MATHEMATICAL FINANCE
THEORY, MODELING, IMPLEMENTATION

CHRISTIAN FRIES
Sample 1

The following is Chapter 1 of Mathematical Finance: Theory, Modeling, Implementation. ISBN 0-470-04722-4.
CHAPTER 1

Introduction

1.1 Theory, Modeling, and Implementation

This book tries to give a balanced representation of the theoretical foundations of mathematical finance, especially derivative pricing, state-of-the-art models, which are actually used in practice, and their implementation.

In practice, none of the three aspects—theory, modeling, and implementation—can be considered alone. Knowledge of the theory is worthless if it isn’t applied. Theory provides the tools for consistent modeling. A model without implementation is essentially worthless. Good implementation requires a deep understanding of the model and the underlying theory.

With this in mind, the book tries to build a bridge from academia to practice and from theory to object-oriented implementation.

1.2 Interest Rate Models and Interest Rate Derivatives

The text concentrates on the modeling of interest rates as stochastic (undetermined) quantities and the evaluation of interest rate derivatives under such models. However, this is not a specialization! Although the mathematical modeling of stock prices was the historical starting point and interest rates were assumed to be constant, some important theoretical aspects are significant only for stochastic interest rates (e.g. the change of numéraire technique). So for didactic reasons it is meaningful to start with interest rate models. Another reason to start with interest rate models is that interest rate models are the foundation of hybrid models. Since the numéraire, the reference asset, is most likely an interest-rate-related product, a need for stochastic interest
rates implies the need to build upon an interest rate model; see Figures 1.1 and 1.2. We will do so in Chapter 29. Nevertheless, the first model studied will be, of course, the Black-Scholes model for a single stock, after which we will move to stochastic interest rates.

Figure 1.1. Hybrid Models: The numéraire, the reference asset in the modeling of price processes, is most likely an interest rate product. This choice is not mathematically necessary but common for almost all models. Interest rate processes are the natural starting point for the modeling of price processes.

Figure 1.2. The Black-Scholes model may be interpreted as a hybrid model with deterministic interest rates. The solution of \( dB(t) = rB(t) dt \) is \( B(0) \exp (r t) \), i.e. it is deterministic and given in closed form. Thus the interest rate component is trivial. Within a LIBOR market model the interest rate is a stochastic quantity. This also changes properties of the stock process.
1.3 About This Book

1.3.1 How to Read This Book

The text may be read in a nonlinear way, i.e., the chapters have been kept as free-standing as possible. Chapter 2 provides the foundations in the order of their dependence. The reader familiar with the concepts of stochastic processes and martingales may skip the chapter and use it as reference only. To get a feeling for the mathematical concepts, one should read the special sections Interpretation and Motivation. Readers familiar with programming and implementation may prefer Chapter 13 as an illustration of the basic concepts.

The appendix gives a selection of the results and techniques from diverse areas (linear algebra, calculus, optimization), which are used in the text and in the implementation, but which are less important for understanding the essential concepts.

1.3.2 Abridged Versions

For a crash course focusing on particular aspects some chapters may be skipped. What follows are a few suggestions in this direction.

1.3.2.1 Abridged version “Monte Carlo Pricing”

Foundations (Chapter 2) → Replication (Chapter 3)
→ Black-Scholes Model (Chapter 4)
→ Discretization / Monte-Carlo Simulation (Chapter 13)

1.3.2.2 Abridged version “LIBOR Market Model”

Foundations (Chapter 2) → Replication (Chapter 3)
→ Interest Rate Structures (Chapter 8) → Black Model (Chapter 10)
→ LIBOR Market Model (Chapter 19)
→ Instantaneous and Terminal Correlation (Chapter 25)
→ Shape of the Interest Rate Curve (Chapter 24)

1.3.2.3 Abridged version “Markov Functional Model”

Foundations (Chapter 2) → Replication (Chapter 3)
→ Interest Rate Structures (Chapter 8) → Black Model (Chapter 10)
→ The Density of the Underlying of a European Option (Chapter 5)
→ Markov Functional Models (Chapter 27)
1.3.3 Special Sections

The text contains special sections giving notes on interpretation, motivation, and practical aspects. These are marked by the following symbols:

- **Interpretation:** Provides an interpretation of the preceding topic. Casts light on purposes and practical aspects.

- **Motivation:** Provides motivation for the following topic. Sometimes notes deficiencies in the previous results.

- **Further Reading:** Suggested literature and associated topics.

- **Experiment:** Guide for a software experiment where aspects of the preceding topic can be explored.

- **Tip:** Hints for practical use and software implementation of the preceding topics.

1.3.4 Notation

We will model the time evolution of stocks or interest rates with random variables parametrized through a time parameter $t$. Such stochastic processes may depend on other parameters like maturity or interest rate period. We will separate these two different kinds of parameters by a semicolon—see Figure 1.3.
1.3. ABOUT THIS BOOK

\[ \begin{array}{cccc}
F(T; t, \omega) & F(T; t) & F(T) & F \\
\text{Value} & \text{Random Variable} & \text{Stochastic Process} & \text{Interest Rate Curve} \\
\end{array} \]

**Figure 1.3. On the notation.**

### 1.3.5 Feedback

Please help to improve this work! Please send error reports and suggestions to Christian Fries &lt;email@christian-fries.de&gt;.

Thank you.

### 1.3.6 Resources

In connection with this book the following resources are available:

- Interactive experiments and exercises:  
  [http://www.christian-fries.de/finmath/applets](http://www.christian-fries.de/finmath/applets)

- Java™ source code:  
  [http://www.finmath.net/](http://www.finmath.net/)

- Figures (in Color): The figures in this book are reproduced in black and white. The original color figures may be obtained from [http://www.christian-fries.de/finmath/book](http://www.christian-fries.de/finmath/book)

Sample 2

CHAPTER 2

Foundations

2.1 Probability Theory

Definition 1 (Probability Space, σ-Algebra):
Let Ω denote a set and \( \mathcal{F} \) a family of subsets of Ω. \( \mathcal{F} \) is a σ-algebra if

1. \( \emptyset \in \mathcal{F} \).
2. \( F \in \mathcal{F} \Rightarrow \Omega \setminus F \in \mathcal{F} \).
3. \( F_1, F_2, F_3, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} F_i \in \mathcal{F} \).

The pair \((\Omega, \mathcal{F})\) is a measurable space. A function \( P : \mathcal{F} \to [0, \infty) \) is a probability measure if

1. \( P(\emptyset) = 0, P(\Omega) = 1 \).
2. For \( F_1, F_2, F_3, \ldots \in \mathcal{F} \) mutually disjoint (i.e. \( i \neq j \Rightarrow F_i \cap F_j = \emptyset \)), we have

\[
P\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P(F_i).
\]

The triple \((\Omega, \mathcal{F}, P)\) is called probability space (if instead of 1 we require only \( P(\emptyset) = 0 \), then \( P \) is called measure and \((\Omega, \mathcal{F}, P)\) is called measure space).
Interpretation: The set \( \Omega \) may be interpreted as the set of elementary events. Only one such event may occur. The subset \( F \subset \Omega \) may then be interpreted as an event configuration, e.g. as if one asked only for a specific property of an event, a property that might be shared by more than one event. Then the complement of a set of events corresponds to the negation of the property in question, and the union of two subsets \( F_1, F_2 \subset \Omega \) corresponds to combining the questions for the two corresponding properties with an “or”. Likewise the intersection corresponds to an “and”: only those events that share both properties are part of the intersection. A \( \sigma \)-algebra may then be interpreted as a set of properties, e.g., the set of properties by which we may distinguish the events or the set of properties on which we may base decisions and answer questions. Thus the \( \sigma \)-algebra may be interpreted as information (on properties of events).

Thus a probability space \((\Omega, \mathcal{F}, P)\) may be interpreted as a set of elementary events, a family of properties of the events, and a map that assigns a probability to each property of the events, the probability that an event with the respective properties will occur.

Since conditional expectation will be one of the central concepts, we remind the reader of the notions of *conditional probability* and *independence*.

**Definition 2 (Independence, Conditional Probability):**

Let \((\Omega, \mathcal{F}, P)\) denote a probability space and \(A, B \in \mathcal{F}\).

1. We say that \(A\) and \(B\) are independent, if
   \[
P(A \cap B) = P(A) P(B).
   \]

2. For \(P(B) > 0\) we define the conditional probability of \(A\) under the hypothesis \(B\) as
   \[
P(A|B) := \frac{P(A \cap B)}{P(B)}.
   \]

The Borel \(\sigma\)-algebra \(\mathcal{B}(\mathbb{R})\) or \(\mathcal{B}(\mathbb{R}^n)\) plays a special role in integration theory. We define it next.

**Definition 3 (Borel \(\sigma\)-Algebra, Lebesgue Measure):**

Let \(n \in \mathbb{N}\) and \(a_i < b_i (i = 1, \ldots, n)\). By \(\mathcal{B}(\mathbb{R}^n)\) we denote the smallest \(\sigma\)-algebra for which
\[
(a_1, b_1) \times \cdots \times (a_n, b_n) \in \mathcal{B}(\mathbb{R}^n).
\]

\(\mathcal{B}(\mathbb{R}^n)\) is called the Borel \(\sigma\)-algebra. The measure \(\lambda\) defined on \(\mathcal{B}(\mathbb{R}^n)\) with
\[
\lambda((a_1, b_1) \times \cdots \times (a_n, b_n)) := \prod_{i=1}^{n} (b_i - a_i)
\]
is called a *Lebesgue measure* on \(\mathcal{B}(\mathbb{R}^n)\).
**Remark 4 (Lebesgue Measure):** Obviously the Lebesgue measure is not a probability measure on \((\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))\) since \(\lambda(\mathbb{R}^n) = \infty\). It will be needed in the discussion of Lebesgue integration and we give the definition merely for completeness.\(^1\)

**Definition 5 (Measurable, Random Variable):**

Let \((\Omega, \mathcal{F})\) and \((S, \mathcal{S})\) denote two measurable spaces.

1. A map \(T : \Omega \mapsto S\) is called \((\mathcal{F}, \mathcal{S})\)-measurable if\(^2\)

   \[T^{-1}(A) \in \mathcal{F} \text{ for all } A \in \mathcal{S}.\]

   If \(T : \Omega \mapsto S\) is a \((\mathcal{F}, \mathcal{S})\)-measurable map we write more concisely

   \[T : (\Omega, \mathcal{F}) \mapsto (S, \mathcal{S}).\]

2. A measurable map \(X : (\Omega, \mathcal{F}) \mapsto (S, \mathcal{S})\) is also called a random variable. A random variable \(X : (\Omega, \mathcal{F}) \mapsto (S, \mathcal{S})\) is called a \(n\)-dimensional real-valued random variable if \(S = \mathbb{R}^n\) and \(\mathcal{S} = \mathcal{B}(\mathbb{R}^n)\).

We are interested in the probability for which a given random variable attains a certain value or range of values. This is given by the following definition.

**Definition 6 (Image Measure):**

Let \(X : (\Omega, \mathcal{F}) \mapsto (S, \mathcal{S})\) denote a random variable and \(P\) a measure on the measurable space \((\Omega, \mathcal{F})\). Then

\[P_X(A) := P(X^{-1}(A)) \quad \forall A \in \mathcal{S}\]

defines a probability measure on \((S, \mathcal{S})\), which we call the image measure of \(P\) with respect to \(X\).

