}\right)}{\partial t\partial Z}\left(t_{i+1}-t_i\right)\left(Z_{t_{i+1}}-Z_{t_i}\right) = \lim_{n\to\infty}\sum_{i=0}^{2^n-1}\left(t_{i+1}-t_i\right)\left(Z_{t_{i+1}}-Z_{t_i}\right) = 0 \quad ,$$

[6] In particular, this implies that Brownian motion has non differentiable trajectories, because otherwise one would get

$$\sum_{i=0}^{2^n-1}\left(Z_{t_{i+1}}-Z_{t_i}\right)^2 = \sum_{i=0}^{2^n-1}\left(\frac{d}{dt}Z_{t_i^*}\left(t_{i+1}-t_i\right)\right)^2 = \frac{1}{2^n}\sum_{i=0}^{2^n-1}\left(\frac{d}{dt}Z_{t_i^*}\right)^2\left(t_{i+1}-t_i\right) \quad ,$$

where $t_i^* \in [t_i, t_{i+1}]$, and, in the limit,

$$\sum_{i=0}^{2^n-1}\left(Z_{t_{i+1}}-Z_{t_i}\right)^2 \underset{n\to\infty}{\rightarrow} \lim_{n\to\infty}\frac{1}{2^n}\cdot\int_0^t\left(\frac{d}{dt}Z_s\right)^2 ds = 0 \quad ,$$

i.e. a contradiction with (43).

i.e. they also do not contribute to Itô's formula. In that case, the contribution is zero because the quadratic cross variation

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}(t_{i+1}-t_i)\left(Z_{t_{i+1}}-Z_{t_i}\right)$$

between t and Z_t is zero.

Based on these considerations, a simple mechanical rule can be motivated to compute Itô differentials. It consists in computing first a second order Taylor "differential" and then in defining second order differentials in the single variables according to the simple "multiplications rule":

	dt	dZ_t
dt	0	0
dZ_t	0	dt

.

In a mechanical way, this gives Itô's formula as

$$\begin{aligned} dg(t,Z_t) &= \frac{\partial g(t,Z_t)}{\partial t}dt+\frac{\partial g(t,Z_t)}{\partial Z}dZ_t+\frac{1}{2}\frac{\partial^2 g(t,Z_t)}{\partial^2 t}(dt)^2+\frac{1}{2}\frac{\partial^2 g(t,Z_t)}{\partial^2 Z}(dZ_t)^2 \\ &\quad+\frac{\partial^2 g(t,Z_t)}{\partial t\partial Z}dtdZ_t \\ &= \left(\frac{\partial g(t,Z_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,Z_t)}{\partial^2 Z}\right)dt+\frac{\partial g(t,Z_t)}{\partial Z}dZ_t \quad , \end{aligned}$$

by using the multiplication rule $(dt)^2 = dtdZ_t = 0$ and $(dZ_t)^2 = dt$.

Example 114 *Consider a process $X := (X_t, \mathcal{G}_t)_{0\le t\le T}$ satisfying the stochastic differential*

$$dX_t = K_t dt + H_t dZ_t \quad ,$$

for given X_0 and for some adapted processes $K := (K_t, \mathcal{G}_t)_{0\le t\le T}$ and $H := (H_t,\mathcal{G}_t)_{0\le t\le T}$ such that $H \in \mathcal{H}$ and $K \in \mathcal{K}$, where

$$\mathcal{K} := \left\{(\mathcal{G}_t)_{0\le t\le T} - \text{adapted processes } (K_t)_{0\le t\le T} \mid \int_0^T K_t dt < \infty \ P-a.s\right\}.$$

X is called an Itô process. By applying the above multiplication rules it then follows for any function f of class C^2:

$$\begin{aligned} df(X_t) &= f'(X_t)\,dX_t+\frac{1}{2}f''(X_t)(dX_t)^2 \\ &= f'(X_t)(K_t dt+H_t dZ_t)+\frac{1}{2}f''(X_t)\left(K_t^2(dt)^2+H_t^2(dZ_t)^2+2K_tH_tdtdZ_t\right) \\ &= f'(X_t)(K_t dt+H_t dZ_t)+\frac{1}{2}f''(X_t)H_t^2 dt \quad . \end{aligned}$$

5.4 An Application of Stochastic Calculus: the Black-Scholes Model

By means of Itô's calculus we are now endowed with the analytical tool that permits us to extend the set of self-financing strategies in Black and Scholes model in a way that will make any European contingent claim in the model perfectly hedgeable.