**Interpretation:** A real-valued random variable assigns a real value (or vector of values) to each elementary event \(\omega\). This value may be interpreted as the result of an experiment, depending on the events. In our context the random variables mostly stand for payments or values of financial products depending on the state of the world. How random a random variable is depends on the random variable itself. The random variable that assigns the same value to all events \(\omega\) exhibits no randomness at all. If we could observe

\(^1\) The Lebesgue measure measures intervals \((n = 1)\) according to their length, rectangles \((n = 2)\) according to their area, and cubes \((n = 3)\) according to their volume.

\(^2\) We define \(T^{-1}(A) := \{\omega \in \Omega \mid T(\omega) \in A\}\).
only the result of such an experiment (random variable), we would not be able to say anything about the state of the world $\omega$ that led to the result.

The image measure is the probability measure induced by the probability measure $P$ (a probability measure on $(\Omega, \mathcal{F})$) and the map $X$ on the image space $(S, S)$.

The property of being measurable may be interpreted as the property that the distinguishable events in the image space $(S, S)$ are not finer (better distinguishable) than the events in the preimage space $(\Omega, \mathcal{F})$. Only then it is possible to use the probability measure $P$ on $(\Omega, \mathcal{F})$ to define a probability measure on $(S, S)$, see Figure 2.1.

![Figure 2.1](http://www.christian-fries.de/finmath/book)

**Figure 2.1.** Illustration of measurability: The random variables $X$ and $Z$ assign a gray value to each elementary event $\omega_1, \ldots, \omega_{10}$ as shown. The $\sigma$-algebra $\mathcal{F}$ is generated by the sets $F_1 = \{\omega_1, \omega_2, \omega_3\}$, $F_2 = \{\omega_4, \omega_5, \omega_6\}$, $F_3 = \{\omega_7, \ldots, \omega_{10}\}$. The random variable $X$ is measurable with respect to $\mathcal{F}$, the random variable $Z$ is not measurable with respect to $\mathcal{F}$.

**Exercise:** Let $X$ be as in Definition 6. Show that

$$\{X^{-1}(A) \mid A \in S\}$$

is a $\sigma$-algebra. What would be an interpretation of $X^{-1}(A)$?

**Motivation:** We will now define the Lebesgue integral and give an interpretation and a comparison to the (possibly more familiar) Riemann integral.

The definition of the Lebesgue integral is not only given to prepare the definition of the conditional expectation (Definition 15). The definition will also show the construction of the Lebesgue integral and we will later use similar steps to construct the Itô integral.
Definition 7 (Integral, Lebesgue Integral): Let \((\Omega, \mathcal{F}, \mu)\) denote a measure space.

1. Let \(f\) denote a \((\mathcal{F}, \mathcal{B}(\mathbb{R}))\)-measurable real-valued, nonnegative map. \(f\) is called an **elementary function** if \(f\) takes on only a finite number of values \(a_1, \ldots, a_n\). For an elementary function we define
   \[
   \int_{\Omega} f(\omega) \, d\mu(\omega) := \sum_{i=1}^{n} a_i \mu(A_i)
   \]
   where \(A_i := f^{-1}([a_i]) \Rightarrow A_i \in \mathcal{F}\) as the (Lebesgue) integral of \(f\).

2. Let \(f\) denote a nonnegative map defined on \(\Omega\), such that a monotonically increasing sequence \((u_k)_{k \in \mathbb{N}}\) of elementary maps with \(f := \sup_{k \in \mathbb{N}} u_k\) exists. Then
   \[
   \int_{\Omega} f(\omega) \, d\mu(\omega) := \sup_{k \in \mathbb{N}} \int_{\Omega} u_k(\omega) \, d\mu(\omega)
   \]
   is unique and is called the (Lebesgue) integral of \(f\).

3. Let \(f\) denote a map on \(\Omega\) such that we have for \(f^+ := \max(f, 0)\) and \(f^- := \max(-f, 0)\), respectively, a monotone increasing sequence of elementary maps as in the previous definition. Furthermore we require that \(\int_{\Omega} f^\pm \, d\mu < \infty\). Then \(f\) is called **integrable** with respect to \(\mu\) and we define
   \[
   \int_{\Omega} f \, d\mu := \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu
   \]
   as the (Lebesgue) integral of \(f\).

Remark 8 (\(\int f(x) \, dx, \int f(t) \, dt\)): If the measure \(\mu\) is the Lebesgue measure \(\mu = \lambda\) we use the shortened notation
   \[
   \int_A f(x) \, d\lambda(x) := \int_A f(x) \, dx.
   \]
In this case \(\Omega = \mathbb{R}^n\) and we denote the elements of \(\Omega\) by latin letters, e.g. \(x\) (instead of \(\omega\)). If the elements of \(\Omega = \mathbb{R}\) have the interpretation of a time we usually denote them by \(t\).

---

\(^3\) We have \((a_i - \frac{1}{n}, a_i + \frac{1}{n}) \in \mathcal{B}(\mathbb{R})\) by definition. Then \([a_i] = \bigcap_{n=1}^{\infty} (a_i - \frac{1}{n}, a_i + \frac{1}{n}) \in \mathcal{B}(\mathbb{R})\).
**Theorem 9 (measurable ⇔ integrable for nonnegative maps):** For a non-negative map $f$ on $\Omega$ a monotone increasing sequence $(u_k)_{k \in \mathbb{N}}$ of elementary maps with $f = \sup_{k \in \mathbb{N}} u_k$ exists if and only if $f$ is $\mathcal{F}$-measurable.

**Proof:** See [1], §12.

**Interpretation:** To develop an understanding of the (Lebesgue) integral we consider its definition for elementary maps:

$$
\int_{\Omega} f(\omega) \, d\mu(\omega) = \sum_{i=1}^{n} a_i \mu(A_i), \quad A_i := f^{-1}(\{a_i\}).
$$

The Lebesgue integral of an (elementary) map $f$ is the weighted sum of the function values $a_i$ of $f$, each weighted by the measure $\mu(A_i)$ of the set on which this value is attained (i.e., $A_i = f^{-1}(\{a_i\})$).

If in addition we have $\mu(\Omega) = 1$, where $\Omega$ is the domain of $f$, then the integral is a weighted average of the function values $a_i$ of $f$.

For a real-valued (elementary) function of a real-valued argument, e.g., $f : [a, b] \mapsto \mathbb{R}$, and the Lebesgue measure, the integral corresponds to the naive concept of an integral as being the sum all rectangles given by

$$
\text{base area (Lebesgue measure of the interval)} \times \text{height (function value)}
$$

which is also the concept behind the Riemann integral.

Part 2 and 3 of Definition 7 extend this concept via a limit approximation to more general functions.

**Excursus: On the Difference between Lebesgue and Riemann Integrals**

The construction of the Lebesgue integral differs from the construction of the Riemann integral (which is perhaps more familiar) in the way the sets $A_i$ are chosen. The Riemann integral starts from a given partition of the domain and multiplies the

---

4 An understanding of the difference between the Lebesgue and Riemann integrals does not play a major role in the following text. The excursus can safely be skipped. It should serve to satisfy curiosity, e.g., if the concept of a Riemann integral is more familiar.
2.1. PROBABILITY THEORY

Figure 2.2. Lebesgue integral versus Riemann integral.

size of each subinterval by a corresponding functional value of (any) chosen point belonging to that interval (e.g., the center point). The Lebesgue integral chooses the partition as preimage $f^{-1}([a_i])$ of given function values $a_i$. In short: the Riemann integral partitions the domain of $f$, the Lebesgue integral partitions the range of $f$. For elementary functions both approaches give the same integral value; see Figure 2.2. For general functions the corresponding integrals are defined as the limit of a sequence of approximating elementary functions (if it exists). Here, the two concepts are different: In the limit, all Riemann integrable functions are Lebesgue integrable, and the two limits give the same value for the integral. However, there exist Lebesgue integrable functions for which the Riemann integral is not defined (its limit construction does not converge).

Definition 10 (Distribution):
Let $P$ denote a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (e.g., the image measure of a random variable). The function

$$F_P(x) := P((-\infty, x))$$

is called the distribution function\(^5\) of $P$. If $P$ denotes a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ the $n$-dimensional distribution function is defined as

$$F_P(x_1, \ldots, x_n) := P((-\infty, x_1) \times \cdots \times (-\infty, x_n)).$$

\(^5\) We have $(-\infty, x) \in \mathcal{B}(\mathbb{R})$ since $(-\infty, x) = \bigcup_{i=1}^{\infty} (x - i, x)$, see Definition 1.
The distribution function of a random variable $X$ is defined as the distribution function of its image measure $P_X$ (see Definition 6).

**Definition 11 (Density):**

Let $F_P$ denote the distribution function of a probability measure $P$. If $F_P$ is differentiable, we define

$$
\phi(x) := \frac{\partial}{\partial x} F_p(x)
$$

as the density of $P$. If $P$ is the image measure of some random variable $X$, we also say that $\phi$ is the density of $X$.

**Remark 12 (Integration Using a Known Density/Distribution):** To calculate the integral of a function of a random variable (e.g., to calculate expectation or variance), it is sufficient to know the density or distribution function of the random variable. Let $g$ denote a sufficiently smooth function and $X$ a random variable on $(\Omega, \mathcal{F}, P)$; then we have

$$
\int_{\Omega} g(X(\omega)) \, dP(\omega) = \int_{-\infty}^{\infty} g(x) \, dF_{X}(x) = \int_{-\infty}^{\infty} g(x)\phi(x) \, dx,
$$

where $F_{X}$ denotes the distribution function of $X$ and $\phi$ the density of $X$ (i.e., of $P_X$). In this case it is neither necessary to know the underlying space $\Omega$, the measure $P$, nor how $X$ is modeled (i.e., defined) on this space.

**Definition 13 (Independence of Random Variables):**

Let $X : (\Omega, \mathcal{F}) \mapsto (S, S)$ and $Y : (\Omega, \mathcal{F}) \mapsto (S, S)$ denote two random variables. $X$ and $Y$ are called independent, if for all $A, B \in S$ the events $X^{-1}(A)$ and $Y^{-1}(B)$ are independent in the sense of Definition 2.

**Remark 14 (Independence):** For $i = 1, \ldots, n$ let $X_i : \Omega \mapsto \mathbb{R}$ denote random variables with distribution functions $F_{X_i}$ and let $F_{(X_1, \ldots, X_n)}$ denote the distribution function of $(X_1, \ldots, X_n) : \Omega \mapsto \mathbb{R}^n$. Then the $X_i$ are pairwise independent if and only if

$$
F_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n).
$$

**Definition 15 (Expectation, Conditional Expectation):**

Let $X$ denote a real-valued random variable on the probability space $(\Omega, \mathcal{F}, P)$.

1. If $X$ is $P$-integrable, we define

$$
E^P(X) := \int_{\Omega} X \, dP
$$

as the expectation of $X$. 

http://www.christian-fries.de/finmath/book
2. Furthermore, let \( F_i \in \mathcal{F} \) with \( P(F_i) > 0 \). Then
\[
E^P(X|F_i) := \frac{1}{P(F_i)} \int_{F_i} X \, dP
\]
is called the \textit{conditional expectation of } \( X \text{ under (the hypothesis) } F_i \).

\[\text{Theorem 16 (Conditional Expectation}\,^6\,\text{:}} \quad \text{Let } X \text{ denote a real-valued random variable on } (\Omega, \mathcal{F}, P), \text{ either nonnegative or integrable. Then we have for each } \sigma\text{-algebra } C \subset \mathcal{F} \text{ a nonnegative or integrable real-valued random variable } X|_C \text{ on } \Omega, \text{ unique in the sense of almost sure equality,}^7, \text{ such that } X|_C \text{ is } C\text{-measurable and}
\[
\forall C \in C : \int_C X|_C \, dP = \int_C X \, dP, \quad \text{i.e., } E^P(X|C) = E^P(X|C).
\]

We will discuss the interpretation of this theorem after giving a name to \( X|_C \):

\[\text{Definition 17 (Conditional Expectation (continued)):}} \quad \text{Under the assumptions and with the notation of Theorem 16 we define:}
\]
1. The random variable \( X|_C \) is called the \textit{conditional expectation of } \( X \text{ under (the hypothesis) } C \) and is denoted by
\[
E^P(X|C) := X|_C.
\]
2. Let \( Y \) denote another random variable on the same measure space. We define:
\[
E^P(X|Y) := E(X|\sigma(Y)),
\]
where \( \sigma(Y) \) is the \( \sigma\)-algebra generated by \( Y \), i.e., the smallest \( \sigma\)-algebra, with respect to which \( Y \) is measurable, i.e., \( \sigma(Y) := \sigma(Y^{-1}(S)) \).

\[\text{Interpretation:}} \quad \text{First note that the two concepts of expectations from Definition 15 are just special cases of the conditional expectation defined in Definition 17, namely:}
\]
- Let \( C = \{\emptyset, \Omega\} \). Then \( E(X | C) = X|_C \) where \( X|_C(\omega) = E(X) \forall \omega \in \Omega. \)

\[\text{See [2], Chapter 15}\]
\[\text{A property holds } P\text{-\textit{almost surely} if the set of } \omega \in \Omega \text{ for which the property does not hold has measure zero.}\]
For $C = \{\emptyset, F, \Omega \setminus F, \Omega\}$ we have $X|C(\omega) = \begin{cases} E(X | F) & \text{if } \omega \in F \\ E(X | \Omega \setminus F) & \text{if } \omega \in \Omega \setminus F \end{cases}$, with $E(X) = P(F) E(X|F) + (1 - P(F)) E(X|\Omega \setminus F)$.

The conditional expectation is a random variable that is derived from $X$ such that only events (sets) in $C$ can be distinguished. In the first case we have a very coarse $C$ and the image of $X|C$ contains only the expectation $E(X)$. This is the smallest piece of information on $X$. As $C$ becomes finer, more and more information about $X$ becomes visible in $X|C$. Furthermore, if $X$ itself is $C$-measurable, then $X$ and $X|C$ are ($P$-almost surely) indistinguishable.

![Figure 2.3](http://www.christian-fries.de/finmath/book)

**Figure 2.3.** Conditional expectation: Let the $\sigma$-algebra $C$ be generated by the sets $C_1 = \{\omega_1, \omega_2, \omega_3\}$, $C_2 = \{\omega_4, \omega_5, \omega_6\}$, $C_3 = \{\omega_7, \ldots, \omega_{10}\}$.

In this sense $C$ may be interpreted as an information set and $X|C$ as a filtered version of $X$. If it is only possible to make statements about events in $C$, then we can only make statements about $X$ which could also be made about $X|C$, see Figure 2.3.

### 2.2 Stochastic Processes

**Definition 18 (Stochastic Process):**

A family $X = \{X_t | 0 \leq t < \infty\}$ of random variables $X_t : (\Omega, \mathcal{F}) \to (S, S)$ is called (time continuous) stochastic process. If $(S, S) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we say that $X$ is a $d$-dimensional stochastic process. The family $X$ may also be interpreted as a function $X : [0, \infty) \times \Omega \to S$:

$$X(t, \omega) := X_t(\omega) \quad \forall (t, \omega) \in [0, \infty) \times \Omega.$$ 

If the range $(S, S)$ is not given explicitly, we assume $(S, S) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. 

http://www.christian-fries.de/finmath/book
Interpretation: The parameter $t$ obviously refers to time. For fixed $t \in [0, \infty)$ we view $X(t)$ as the outcome of an experiment at time $t$. Note that all random variables $X(t)$ are modeled over the same measurable space $(\Omega, \mathcal{F})$. Thus we do not assume a family $(\Omega_t, \mathcal{F}_t)$ of measurable spaces, one for each $X_t$. The stochastic process $X$ assigns a path to each $\omega \in \Omega$: For a fixed $\omega \in \Omega$ the path $X(\cdot, \omega) = \{ (t, X(t, \omega)) \mid t \in [0, \infty) \}$ is a sequence of outcomes of the random experiments $X_t$ (a trajectory) associated with a state $\omega$. Knowledge about $\omega \in \Omega$ implies knowledge of the whole history (past, present, and future) $X(\omega)$.

To model the different levels of knowledge and thus distinguish between past and future, we will define in Section 2.3 the concept of a filtration and an adapted process.

Definition 19 (Path):
Let $X$ denote a stochastic process. For a fixed $\omega \in \Omega$ the mapping $t \mapsto X(t, \omega)$ is called the path of $X$ (in state $\omega$).

Definition 20 (Equality of Stochastic Processes):
We define three notions of equality of stochastic processes:

1. Two stochastic processes $X$ and $Y$ are called indistinguishable if

$$P(X_t = Y_t : \forall 0 \leq t < \infty) = 1.$$ 

2. A stochastic process $Y$ is a modification of $X$ if

$$P(X_t = Y_t) = 1 : \forall 0 \leq t < \infty.$$ 

3. Two stochastic processes $X$ and $Y$ have the same finite-dimensional distributions, if

$$\forall n : \forall 0 \leq t_1 < t_2 < \cdots < t_n < \infty : \forall A \in \mathcal{B}(S^n) : P((X_{t_1}, \ldots, X_{t_n}) \in A) = P((Y_{t_1}, \ldots, Y_{t_n}) \in A).$$

Remark 21 (On the Equality of Stochastic Processes): While in Definition 20.3 only the distributions generated by the processes are considered, Definitions 20.1 and 20.2 consider the pointwise differences between the processes. The difference between 20.1 and 20.2 will become apparent in the following example:

Let $Z : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to ([-1, 1], \mathcal{B}([-1, 1]))$ be a random variable on $(\Omega, \mathcal{F}, P) = (\mathbb{R}, \mathcal{B}, \lambda)$ and $t \mapsto X(t) := t \cdot Z$ be a stochastic process. An interpretation of this process would be the position of a moving particle, having at time 0 the position 0 and the random speed $Z$. 

---

8 An interpretation of this process would be the position of a moving particle, having at time 0 the position 0 and the random speed $Z$. 

---

http://www.christian-fries.de/finmath/book
∀ \( A \in B([−1, 1]) \) and \( P(\{Z = x\}) = 0 \forall x \in [−1, 1] \), e.g., an equally distributed \( Z \).
Furthermore let \( \forall (t, \omega) \in [0, \infty) \times \mathbb{R} \)

\[
Y_1(t, \omega) := \begin{cases} 
X(t, \omega) & \text{for } t \neq \omega \\
-X(t, \omega) & \text{for } t = \omega
\end{cases}, \quad Y_2(t, \omega) := -X(t, \omega).
\]

The \( Y_1 \) is a modification of \( X \), since \( Y_1(t) \) differs from \( X(t) \) (for fixed \( t \)) only on a set with probability 0. However, \( X \) and \( Y_1 \) are not indistinguishable, since \( P(X(t) = Y_1(t) : \forall 0 \leq t < \infty) = \frac{1}{2} \) (the two processes are different on 50% of all paths). \( Y_2 \) is neither indistinguishable nor a modification of \( X \), but due to \( P(\{Z \in A\}) = P(\{Z \in -A\}) \forall A \in \mathcal{B}([-1, 1]) \) it fulfills condition 3 in Definition 20.

To summarize, condition 1 in Definition 20 considers the equality of the processes \( X, Y \), condition 2 in Definition 20 considers equality of the random variables \( X(t), Y(t) \) for fixed \( t \), and condition 3 in Definition 20 considers the equality of distributions. In our applications we are interested only in the distributions of processes.

### 2.3 Filtration

**Definition 22 (Filtration):**

Let \((\Omega, \mathcal{F})\) denote a measurable space. A family of \(\sigma\)-algebras \(\{\mathcal{F}_t | t \geq 0\}\), where

\[
\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \quad \text{for } 0 \leq s \leq t,
\]

is called a *filtration* on \((\Omega, \mathcal{F})\).

**Definition 23 (Generated Filtration):**

Let \(X\) denote a stochastic process on \((\Omega, \mathcal{F})\). We define

\[
\mathcal{F}_t^X := \sigma(X_s; 0 \leq s \leq t)
\]

:= the smallest \(\sigma\)-algebra with respect to which \(X_s\) is measurable \(\forall s \in [0, t]\).

**Definition 24 (Adapted Process):**

Let \(X\) denote a stochastic process on \((\Omega, \mathcal{F})\) and \(\{\mathcal{F}_t\}\) a filtration on \((\Omega, \mathcal{F})\). The process \(X\) is called \(\{\mathcal{F}_t\}\)-adapted, if \(X_t\) is \(\mathcal{F}_t\)-measurable for all \(t \geq 0\).
2.3. FILTRATION

Interpretation: In Figure 2.4 we depict a filtration of four \( \sigma \)-algebras with increasing refinement (left to right). The black borders surround the generators of the corresponding \( \sigma \)-algebra. If a stochastic process maps a gray value for each elementary event (or path) \( \omega_i \) of \( \Omega \) (left), then the process is adapted if it takes a constant gray value on the generators of the respective \( \sigma \)-algebra. If at time \( t_2 \) the process assigns to \( \omega_7 \) the same dark gray as to \( \omega_8 \), then the process is adapted, otherwise it is not.

By means of the conditional expectation (see Theorem 16 and the interpretation of Figure 2.3) we may create an adapted process from a given filtration \( \{ \mathcal{F}_t \mid t \geq 0 \} \) and an \( \mathcal{F} \)-measurable random variable \( Z \):

**Lemma 25 (Process of the Conditional Expectation):** Let \( \{ \mathcal{F}_t \mid t \geq 0 \} \) denote a filtration \( \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \) and \( Z \) an \( \mathcal{F} \) measurable random variable. Then

\[
X(t) := \mathbb{E}(Z \mid \mathcal{F}_t)
\]

is a \( \{ \mathcal{F}_t \} \)-adapted process.

This lemma shows how the filtration (and the corresponding adapted process) may be viewed as a model for information: The random variable \( X(t) \) in Lemma 25 allows with increasing \( t \) more and more specific statements about the nature of \( Z \). Compare this to the illustrations in Figure 2.1 and 2.3.

Figure 2.4. Illustration of a filtration and an adapted process.
The concepts of an adapted process only links random variables $X(t)$ to $\sigma$-algebras $F_t$ for any $t$. It does not necessarily imply that the stochastic process $X$ (interpreted as a random variable on $[0, \infty) \times \Omega$) is measurable. A stronger requirement is given by the following Definition.