5.4.1 The Black-Scholes Market

The model structure is:

- $I := [0, T]$ is a continuous time index representing the available transaction dates in the model
- The sample space is given by $\Omega := \mathbb{R}^{[0,T]}$ with single outcomes ω of the form

$$\omega = (\omega_t)_{t\in[0,T]} \quad ,$$

  where $\omega_t \in \mathbb{R}$, $t \in [0, T]$.
- A Brownian motion process $Z := (Z_t, \mathcal{G}_t)_{t\in[0,T]}$ on $(\Omega, \mathcal{G}, P)$, where $(\mathcal{G}_t)_{t\in[0,T]}$ is the natural filtration associated to Z.
- Dynamics of the stock price and money account:

$$\begin{aligned} S_t &= S_0 \exp\left(\sigma Z_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right) \quad , \qquad (45) \\ B_t &= B_0 \exp(rt) \quad , \end{aligned}$$

  for some μ, r, $\sigma > 0$, for given $B_0 = 1$, S_0. In differential form this gives

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma S_t dZ_t \quad , \\ dB_t &= r B_t dt \quad . \end{aligned}$$

5.4.2 Self Financing Portfolios and Hedging in the Black-Scholes Model

Definition 115 *A self-financing strategy in the Black and Scholes model is an adapted process* $\Delta := (\Delta_t, \mathcal{G}_t)_{t\in[0,T]} \in \mathcal{H}$ *with value process* $X := (X_t, \mathcal{G}_t)_{t\in[0,T]}$ *such that*

$$dX_t = \Delta_t dS_t + \frac{(X_t - \Delta_t S_t)}{B_t} dB_t = \Delta_t dS_t + (X_t - \Delta_t S_t)\, r dt \quad ,$$

for given X_0. We will implicitly require in the sequel the integrability condition $\Delta \cdot S \in \mathcal{H}$.

Remark 116 *The continuous time definition of a self-financing strategy is the direct extension of the one introduced for the discrete time setting. In particular, one has, using the risky asset dynamics:*

$$dX_t = \Delta_t (dS_t - S_t r dt) + X_t r dt = [rX_t + \Delta_t (\mu - r) S_t] dt + \Delta_t \sigma S_t dZ_t \quad . \tag{46}$$

Our goal is to hedge European derivatives defined by some $\mathcal{G}_T-$measurable pay-off given by

$$v(T, S_T) = g(S_T) \quad ,$$

for a given continuous function g. We denote by $v(t, S_t)$ the price of the derivative at time $t \in [0, T]$ and assume that v is of class $C^{1,2}$ (in order to apply Itô's Lemma). By Itô's Lemma the dynamics of $v_t := v(t, S_t)$ are

$$\begin{aligned} dv_t &= \partial_t v_t dt + \frac{1}{2} \partial_{SS}^2 v_t \cdot (dS_t)^2 + \partial_S v_t \cdot dS_t \\ &= \left[\partial_t v_t + \frac{\sigma^2 S_t^2}{2} \partial_{SS}^2 v_t + \mu S_t \partial_S v_t \right] dt + \sigma S_t \partial_S v_t dZ_t \quad . \end{aligned} \tag{47}$$

In order for a self-financed portfolio Δ with value process X to be a perfect hedge for $(v_t)_{t \in [0,T]}$ the following hedging condition has to be satisfied:

$$X_t = v_t \quad , \quad t \in [0, T] \quad . \tag{48}$$

This imposes a strong restriction on the joint dynamics (46), (47), which have to coincide, implying:

$$\begin{aligned} \Delta_t \sigma S_t &= \partial_S v \sigma S_t \quad , \\ rX_t + \Delta_t (\mu - r) S_t &= \partial_t v + \mu S_t \partial_S v + \frac{\sigma^2 S_t^2}{2} \partial_{SS}^2 v \quad . \end{aligned}$$

Therefore we get, from the first of these two equations,

$$\Delta_t = \partial_S v_t \quad ,$$

i.e the delta of the portfolio. Inserting the delta in the second equation together with the perfect hedging condition (48) finally gives the partial differential equation (PDE)

$$\partial v + rS \partial_S v + \frac{\sigma^2 S_t^2}{2} \partial_{SS}^2 v = rv \quad , \tag{49}$$

for the function $v(t,S)$, subject to the boundary condition

$$v(T,S) = g(S) \quad , \quad S > 0 \quad .$$

This is Black-Scholes partial differential equation for the price of an European derivative. Solving this equation for the case

$$g(S) = (S-K)^{+} \quad ,$$

gives the call Black-Scholes pricing formula $v_{BS}(t,S)$ in Proposition 93. Computing $\partial_S v_{BS}$ based on the formula in Proposition 93 gives the delta of the call as

$$\Delta_t = \partial_S v_{BS}(t, S_t) = \mathcal{N}(d_{1t}) \quad ,$$

where

$$d_{1t} = \left. \frac{\log\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{(T-t)}} \right|_{S=S_t} \quad .$$

We remark that since $|\mathcal{N}(d_{1t})| \leq 1$ one has $(\mathcal{N}(d_{1t}) S_t, \mathcal{G}_t)_{t\in[0,T]} \in \mathcal{H}$, as initially assumed.

5.4.3 Probabilistic Interpretation of Black-Scholes Prices: Girsanov Theorem once more

It is important to remark that the fundamental PDE (49) does not depend on the expected return parameter[7] μ. This gives us the possibility to provide a probabilistic interpretation of Black and Scholes formula, which can be written as a discounted conditional expectation under a risk neutral martingale measure in the model. In other words, this allows us to give a probabilistic interpretation of pricing functions that are solutions of specific PDE's.