**Definition 26 (Progressively Measurable):**

An $(n$-dimensional) stochastic process $X$ is called *progressively measurable* with respect to the filtration $\{F_t\}$ if for each $T > 0$ the mapping

$$X : ([0, T] \times \Omega, \mathcal{B}([0, T] \otimes F_t)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}^n))$$

is measurable.

**Remark 27:** Any progressively measurable process is measurable and adapted. Conversely a measurable and adapted process has a progressively measurable modification; see [20].

Another regularity requirement for stochastic processes is that of being *previsible*:

**Definition 28 (Previsible Process):**

Let $X$ denote a (real-valued) stochastic process on $(\Omega, \mathcal{F})$ and $\{F_t\}$ a filtration on $(\Omega, \mathcal{F})$. The process $X$ is called $\{F_t\}$-previsible, if $X$ is $\{F_t\}$-adapted and bounded with left continuous paths.

### 2.4 Brownian Motion

**Definition 29 (Brownian Motion):**

Let $W : [0, \infty) \times \Omega \to \mathbb{R}^n$ denote a stochastic process with the following properties:

1. $W(0) = 0$ ($P$-almost surely).
2. The map $t \mapsto W(t)$ is continuous ($P$-almost surely).
3. For given $t_0 < t_1 < \cdots < t_k$ the increments $W(t_1) - W(t_0), \ldots, W(t_k) - W(t_{k-1})$ are mutually independent.
4. For all $0 \leq s \leq t$ we have $W(t) - W(s) \sim N(0, (t-s)I_n)$, i.e., the increment is normally distributed with mean 0 and covariance matrix $(t-s)I_n$, where $I_n$ denotes the $n \times n$ identity matrix.

Then $W$ is called $(n$-dimensional) $P$-Brownian motion or a $(n$-dimensional) $P$-Wiener process.

We have not yet discussed the question of whether a process with such properties exists (it does). The question for its existence is nontrivial. For example, if we want to
replace normally distributed by lognormally distributed in property 4 in Definition 29 there would be no such process. If we set \( s = 0 \) in property 4, we see that we have prescribed the distribution of \( W(t) \) as well as the distribution of the increments \( W(t) - W(s) \).

**Remark 30 (Brownian Motion):** Property 4 is less axiomatic than one might assume: The central requirement is the independence of the increments together with the requirement that increments of the same time step size \( t - s \) have the same nonnegative variance (here \( t - s \)) and mean 0. That the increments are normally distributed is more a consequence than an requirement, see Theorem 31. This theorem also gives a construction of the Brownian motion.

Figure 2.5. *Time discretization of a Brownian motion: The transition \( \Delta W(T_i) \) from time \( T_i \) to \( T_{i+1} \) is normally distributed. The mean of the transition is 0, i.e., under the condition that at time \( T_i \) the state \( W(T_i) = x^* \) was attained, the (conditional) expectation of \( W(T_{i+1}) \) is \( x^* \): \( \mathbb{E}(W(T_{i+1}) \mid W(T_i) = x^*) = x^* \).*

**Tip (Time-Discrete Realizations):** In the following we will often consider the realizations of a stochastic process at discrete times \( 0 = T_0 < T_1 < \ldots < T_N \) only (e.g., this will be the case when we consider the implementation). If we need only the realizations \( W(T_i) \), we may generate them by the time-discrete increments \( \Delta W(T_i) = W(T_{i+1}) - W(T_i) \) since from Definition 29 we have \( W(T_i) = \sum_{k=0}^{i-1} \Delta W(T_k), W(T_0) := 0 \). See Figure 2.4. \(<|\)

---

\( ^9 \) Note that the sum of two (independent) normally distributed random variables is normally distributed, but the sum of two lognormally distributed random variables is not lognormal.
Sample 3

CHAPTER 13

Discretization of Time and State Space

In this chapter we present methods for discretization and implementation of Itô stochastic processes. We give an integrated presentation of path simulation (Monte Carlo simulation) and lattice methods (e.g., trees). Finally, we show how both methods can be combined; see Section 13.4.

Throughout our discussion of the discretization and implementation we will repeat some of the terms from Chapter 2, e.g., path, \( \sigma \)-algebra, filtration, process, and \( \mathcal{F}_t \)-adapted. Thus, this chapter will also serve as an illustration of some of the mathematical concepts from Chapter 2.

The discretization and implementation should not be seen as a minor additional step after the mathematical analysis and it should not be underestimated. The discretization and implementation allow us a second look, possibly providing further insights into a model.\(^1\)

In Figure 13.1 we give an overview of the steps involved in the discretization and implementation of Itô processes.

13.1 Discretization of Time: The Euler and the Milstein Schemes

As a first step we shall consider the discretization of time and present the Euler scheme and the Milstein scheme.

\(^1\) Indeed, it is common in mathematics to prove analytical results as a limit of a numerical, i.e., discrete, procedure.

http://www.christian-fries.de/finmath/book
Figure 13.1. Discretization and implementation of Itô processes
13.1. Definitions

Definition 177 (Euler Scheme):

Given an Itô process

\[ dX(t) = \mu(t, X(t)) \, dt + \sigma(t, X(t)) \, dW(t), \]

and a time discretization \( \{ t_i \mid i = 0, \ldots, n \} \) with \( 0 = t_0 < \ldots < t_n \), then the time-discrete stochastic process \( \tilde{X} \) defined by

\[ \tilde{X}(t_{i+1}) = \tilde{X}(t_i) + \mu(t_i, \tilde{X}(t_i)) \Delta t_i + \sigma(t_i, \tilde{X}(t_i)) \Delta W(t_i) \]

is called an Euler scheme of the process \( X \) (where \( \Delta t_i := t_{i+1} - t_i \) and \( \Delta W(t_i) := W(t_{i+1}) - W(t_i) \)).

Interpretation (Euler Discretization): The Euler scheme derives from a simple integration rule. From the definition of the Itô process we have

\[ X(t_{i+1}) = X(t_i) + \int_{t_i}^{t_{i+1}} \mu(t, X(t)) \, dt + \int_{t_i}^{t_{i+1}} \sigma(t, X(t)) \, dW(t). \]

Obviously, the Euler scheme is given by the approximation of the integrals

\[ \int_{t_i}^{t_{i+1}} \mu(t, X(t)) \, dt \approx \int_{t_i}^{t_{i+1}} \mu(t_i, X(t_i)) \, dt = \mu(t_i, X(t_i)) \Delta t_i \]

\[ \int_{t_i}^{t_{i+1}} \sigma(t, X(t)) \, dW(t) \approx \int_{t_i}^{t_{i+1}} \sigma(t_i, X(t_i)) \, dW(t) = \sigma(t_i, X(t_i)) \Delta W(t_i). \]

The following Milstein scheme improves the approximation of the stochastic integral \( \int dW \).

Definition 178 (Milstein Scheme):

Given an Itô process

\[ dX(t) = \mu(t, X(t)) \, dt + \sigma(t, X(t)) \, dW(t), \]

and a time discretization \( \{ t_i \mid i = 0, \ldots, n \} \) with \( 0 = t_0 < \ldots < t_n \), then the time-discrete stochastic process \( \tilde{X} \) defined by

\[ \tilde{X}(t_{i+1}) = \tilde{X}(t_i) + \mu(t_i, \tilde{X}(t_i)) \Delta t_i + \sigma(t_i, \tilde{X}(t_i)) \Delta W(t_i) \]

\[ + \frac{1}{2} \sigma(t_i, \tilde{X}(t_i)) \sigma'(t_i, \tilde{X}(t_i))(\Delta W(t_i)^2 - \Delta t_i) \]
is called a Milstein Scheme of the process $X$ (where $\Delta t_i := t_{i+1} - t_i$ and $\Delta W(t_i) := W(t_{i+1}) - W(t_i)$ and $\sigma' := \frac{\partial}{\partial x}\sigma$).

**Remark 179 (Milstein Scheme):** The Milstein scheme gives an “improvement” only if $\sigma$ depends on $X$.

Let us consider another discretization scheme:

**Definition 180 (Euler Scheme with Predictor-Corrector Step):** Given an Itô process

$$dX(t) = \mu(t, X(t)) \, dt + \sigma(t, X(t)) \, dW(t),$$

and a time discretization $\{t_i \mid i = 0, \ldots, n\}$ with $0 = t_0 < \ldots < t_n$, then the time-discrete stochastic process $\tilde{X}$ defined by

$$\tilde{X}(t_{i+1}) = \tilde{X}(t_i) + \frac{1}{2} (\mu(t_i, \tilde{X}(t_i)) + \mu(t_{i+1}, \tilde{X}(t_{i+1}))) \Delta t_i + \sigma(t_i, \tilde{X}(t_i)) \Delta W(t_i)$$

is called an Euler scheme with predictor-corrector step of the process $X$ (where $\Delta t_i := t_{i+1} - t_i$ and $\Delta W(t_i) := W(t_{i+1}) - W(t_i)$).

**Interpretation (Predictor-Corrector Scheme):** The predictor-corrector scheme improves the integration of the drift term $\int \mu \, dt$, not of the stochastic integral $\int dW$. Instead of approximating the integral $\int_{t_i}^{t_{i+1}} \mu(t, X(t)) \, dt$ by a rectangular rule $\mu(t_i, X(t_i)) \Delta t_i$ the method aims to use a trapezoidal rule. With a trapezoidal rule the integral $\int_{t_i}^{t_{i+1}} \mu(t, X(t)) \, dt$ would be approximated as $\frac{1}{2} (\mu(t_i, X(t_i)) + \mu(t_{i+1}, X(t_{i+1}))) \Delta t_i$. Since the realization $X(t_{i+1})$ and thus $\mu(t_{i+1}, X(t_{i+1}))$ is unknown, it is approximated by an Euler step $\tilde{X}(t_{i+1})$ (predictor step) and the trapezoidal rule is applied with this approximation. This corresponds to correcting $\tilde{X}(t_{i+1})$ (corrector step). We have:

$$\tilde{X}(t_{i+1}) = \tilde{X}(t_{i+1}) - \mu(t_i, \tilde{X}(t_i)) \Delta t_i + \frac{1}{2} (\mu(t_i, \tilde{X}(t_i)) + \mu(t_{i+1}, \tilde{X}(t_{i+1}))) \Delta t_i$$

$$= \tilde{X}(t_{i+1}) + \frac{1}{2} (\mu(t_{i+1}, \tilde{X}(t_{i+1})) - \mu(t_i, \tilde{X}(t_i))) \Delta t_i \quad \text{(13.1)}$$

\[<|\]
Tip (Implementation of the Predictor-Corrector Scheme):

Note that for an implementation formula (13.1) is more efficient than the two Euler steps in the original Definition (180) of the scheme. The second Euler step is replaced by a correction term applied to \( \tilde{X}^* (t_{i+1}) \) and requires only the additional calculation of \( \mu(t_{i+1}, \tilde{X}^* (t_{i+1})) \).

The schemes presented give a time-discrete stochastic process \( \tilde{X} \) such that \( \tilde{X} (t_i) \) is an approximation of \( X(t_i) \). An in-depth discussion of numerical methods of approximating stochastic processes can be found in [21].

### 13.1.2 Time Discretization of a Lognormal Process

Consider the process

\[
\mathrm{d}X = \mu(t, X(t)) X(t) \, \mathrm{d}t + \sigma(t) X(t) \, \mathrm{d}W(t),
\]

where \((t, x) \mapsto \mu(t, x)\) and \(t \mapsto \sigma(t)\) are given deterministic functions. With Lemma 50 we have

\[
\mathrm{d} \log(X) = (\mu(t, X(t)) - \frac{1}{2} \sigma^2(t)) \, \mathrm{d}t + \sigma(t) \, \mathrm{d}W(t).
\]

In the following we discuss several possible time discretizations of the process \( X \). The discussion is of special importance since the Black-Scholes model, the Black model, and the LIBOR market model are all of the form (13.2).

#### 13.1.2.1 Discretization via Euler Scheme

The Euler scheme for the stochastic differential equation (13.2) is given by

\[
\tilde{X}(t_{i+1}) = \tilde{X}(t_i) + \mu(t_i, \tilde{X}(t_i)) \tilde{X}(t_i) \Delta t_i + \sigma(t_i) \tilde{X}(t_i) \Delta W(t_i).
\]

The random variables \( \tilde{X}(t) \) generated by this scheme differ from the random variables \( X(t) \) of the time-continuous process by a discretization error \( X(t) - \tilde{X}(t) \). This discretization error might be relatively large. Take, for example, the even simpler case of a vanishing drift \( \mu = 0 \). Then \( \tilde{X}(t_1) \) is normally distributed, while \( X(t_1) \) is lognormally distributed. Note that \( \tilde{X} \) can attain negative values, while \( X \) cannot (this follows from (13.3)).

#### 13.1.2.2 Discretization via Milstein scheme

One way of reducing the discretization error is to use the Milstein scheme (Definition 178):

\[
\tilde{X}(t_{i+1}) = \tilde{X}(t_i) + (\mu(t_i, \tilde{X}(t_i)) - \frac{1}{2} \sigma(t_i)^2) \tilde{X}(t_i) \Delta t_i + \sigma(t_i) \tilde{X}(t_i) \Delta W(t_i) + \frac{1}{2} \sigma(t_i)^2 \tilde{X}(t_i) \Delta W(t_i)^2.
\]

http://www.christian-fries.de/finmath/book

185
13.1.2.3 Discretization of the Log Process

A much better discretization than the two previous schemes is given by the Euler discretization of the Itô process of log($X$). The Euler scheme of (13.3) is given by

$$\log(\tilde{X}(t_{i+1})) = \log(\tilde{X}(t_i)) + (\mu(t_i, \tilde{X}(t_i)) - \frac{1}{2} \sigma(t_i)^2) \Delta t_i + \sigma(t_i) \Delta W(t_i),$$  \hspace{1cm} (13.4)

and applying the exponential we have

$$\tilde{X}(t_{i+1}) = \tilde{X}(t_i) \exp((\mu(t_i, \tilde{X}(t_i)) - \frac{1}{2} \sigma(t_i)^2) \Delta t_i + \sigma(t_i) \Delta W(t_i)).$$

Using this scheme will give a lognormal random variable $\tilde{X}(t_1)$.

13.1.2.4 Exact Discretization

For the special case where $\mu$ does not depend on $X$, e.g., if $X$ is a relative price under the corresponding martingale measure and thus even drift-free, then we can take the exact solution as a discretization scheme. We then have

$$X(t_{i+1}) = X(t_i) \exp((\mu_i - \frac{1}{2} \sigma_i^2) \Delta t_i + \sigma_i \Delta W(t_i)), \hspace{1cm} (13.5)$$

where

$$\mu_i := \frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} \mu(\tau) \, d\tau, \hspace{1cm} \sigma_i := \sqrt{\frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} \sigma^2(\tau) \, d\tau}.$$ 

13.2 Discretization of Paths (Monte Carlo Simulation)

Consider the time-discrete stochastic process

$$X(t_{i+1}) = X(t_i) + \mu(t_i, X(t_i)) \Delta t_i + \sigma(t_i, X(t_i)) \Delta W(t_i). \hspace{1cm} (13.6)$$

This is an Euler scheme. The considerations below apply to any other discretization scheme. Furthermore, we do not apply a tilde to the process $X$ since we are only considering the time-discrete process, and so do not have to distinguish it from the original time-continuous process.
13.2.1 Monte Carlo Simulation

The random variables $\Delta W(t_i)$ of the respective time steps are mutually independent; see Definition 29. At every time step $t_i$ a random number is drawn according to the distribution of $\Delta W(t_i)$, (i.e., a vector of random numbers if $\Delta W(t_i)$ is vector valued), which we denote by $\Delta W(t_i, \omega_j)$. Then

$$X(t_{i+1}, \omega_j) = X(t_i, \omega_j) + \mu(t_i, X(t_i, \omega_j)) \Delta t_i + \sigma(t_i, X(t_i, \omega_j)) \Delta W(t_i, \omega_j)$$

determines the process $X$ on a path, which we denote by $\omega_j$. Here $\Delta W(t_i, \omega_j)$ and $\Delta W(t_k, \omega_j)$ ($i \neq k$) are independent random numbers, following the definition of the Brownian motion. If we follow this rule to generate paths $\omega_1, \ldots, \omega_{n_{\text{paths}}}$, where $\Delta W(t_i, \omega_j)$ and $\Delta W(t_i, \omega_k)$ ($j \neq k$) are independent, then we say that the set

$$\{X(t_i, \omega_k) \mid i = 0, 1, \ldots, n_{\text{times}}; k = 0, 1, \ldots, n_{\text{paths}}\}$$

is a Monte Carlo simulation of the process $X$.

![Figure 13.2. Monte-Carlo Simulation](http://www.christian-fries.de/finmath/book)

An approximation of the expectation of some function $f$ of the $X(t_i)$’s is then given by

$$E^P(f(X(t_0), \ldots, X(t_{n_{\text{times}}}))) \mid F_{t_0} \approx \frac{1}{n_{\text{paths}}} \sum_{j=1}^{n_{\text{paths}}} f(X(t_0, \omega_j), \ldots, X(t_{n_{\text{times}}}, \omega_j)).$$

The generation of random numbers is discussed in Section B.1.

13.2.2 Weighted Monte Carlo Simulation

A generalization of the procedure is to generate the random numbers $\Delta W(t_i, \omega_j)$ not according to the distribution $\Delta W(t_i)$, which means that all paths $\omega_j$ are generated with
nonuniform weights \( p_j (\sum_{j=1}^{n_{\text{paths}}} p_j = 1) \). In this case we call the simulation weighted Monte Carlo simulation. For the expectation we have

\[
E^\mathbb{P}(f(X(t_0), \ldots, X(t_{n_{\text{times}}})) \mid \mathcal{F}_{t_0}) \approx \sum_{j=1}^{n_{\text{paths}}} p_j f(X(t_0, \omega_j), \ldots, X(t_{n_{\text{times}}}, \omega_j)).
\]

To summarize, the Monte Carlo simulation consists of the time-discrete process \( X \) in (13.6), represented over a discrete probability space \( (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) \), where

\[
\tilde{\Omega} = \{\omega_1, \ldots, \omega_{n_{\text{paths}}} \} \subset \Omega, \quad \tilde{\mathcal{F}} = \sigma(\{\omega_j\}|j = 1, \ldots, n_{\text{paths}}), \quad \tilde{\mathbb{P}}(\{\omega_j\}) = p_j.
\]

### 13.2.3 Implementation

Figures 13.3 and 13.4 show an example for an object-oriented design. The figures follow the Unified Modeling Language (UML) 1.3; see [25].

The generation of a Monte Carlo simulation of a lognormal process is realized through the abstract base class LogNORMALPROCESS. The class defines abstract methods for initial conditions, drift, and volatility. A specific model has to be derived from this class and implement the three methods. The abstract base class LogNORMALPROCESS provides the implementation of the discretization scheme, using the methods for initial conditions, drift, and volatility.

The calculation of the Brownian increments, i.e., the random numbers, is given by an additional class: BROWNIANMOTION.

#### Figure 13.3. UML Diagram: Monte Carlo simulation/lognormal process.

### 13.2.3.1 Example: Valuation of a Stock Option under the Black-Scholes Model Using Monte Carlo Simulation

Consider the model from Chapter 4, the Black-Scholes model: We have to simulate the process

\[
\frac{dS(t)}{S(t)} = r(t) dS(t) \, dt + \sigma(t) S(t) \, dW^E(t) \quad \text{under the measure } \mathbb{Q}^N, \quad S(0) = S_0
\]
together with the numéraire
\[ dN(t) = r(t)N(t) \, dt, \]
which is not stochastic here. For this example we have to set \( X = (X_1, X_2) = (S, N) \) in the previous section. We choose \( r \) and \( \sigma \) constant and apply the Euler scheme to \( \log(S) \), following Section 13.1.2.3:
\[
S(t_{i+1}) = S(t_i) \exp((r - \frac{1}{2}\sigma^2) \Delta t_i + \sigma \Delta W(t_i)), \quad S(0) = S_0,
\]
\[
N(t_{i+1}) = N(t_i) \exp(r \Delta t_i), \quad N(0) = 1.
\]
In this example the time discretization does not introduce an approximation error, because we are in the special situation of Section 13.1.2.4. If \( \omega_1, \ldots, \omega_{\text{paths}} \) are paths of a Monte Carlo simulation, then we have for the price \( V \) of a European option with maturity \( t_k \) and strike \( K \)
\[
V(0) \approx N(0) \sum_{j=1}^{\text{paths}} \frac{1}{\text{paths}} \max(S(t_k, \omega_j) - K, 0) \frac{N(t_k)}{N(0)}.
\]
We can extend the object-oriented design from Figure 13.3 to derive the class \texttt{BlackScholesModel} from the abstract base class \texttt{LogNormalProcess}. The class \texttt{BlackScholesModel} implements the methods providing the initial value (returning \( S(0) \)), the drift (returning \( r \)), and the factor loading (returning \( \sigma \)). In this context the factor loading is identical to the volatility.\(^2\) In addition the class implements a method that returns the corresponding numéraire.

\subsection*{13.2.3.2 Separation of Product and Model}

The evaluation of a derivative product, in our case a simple European option, is realized in its own class \texttt{StockOption}. This class does not communicate directly with the \texttt{BlackScholesModel}. Instead it expects an \texttt{interface} \texttt{MonteCarloStockProcessModel} and the model implements this \texttt{interface}.

The \texttt{interface} \texttt{MonteCarloStockProcessModel} means that the stock model makes the stock process and the numéraire available to the stock product as a Monte Carlo simulation. All corresponding Monte Carlo evaluations of stock products expect this interface only. All corresponding Monte Carlo stock models implement this interface. This produces a separation of product and model. The model used to evaluate the products may be exchanged for another, as long as the interface is respected.

We will use this principle in the object-oriented design of the LIBOR market model, a multidimensional interest rate model. There we will reuse the classes \texttt{BrownianMotion} and \texttt{LogNormalProcess}; see Section 19.6.

\(^2\) In a multi factor model the factor loading is given by the square root of the covariance matrix.
Figure 13.4. **UML Diagram: Evaluation under a Black-Scholes model via Monte Carlo simulation.**
Sample 4

The pure and simple truth
is rarely pure and never simple.

Oscar Wilde
The Importance of Being Earnest [40].