To highlight this point, rewrite the stock price dynamics as

$$dS_t = \mu S_t dt + \sigma S_t dZ_t = r S_t dt + \sigma S_t (dZ_t + \theta dt) = r S_t dt + \sigma S_t d\widetilde{Z}_t \quad ,$$

where $\widetilde{Z}_t = Z + \theta t$ and $\theta = (\mu - r)/\sigma$ is the market price of risk. Notice, that if we could find an equivalent probability measure $\widetilde{P}$ such that the process $\left(\widetilde{Z}_t, \mathcal{G}_t\right)_{t\in[0,T]}$ is Brownian motion under

[7] Therefore, the pricing of a derivative does not depend on the market expectations on the risky asset returns.

$\widetilde{P}$, then we would have that under $\widetilde{P}$ the stock price process is a geometric Brownian motion with dynamics

$$dS_t = rS_t dt + \sigma S_t d\widetilde{Z}_t \quad . \tag{50}$$

By replicating all arguments in the above section when using the $\widetilde{P}$ dynamics (50) with drift r we would then obtain again precisely the PDE (49) for the price function $v(t, S)$ of an European derivative. Therefore, changing in this way the measure does not alter the functional form for the pricing formula $v(t, S)$.

As usual the desired change of probability measure is provided by a version of Girsanov Theorem. We give a version of this theorem for the present setting, which is an immediate consequence of the proofs developed in the semicontinuous model setting.

Corollary 117 *In the Black and Scholes model with stock price dynamics (45) a risk neutral martingale measure $\widetilde{P}$ on $(\Omega, \mathcal{G}_T)$ is obtained by setting for any $A \in \mathcal{G}_T$,*

$$\widetilde{P}(A) := \int_A \exp\left(-\theta Z_T - \frac{\theta^2 T}{2}\right) dP \quad ,$$

where

$$\theta = \frac{\mu - r}{\sigma} \quad ,$$

is the market price of risk in the model.

Proof. The proof follows the same arguments as those for the version of Girsanov Theorem provided in the semicontinuous setting. ■

The key feature of risk neutral probabilities is that discounted prices of self-financed portfolios (and thus also discounted prices of hedge portfolios) are martingales. This allows us to write today's price function of a derivative as the discounted risk neutral expectation of its terminal pay off.

Specifically, let X_t be the $t-$time value of a self-financed portfolio in the Black Schole model and define by $\widetilde{X}_t := X_t/B_t = X_t \exp(-rt)$ the discounted portfolio value. By Itô's Lemma we then have under $\widetilde{P}$ (cf. also (46)):

$$\begin{aligned} d\widetilde{X}_t &= X_t \frac{d}{dt}(\exp(-rt))\,dt + \frac{1}{B_t} dX_t + \frac{1}{B_t}\cdot 0 \\ &= -\frac{X_t}{B_t} r dt + \frac{1}{B_t}\left(rX_t dt + \Delta_t \sigma S_t d\widetilde{Z}_t\right) \\ &= \frac{\Delta_t S_t}{B_t}\sigma d\widetilde{Z}_t \quad , \end{aligned}$$

that is $\left(\widetilde{X}_t, \mathcal{G}_t\right)_{t\in[0,T]}$ is a martingale under $\widetilde{P}$, provided that $\Delta \cdot S \in \mathcal{H}$. For the hedge portfolio

$$\Delta_t = \partial_S v(t, S_t) \quad ,$$

this gives

$$v(0,S_0) = \frac{v(0,S_0)}{B_0} = \frac{X_0}{B_0} = \widetilde{E}\left(\left.\frac{X_T}{B_T}\right| \mathcal{G}_0\right) = \widetilde{E}\left(\frac{v(T,S_T)}{B_T}\right) = \frac{1}{B_T}\widetilde{E}\left(g(S_T)\right|\mathcal{G}_0) \quad , \tag{51}$$

which writes $v(0, S_0)$ as a discounted expectation of the terminal pay-off $g(S_T)$, conditional on the initial condition $S = S_0$.

For the case $g(S) = (S-K)^+$ we computed this expectation in the semicontinuous model, providing Black and Scholes call price formula. Moreover, this formula is at the same time the solution of the PDE (49) to which we already gave the probabilistic interpretation (51). Finally, to compute the hedging strategy in the Black and Scholes model we just have to compute the derivative

$$\left.\frac{\partial v(t,S)}{\partial S}\right|_{S=S_t} = \frac{B_t}{B_T}\cdot\frac{\partial \widetilde{E}\left(g(S_T)\right| S_t = S)}{\partial S} \quad .$$

In the call option case, this gives after some algebra

$$\left.\frac{\partial v(t,S)}{\partial S}\right|_{S=S_t} = \mathcal{N}(d_{1t}) \quad ,$$

using the explicit expression for

$$\frac{1}{B_T}\widetilde{E}\left(\left.(S_T-K)^+\right|\mathcal{G}_0\right)$$

obtained in Proposition 93.