We assume a time discretization (tenor structure)

\[ 0 = T_0 < T_1 < \cdots < T_n. \]

We model the forward rates \( L_i := L(T_i, T_{i+1}) \) for \( i = 0, \ldots, n - 1 \); see Definition 99. This represents a discretization of the interest rate curve, where the continuum of maturities has been discretized.\(^1\)

The LIBOR market model assumes a lognormal dynamic for LIBORs \( L_i := L(T_i, T_{i+1}) \), i.e.,\(^2\)

\[
\frac{dL_i(t)}{L_i(t)} = \mu^\mathbb{P}_i(t) \, dt + \sigma^\mathbb{P}_i(t) \, dW^\mathbb{P}_i(t) \quad \text{for } i = 0, \ldots, n - 1, \text{ under } \mathbb{P}, \tag{19.1}
\]

with initial conditions

\[ L_i(0) = L_{i,0}, \quad \text{with } L_{i,0} \in [0, \infty), i = 0, \ldots, n - 1, \]

where \( W^\mathbb{P}_i \) denote (possibly instantaneously correlated) \( \mathbb{P} \)-Brownian motions with

\[
dW^\mathbb{P}_i(t) \, dW^\mathbb{P}_j(t) = \rho_{i,j}(t) \, dt.
\]

Let \( \sigma^\mathbb{P} : [0, T] \to \mathbb{R} \) and \( \rho_{i,j} : [0, T] \to \mathbb{R} \) be deterministic functions and \( \mu^\mathbb{P}_i \) the drift as \( \mathcal{F}_t \)-adapted process. By \( R(t) := (\rho_{i,j}(t))_{i,j=0,\ldots,n-1} \) we denote the correlation matrix.

\(^1\) In practice it is normal to model semiannual or quarterly rates \( T_{i+1} - T_i = 0.25 \) and to consider these up to a maturity of 20 or 30 years, giving 80 or 120 interest rates to model.

\(^2\) We denote the simulation time parameter of the stochastic process by \( t \).
Motivation: Equation (19.1) is a lognormal model for the forward rates $L_i$. If we consider only a single equation, i.e., fix $i \in \{1, \ldots, n-1\}$, it represents the Black model considered in Chapter 10: Equation (19.1) is identical with Equation (10.1). If we change the measure such that $L_i$ is drift-free (see Chapter 10), we see that the terminal distribution of $L_i$ is lognormal. Thus, the LIBOR market model is equivalent to the consideration of $n$ Black models under a unified measure.

As was discussed in Chapter 10, to evaluate a caplet under this model it is not relevant that $\sigma_i$ is time dependent (we have assumed time dependency of $\sigma_i$ in Chapter 10 for didactical reasons). However, for the value of complex derivatives the time dependency matters. A further degree of freedom introduced in (19.1) is the instantaneous correlation $\rho_{i,j}$ of the driving Brownian motions. For the value of a caplet the instantaneous correlation is insignificant (indeed, it does not enter in the Black model). For the evaluation of swaptions the correlation of the forward rates is significant.

For further generalizations of the model, consider nondeterministic $\sigma_i$, i.e., stochastic volatility models. In this case the terminal LIBOR distributions no longer correspond to the Black model ones, which is, of course, intended. Equation (19.1) is to be seen as a starting point of a whole model family. The model (19.1) has been chosen as the starting point, because (historically) the lognormal (Black) model is well understood, especially by traders.\(^3\)

Remark 212 (Interest Rate Structure): Equation (19.1) models the evolution of the LIBOR $L(T_i, T_{i+1})$. Without further interpolation assumption, these are the shortest forward rates that can be considered in our time discretization (tenor structure). The equation system (19.1) thus determines the evolution of all bond prices with maturities $T_i$ and all forward rates for the periods $[T_i, T_k]$, since

$$ 1 + L(T_i, T_k)(T_k - T_i) = \frac{P(T_i)}{P(T_k)} = \prod_{j=i}^{k-1} \frac{P(T_j)}{P(T_{j+1})} = \prod_{j=i}^{k-1} (1 + L(T_j, T_{j+1})(T_{j+1} - T_j)). $$

To shorten notation we write $\delta_i := T_{i+1} - T_i$, $i = 0, \ldots, n-1$ for the period length.

\(^3\) Caplet prices are quoted by traders by the implied Black volatility. This is, of course, just another unit of the price, since the Black model is a one-to-one map from price to implied volatility.
19.1 Derivation of the Drift Term

As in Chapters 10 and 11, our first step is to choose some numéraire \( N \) and derive the drift under a martingale measure \( Q^N \). If the processes have been derived under the martingale measure \( Q^N \), then the (discretized) interest rate curve may be simulated numerically and a derivative \( V \) may be priced through \( V(0) = N(0)E^Q_N (V | \mathcal{F}_T) \) (see Chapter 13).

We fix a numéraire \( N \). Let the assumptions of Theorem 74 hold such that there exists a corresponding equivalent martingale measure \( Q^N \) such that \( N \)-relative prices are martingales. From Theorem 59 under \( Q^N \) the process (19.1) has a changed drift, namely

\[
\frac{dL_i(t)}{L_i(t)} = \mu^{Q^N}_i (t) \, dt + \sigma_i(t) \, dW^{Q^N}_i (t) \quad \text{for } i = 0, \ldots, n - 1. \tag{19.2}
\]

19.1.1 Derivation of the Drift Term under the Terminal Measure

We fix the \( T_n \)-bond \( N(t) = P(T_n; t) \) as numéraire. From Theorem 59 under \( Q^{P(T_n)} \) the process (19.1) has a changed drift:

\[
\frac{dL_i(t)}{L_i(t)} = \mu^{Q^{P(T_n)}}_i (t) \, dt + \sigma_i(t) \, dW^{Q^{P(T_n)}}_i (t) \quad \text{for } i = 0, \ldots, n - 1. \tag{19.3}
\]

We need to determine \( \mu^{Q^{P(T_n)}}_i \). The martingale measure \( Q^{P(T_n)} \) corresponding to \( N(t) = P(T_n; t) \) is also called terminal measure (since \( T_n \) is the time horizon of our time discretization).

As in Chapter 10, we will construct relative prices with respect to \( P(T_n) \) and obtain equations from which we will derive the drifts \( \mu_i \). From Definition 99

\[
\prod_{k=i}^{n-1} \left( 1 + \delta_k L_k \right) = \prod_{k=i}^{n-1} \frac{P(T_k)}{P(T_{k+1})} = \frac{P(T_i)}{P(T_n)} \quad \text{for } i = 0, \ldots, n - 1. \tag{19.4}
\]

Since we have a \( P(T_n) \)-relative price of a traded product on the right-hand side in (19.4), we have for the drifts:

\[
\text{Drift}_{Q^{P(T_n)}} \left[ \prod_{k=i}^{n-1} (1 + \delta_k L_k) \right] = 0, \quad i = 0, \ldots, n - 1.
\]
We apply Theorem 48 and obtain ∀ \( i = 0, \ldots, n - 1 \)

\[
\begin{align*}
\prod_{k=i}^{n-1} (1 + \delta_k L_k) &= \sum_{j=i}^{n-1} \prod_{k=i}^{n-1} (1 + \delta_k L_k) \delta_j dL_j + \sum_{j=0}^{n-1} \prod_{k=i}^{n-1} (1 + \delta_k L_k) \delta_j dL_j \delta_l dL_l \\
&= \prod_{k=i}^{n-1} (1 + \delta_k L_k) \left( \sum_{j=i}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} + \sum_{j=0}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} \frac{\delta_l dL_l}{1 + \delta_l L_l} \right) \\

&= \prod_{k=i}^{n-1} (1 + \delta_k L_k) \sum_{j=i}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} + \sum_{l\geq j+1}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} \frac{\delta_l dL_l}{1 + \delta_l L_l} \\

&= \prod_{k=i}^{n-1} (1 + \delta_k L_k) \sum_{j=i}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} + \sum_{l\geq j+1}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} \frac{\delta_l dL_l}{1 + \delta_l L_l}.
\end{align*}
\]

Since ∀ \( i = 0, \ldots, n - 1 \)

\[
\text{Drift}_{Q^{PT_n}} \left[ \prod_{k=i}^{n-1} (1 + \delta_k L_k) \right] = 0
\]

(19.5)

it follows that ∀ \( i = 0, \ldots, n - 1 \)

\[
\begin{align*}
\sum_{j=i}^{n-1} \text{Drift}_{Q^{PT_n}} \left[ \frac{\delta_j dL_j}{1 + \delta_j L_j} + \sum_{l\geq j+1}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} \frac{\delta_l dL_l}{1 + \delta_l L_l} \right] &= 0 \\
\end{align*}
\]

(19.6)

and thus ∀ \( j = 0, \ldots, n - 1 \)

\[
\begin{align*}
\text{Drift}_{Q^{PT_n}} \left[ \frac{\delta_j dL_j}{1 + \delta_j L_j} + \sum_{l\geq j+1}^{n-1} \frac{\delta_j dL_j}{1 + \delta_j L_j} \frac{\delta_l dL_l}{1 + \delta_l L_l} \right] &= 0.
\end{align*}
\]

(19.7)

If we now use

\[
dL_j = L_j \mu_j^{Q^{PT_n}} dt + L_j \sigma_j^{Q^{PT_n}} dW_j^{Q^{PT_n}}
\]

and

\[
dL_j dL_l = L_j L_l \sigma_j^{Q^{PT_n}} \sigma_l^{Q^{PT_n}} dt
\]

in (19.7), then we have

\[
\mu_j^{Q^{PT_n}} \frac{\delta_j L_j}{1 + \delta_j L_j} + \sum_{l\geq j+1}^{n-1} \frac{\delta_j L_j}{1 + \delta_j L_j} \frac{\delta_l L_l}{1 + \delta_l L_l} \sigma_j^{Q^{PT_n}} \sigma_l^{Q^{PT_n}} = 0.
\]
i.e.,

\[ \mu_{j}^{Q_{Tn}}(t) = - \sum_{l \geq j+1}^{\beta_{l}L_{l}(t)} \frac{\delta_{l}}{1 + \delta_{l}L_{l}(t)} \sigma_{j}(t)\sigma_{l}(t)\rho_{j,l}(t). \]  

(19.8)

The procedure above may be summarized as follows: To derive the \( n \) drifts we write down \( n \) independent traded assets as a function of the model quantities. By considering the drifts of their relative prices, we obtain \( n \) equations for the drifts of the modeled quantities.

### 19.1.2 Derivation of the Drift Term under the Spot LIBOR Measure

We fix the *rolled over one period bond* as numéraire, i.e., the investment of 1 at time \( T_{0} \) into the \( T_{1} \)-bond and after its maturity the reinvestment of the proceeds into the bond of the next period, i.e., in \( T_{j} \) the reinvestment in the \( T_{j+1} \)-bond. It is

\[
N(t) := P(T_{m(t)+1}^{T_{n}}; t) \prod_{j=1}^{m(t)+1} \left( \frac{1}{P(T_{j}^{T_{j-1}}; t)} \right) = P(T_{m(t)+1}^{T_{n}}; t) \prod_{j=0}^{m(t)} (1 + L_{j}(T_{j}) \delta_{j}), \tag{19.9}
\]

\[
= \frac{P(T_{j-1}^{T_{j}}; t)}{P(T_{j}^{T_{j-1}}; t)} = (1 + \delta_{j-1}(T_{j-1}) \delta_{j})
\]

where \( m(t) := \max[i : T_{i} \leq t] \) and \( \delta_{j} := T_{j+1} - T_{j} \). The corresponding equivalent martingale measure \( Q^{N} \) is called the *spot measure*.

As before, we consider the processes of \( N \)-relative prices of traded products (from which we know that they have drift 0 under \( Q^{N} \)). We consider the \( N \)-relative prices of the bonds \( P(T_{i}) \). It is

\[
\frac{P(T_{i}^{T_{n}}; t)}{N(t)} = \frac{P(T_{i}^{T_{n}}; t)}{P(T_{m(t)+1}^{T_{n}}; t)} \prod_{j=0}^{m(t)} (1 + L_{j}(T_{j}) \delta_{j})^{-1}
\]

\[
= \prod_{j=m(t)+1}^{i-1} (1 + L_{j}(t) \delta_{j})^{-1} \prod_{j=0}^{m(t)} (1 + L_{j}(T_{j}) \delta_{j})^{-1}, \tag{19.10}
\]

thus

\[
\text{Drift}_{Q^{N}} \left[ \prod_{k=m(t)+1}^{i-1} (1 + L_{k}\delta_{k})^{-1} \right] = 0. \tag{19.11}
\]
Since
\[
d\left( \prod_{j=m(t)+1}^{i-1} (1 + L_j(t)\delta_j)^{-1} \prod_{j=0}^{m(t)} (1 + L_j(T_j)\delta_j)^{-1} \right) = d\left( \prod_{j=m(t)+1}^{i-1} (1 + L_j(t)\delta_j)^{-1} \prod_{j=0}^{m(t)} (1 + L_j(T_j)\delta_j)^{-1} \right),
\]
we consider
\[
d\left( \prod_{k=m(t)+1}^{i-1} \frac{1}{1 + \delta_k L_k} \right)
= \sum_{j=m(t)+1}^{i-1} \prod_{k=m(t)+1}^{i-1} \frac{1}{1 + \delta_k L_k} \left( \frac{-\delta_j dL_j}{(1 + \delta_j L_j)^2} + \frac{\delta_j^2 dL_j dL_j}{(1 + \delta_j L_j)^3} \right)
+ \sum_{j=m(t)+1}^{i-1} \prod_{k=m(t)+1}^{i-1} \frac{1}{1 + \delta_k L_k} \left( \frac{\delta_j dL_j}{(1 + \delta_j L_j)^2} + \frac{\delta_j^2 dL_j dL_j}{(1 + \delta_j L_j)^3} \right) \left( \frac{\delta_i dL_i}{(1 + \delta_i L_i)^2} + \frac{\delta_i^2 dL_i dL_i}{(1 + \delta_i L_i)^3} \right).
\]

With (19.11) we have \(\forall i = 0, \ldots, n - 1\)
\[
\sum_{j=m(t)+1}^{i-1} \text{Drift}_{\mathbb{Q}^N} \left[ -\frac{\delta_j dL_j}{(1 + \delta_j L_j)} + \sum_{l=m(t)+1}^{j} \frac{\delta_j dL_j}{(1 + \delta_j L_j)} \frac{\delta_i dL_i}{(1 + \delta_i L_i)} \right] = 0
\]
and thus \(\forall j = 0, \ldots, n - 1\)
\[
\text{Drift}_{\mathbb{Q}^N} \left[ -\frac{\delta_j dL_j}{(1 + \delta_j L_j)} + \sum_{l=m(t)+1}^{j} \frac{\delta_j dL_j}{(1 + \delta_j L_j)} \frac{\delta_i dL_i}{(1 + \delta_i L_i)} \right] = 0. \tag{19.12}
\]

If we now use
\[
dL_j = L_j \mu_j^{\mathbb{Q}_t} \, dt + L_j \sigma_j \, dW_j \quad \text{and} \quad dL_j \, dL_l = L_j L_l \sigma_j \sigma_l \rho_{j,l} \, dt
\]

http://www.christian-fries.de/finmath/book

302
in (19.7), then we have
\[
-\mu^Q_N \frac{\delta_j L_j}{1 + \delta_j L_j} + \sum_{l=m(t)+1}^j \frac{\delta_l L_l}{1 + \delta_l L_l} \frac{\delta_l L_l}{1 + \delta_l L_l} \sigma_j \rho_{j,l} = 0,
\]
i.e.,
\[
\mu^Q_N(t) = \sum_{l=m(t)+1}^j \frac{\delta_l L_l(t)}{1 + \delta_l L_l(t)} \sigma_j(t) \rho_{j,l(t)}.
\]

### 19.1.3 Derivation of the Drift Term under the $T_k$-Forward Measure

**Exercise:** (Drift under the $T_k$-forward measure) Consider

\[
N(t) := \begin{cases} 
P(T_k; t), & t \leq T_k \\
P(T_{m(t)+1}; t) \prod_{j=k+1}^{m(t)+1} \frac{1}{P(T_j; T_{j-1})}, & t > T_k, 
\end{cases}
\]

where $m(t) := \max\{i : T_i \leq t\}$.

1. Give an interpretation of $N(t)$ as traded product.
2. Derive the drift of the model (19.1) under the $Q^N$ measure with the numéraire $N$.

**Solution:**

\[
\begin{align*}
\mu_j(t) &= -\sum_{l=j+1}^{k-1} \frac{\delta_l L_l}{1 + \delta_l L_l} \sigma_j \rho_{j,l} & \text{for } j < k - 1 \text{ and } t \leq T_k \\
\mu_j(t) &= 0 & \text{for } j = k - 1 \text{ and } t \leq T_k \\
\mu_j(t) &= \sum_{l=k}^{j} \frac{\delta_l L_l}{1 + \delta_l L_l} \sigma_j \rho_{j,l} & \text{for } j \geq k \text{ and } t \leq T_k \\
\mu_j(t) &= \sum_{l=m(t)+1}^{j} \frac{\delta_l L_l}{1 + \delta_l L_l} \sigma_j \rho_{j,l} & \text{for } t > T_k.
\end{align*}
\]

---

\[\text{Since the coefficient of } dt \text{ equals 0.}\]
19.2 The Short Period Bond $P(T_{m(t)+1}; t)$

For $t \notin \{T_1, \ldots, T_n\}$ neither the numéraire $N(t)$ of the terminal measure nor the numéraire of the spot measure is fully described by the processes $L_i(t)$. The unspecified bond $P(T_{m(t)+1}; t)$ occurs in both numéraires. We will now discuss the relevance of $P(T_{m(t)+1}; t)$.

19.2.1 Role of the Short Bond in a LIBOR Market Model

For the modeling of the forward rates $L_i(t) := L(T_i, T_{i+1}; t)$ on the tenor periods $[T_i, T_{i+1}], i = 0, \ldots, n$ the specification of $P(m(t) + 1; t)$ is irrelevant. For the derivation of the corresponding drift terms it was not relevant to specify the stochastic of $P(T_{m(t)+1}; t)$, since the term canceled for the relative prices considered.

Conversely, the LIBOR market model does not describe the stochastic of the short bond $P(T_{m(t)+1}; t)$, since it is not given as a function of the processes $L_i(t)$.

19.2.2 Link to Continuous Time Tenors

The specification of the short bond $P(T_{m(t)+1}; t)$ becomes relevant if the model has to describe interest rates of interest rate periods which are not part of the tenor structure. The specification of $P(m(t) + 1; t)$ will determine how the fractional forward rates $L(T_s, T_e; t)$ with $T_s \notin \{T_1, \ldots, T_n\}$ and/or $T_e \notin \{T_1, \ldots, T_n\}$ will evolve (see Section 19.5). It is the link from a model with discrete tenors (LIBOR market model) to a model with continuous time tenors (Heath-Jarrow-Morton framework). In the special case where $P(m(t) + 1; t)$ has zero volatility, the LIBOR market model under spot measure coincides with a Heath-Jarrow-Morton framework with a special volatility structure under the risk-neutral measure (see Section 23.2).

19.2.3 Drift of the Short Bond in a LIBOR Market Model

Within the LIBOR market model there is no constraint on the drift of $P(m(t) + 1; t)$, because in $\frac{P(m(t)+1; t)}{N(t)}$ the term cancels out. The relative price $\frac{P(m(t)+1; t)}{N(t)}$ is always a martingale for any choice of $P(m(t) + 1; t)$. This might come as a surprise, but we have already encountered this behavior: In the Black-Scholes model the drift $r$ of $B(t)$ is a free parameter, because it is the drift of the numéraire. The parameter $r$ is determined by calibration to a market interest rate. In a short rate model the drift is a free parameter. It is determined by calibration to the market interest rate curve; see Chapter 22. Here, similarly, $P(m(t) + 1; t)$ determines the interpolation of the initial interest rate curve.
The trivial fact that the numéraire-relative price of the numéraire, i.e., \( \frac{N(t)}{N(t)} \), is always a martingale plays a role in Markov functional models. There, the numéraire is postulated to be a functional of some Markov process.

## 19.3 Discretization and (Monte Carlo) Simulation

In this section we will discuss the discretization and implementation of the model. Let us therefore assume that the free parameters \( \sigma_i, \rho_{i,j}, \) and \( L_{i,0} \) \((i, j = 1, \ldots, n)\) are given. Together with the drift formula obtained in the previous section the model is fully specified. Section 19.4 will then discuss how the parameters \( L_{i,0}, \sigma_i, \rho_{i,j} \) are obtained.

### 19.3.1 Generation of the (Time-Discrete) Forward Rate Process

As discussed in Chapter 13, we choose the Euler discretization of the Itô process of \( \log(L_i) \). From Lemma 50 we have

\[
d(\log(L_i(t))) = (\mu_i^{\mathbb{Q}^N}(t) - \frac{1}{2} \sigma_i^2(t)) \, dt + \sigma_i(t) \, dW_i^{\mathbb{Q}^N}(t) \tag{19.13}
\]

and the corresponding Euler scheme of (19.13) is

\[
\log(\tilde{L}_i(t + \Delta t)) = \log(\tilde{L}_i(t)) + (\mu_i(t) - \frac{1}{2} \sigma_i^2(t)) \Delta t + \sigma_i(t) \Delta W_i(t). \tag{19.14}
\]

Applying the exponential gives us the discretization scheme of \( L_i \) as

\[
\tilde{L}_i(t + \Delta t) = \tilde{L}_i(t) \exp\left( (\mu_i(t) - \frac{1}{2} \sigma_i^2(t)) \Delta t + \sigma_i(t) \Delta W_i(t) \right). \tag{19.15}
\]

In the special case that the process \( L_i \) is considered under the measure \( \mathbb{Q}^{P(T_{i+1})} \), i.e., \( \mu_i^{\mathbb{Q}^{P(T_{i+1})}}(t) = 0 \), and that the given \( \sigma_i(t) \) is a known deterministic function, we may use the exact solution for a discretization scheme:

\[
L_i(t + \Delta t) = L_i(t) \exp\left( -\frac{1}{2} \tilde{\sigma}_i^2(t, t + \Delta t) \Delta t + \tilde{\sigma}_i(t, t + \Delta t) \Delta W_i \right),
\]

where

\[
\tilde{\sigma}_i(t, t + \Delta t) := \sqrt{\frac{1}{\Delta t} \int_t^{t+\Delta t} \sigma_i^2(\tau) \, d\tau}.
\]
In the case where $L_i$ is not drift-free, we choose instead of (19.15) the discretization scheme

$$L_i(t + \Delta t) = L_i(t) \exp \left( (\mu_i(t) - \frac{1}{2} \tilde{\sigma}_i(t,t + \Delta t)^2) \Delta t + \tilde{\sigma}_i(t,t + \Delta t) \Delta W_i(t) \right)$$

(19.16)

(we write $L$ in place of $\tilde{L}$, although (19.16) is an approximation of (19.1)). The diffusion $dW$ is discretized by exact solution; the drift $dt$ is discretized by an Euler scheme. The discretization error of this scheme stems from the discretization of the stochastic drift $\mu_i$ only. This discretization error results in a violation of the no-arbitrage requirement of the model (the discretized model does not have the correct, arbitrage-free drift). Methods which do not exhibit an arbitrage due to a discretization error are called arbitrage-free discretization; see [74]).

The volatility functions $\sigma_i$ are usually assumed to be piecewise constant functions on $[T_j, T_{j+1})$, such that $\tilde{\sigma}_i(t, t + \Delta t)$ may be calculated analytically. It is $\tilde{\sigma}_i(t, t + \Delta t) = \sigma_i(t)$.

19.3.2 Generation of the Sample Paths

Equipped with the time discretization (19.16), realizations of the process are calculated for a given number of paths $\omega_1, \omega_2, \omega_3, \ldots$. To do so, normally distributed random numbers $\Delta W_i(t_j)(\omega_k)$, correlated according to $R = (\rho_{ij})$, are generated (see Appendices B.1 and B.2). These are used in the scheme (19.16). The result is a three-dimensional tensor $L_i(t_j, \omega_k)$ parametrized by

- $i$: Index of the interest rate period (tenor structure),
- $j$: Index of the simulation time,
- $k$: Index of the simulation path.

19.3.3 Generation of the Numéraire

Given a simulated interest rate curve $L_i(t_j, \omega_k)$, we can calculate the numéraire. Of course, we have to use the numéraire that was chosen for the martingale measure under which the process was simulated (form of the drift in (19.13)). For the terminal measure we would calculate

$$N(T_i, \omega_k) = \prod_{j=i}^{n-1} (1 + L_j(T_i, \omega_k) (T_{j+1} - T_j))^{-1}.$$
19.4 Calibration—Choice of the Free Parameters

We are now going to explain how the free parameters of the model can be chosen. The free parameters are

- the initial conditions \( L_{i,0} \), \( i = 0, \ldots, n - 1 \),
- the volatility functions or volatility processes \( \sigma_i \), \( i = 1, \ldots, n - 1 \),
- the (instantaneous) correlation \( \rho_{i,j} \), \( i, j = 1, \ldots, n - 1 \).

The determination of the free parameters is also called *calibration of the model*.

**Motivation (Reproduction of Market Prices versus Historical Estimation):** With the LIBOR market model we have a high-dimensional model framework. The main task is the derivation or estimation of the huge amount of free parameters. Two approaches are possible:

- **Reproduction of Market Prices**: The parameters are chosen such that the model reproduces given market prices.

- **Historical Estimation**: The parameters are estimated from historical data, e.g., time series of interest rate fixings.

It may be surprising at first, but the second approach is not meaningful, being in the context of risk-neutral evaluation. The model is considered under the martingale measure \( Q^N \) and its aim is the evaluation and *hedging* (!) of derivatives. An expectation of the numéraire-relative value under the martingale measure corresponds to the numéraire-relative value of the replication portfolio. This replication portfolio has to be set up from traded products, traded at current (!) market prices. If the model did not replicate current market prices, then it would not be possible to buy the replication portfolio of a derivative at the model price of the derivative. The model price would inevitably be wrong.

This remark applies to all free model parameters. In practice, however, it may be difficult or impossible to derive all parameters from market prices. This could be

---

Note: The numéraire is given only at the tenor times \( t = T_i \), since for \( t \neq T_i \) we did not define the short period bond \( P(T_{m(t)+1}; t) \). An interpolation is possible; see Section 19.5 and [98].

---

\(^5\) See Section 19.2.

\(^6\) The parameters \( \sigma_i \) may well be stochastic processes. In this case \( \sigma_i \) is called a stochastic volatility model.
because for a specific product no reliable price is known (low liquidity). It could also be that a corresponding product does not exist. This is often the case for correlation-sensitive products from which we would like to derive the correlation parameters. If a parameter cannot be derived from a market price, a historical estimate becomes an option. If in such a case complete hedge is not possible, the residual risk has to be considered, e.g., by a conservative estimate of the parameter.

For the LIBOR market model a parameter reduction is usually applied first, based on historical estimates of rough market assessment. An example of such parameter reduction is the assumption of a family of functional forms for the volatility \( \sigma_i(t) \) or the correlation \( \rho_{i,j}(t) \). The remaining degrees of freedom are then derived from market prices.

### 19.4.1 Choice of the Initial Conditions

#### 19.4.1.1 Reproduction of Bond Market Prices

Let \( P_{\text{Market}}(T_i) \in (0, 1] \) denote a market observed (i.e., given) price of a \( T_i \)-bond. If we set

\[
L_{i,0} := \frac{P_{\text{Market}}(T_i) - P_{\text{Market}}(T_{i+1})}{P_{\text{Market}}(T_{i+1})(T_{i+1} - T_i)},
\]

then the model reproduces the given market prices of the bonds \( P_{\text{Market}} \). This is ensured by the model having the “right” drift and it is independent of the other parameters.

### 19.4.2 Choice of the Volatilities

#### 19.4.2.1 Reproduction of Caplet Market Prices

We assume here that the \( \sigma_i \)'s are deterministic functions (i.e., not random variables or stochastic processes). The forward rate \( L_i \) follows the Itô process

\[
dL_i(t) = \mu_i^Q(t)L_i(t) \, dt + \sigma_i(t)L_i(t) \, dW_i^Q(t) \quad \text{under } Q := Q^N.
\]

Thus the model corresponds to the Black model discussed in Chapter 10. Under \( Q^{P(T_{i+1})} \) we have \( \mu_{\text{Black,Model}}^Q(T_{i+1}) = 0 \), the distribution of \( L_i(T_i) \) is lognormal, and there exists an analytic evaluation formula for caplets. The only model parameters that enter the caplet price are \( L_0(T_i) \) and

\[
\sigma_i^{\text{Black,Model}} := \left( \frac{1}{T_i} \int_0^{T_i} \sigma_i^2(t) \, dt \right)^{1/2}.
\] (19.17)
If the market price \( V_{\text{Market, Caplet}, i} \) of a caplet on the forward rate \( L_i(T_i) \) is given, then the corresponding implied Black volatility \( \sigma_{i, \text{Black, Market}} \) may be calculated by inverting\(^7\) Equation (10.2). If then \( \sigma_i(t) \) is chosen such that

\[
\sigma_i^{\text{Black, Model}} = \sigma_i^{\text{Black, Market}},
\]

then the model reproduces the given caplet price \( V_{\text{Market, Caplet}, i} \). A possible trivial choice is, e.g., \( \sigma_i(t) = \sigma_i^{\text{Black, Market}} \forall t \).

**Remark 213 (Caplet Smile Modeling):** The fact that the LIBOR market model calibrates to the cap market by a simple boundary condition is one reason for its initial popularity. However, since the model restricted to a single LIBOR is a Black model, the implied volatility does not depend on the strike of an option. Thus, in this form, the model may calibrate to a single caplet per maturity only. It cannot render a caplet smile yet.

To remove this restriction one can extend the model by a local or stochastic volatility or jump-diffusion processes [8]. For an overview on smile modeling in the LIBOR market model see [23].

### 19.4.2.2 Reproduction of Swaption Market Prices

If the correlation \( R = (\rho_{i,j}) \) is given and fixed, then we influence swaption prices through the time structure of the volatility function \( t \mapsto \sigma_i \). We consider swaptions that correspond to our tenor structure, i.e., option on the swap rates:

\[
S(T_i, \ldots, T_j; T_i), \quad 0 < i < j \leq n.
\]

From the definition of the swap and swap rate it is obvious that the price of a corresponding swaption with *exercise date* on or before \( T_i \) and periods \([T_i, T_{i+1}], \ldots, [T_{j-1}, T_j]\) depends only on the behavior of the forward rates \( L_i(t), \ldots, L_{j-1}(t) \) until the *fixing* \( t \leq T_i \); see Figure 19.1.

If we discretize the volatility function corresponding to the tenor structure and define

\[
\sigma_{k,l} := \left( \frac{1}{T_{i+1} - T_i} \int_{T_i}^{T_{i+1}} \sigma_k^2(t) \, dt \right)^{1/2},
\]

the price of an option on the swap rate \( S(T_i, \ldots, T_j; T_i) \) depends only on \( \sigma_{k,l} \) for \( k = i, \ldots, j - 1 \) and \( l = 0, \ldots, i - 1 \).

\(^7\) For inversion of a pricing formula we may use a simple numerical algorithm. For the Black formula (10.2) the price is increasing strictly monotone in the volatility.
Figure 19.1. **Swaption as a function of the forward rates:** The swaption is a function of the forward rates \( L(T_i, T_{i+1}; T_i) \), \( \ldots \), \( L(T_{j-1}, T_j; T_i) \) (all with fixing in \( T_i \)). The corresponding swaption depends only on the joint distribution of these forward rates. Under our model, with given initial conditions \( L_{i,0} \) and correlation \( R = (\rho_{i,j}) \), the swaption price depends on \( \sigma_j(t), \ldots, \sigma_{j-1}(t), \ t \in [0, T_i] \) only. The dynamic of these forward rates beyond the \( t > T_i \) and all other forward rates do not influence the swaption price.

This allows an iterative calculation of \( \sigma_{k,j} \) from given swaption market prices:

For \( i = 1, \ldots, n - 1 \):

For \( j = i + 1, \ldots, n \):

Calculate \( \sigma_{j-1,i-1} \) from the price of an option on \( S(T_i, \ldots, T_j; T_i) \) by considering the already calculated \( \sigma_{k,j} \) with \( k = i, \ldots, j - 1 \) and \( l = 0, \ldots, i - 2 \) from the previous iterations.

To derive \( \sigma_{j-1,i-1} \) from the market price \( V_{\text{swaption}}^{\text{market}}(T_i, \ldots, T_j) \) we have to invert the mapping

\[
\sigma_{j-1,i-1} \leftrightarrow V_{\text{swaption}}^{\text{model}}(T_i, \ldots, T_j).
\]

In principle this mapping may be realized by a Monte Carlo evaluation of the swaption. To allow for faster pricing, and thus faster calibration, analytic approximation formulas for swaption prices within a LIBOR market model have been derived; see also Section 19.4.5.

**Remark 214 (Bootstrapping):** The above procedure of calculating a piecewise constant instantaneous volatility from swaption prices is called **volatility bootstrapping**.

**Remark 215 (Review: Overfitting):** The calculation of a piecewise constant volatility function \( \sigma_{i,j} \) from swaption prices bears the risk of an **overfitting** of the model.
For more see http://www.christian-fries.de/finmath/book.