Statistical Inference for Generalized Mean Reversion Processes

vorgelegt von

Thomas Kott

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)
an der Fakultät für Mathematik
der Ruhr-Universität Bochum
September 2010

Betreuer:
Prof. Dr. Herold Dehling
# Contents

## 1 Preliminaries
1.1 Stochastic Processes ........................................ 5  
1.2 Stochastic Calculus ........................................ 9  
1.3 Statistical Inference for Diffusion Processes .......... 18  
1.4 Linear Operators on Hilbert Spaces ....................... 23

## 2 Time-Discrete Observations
2.1 The Classical Ornstein-Uhlenbeck Process ............... 27  
2.2 Method of Moments: Yule-Walker Estimator ............... 29  
2.3 Maximum Likelihood Estimation ........................... 35  
2.4 Maximum Likelihood in Practice: Numeric Search for Solutions .......... 42  
2.5 Bias through Discretization ............................... 43

## 3 A Maximum Likelihood Approach
3.1 Generalized Ornstein-Uhlenbeck Process ................. 45  
3.2 Maximum Likelihood Estimation ........................... 46  
3.3 Consistency and Asymptotic Normality ................... 48  
3.4 Proofs .................................................... 50

## 4 A Change Point Problem
4.1 Change in the Drift ........................................ 63  
4.2 Generalized Likelihood Ratio Test ......................... 65  
4.3 Asymptotic Behavior of the Test Statistic .............. 67  
4.4 Proof ..................................................... 68

## 5 How to Deal with Jumps?
5.1 Maximum Likelihood vs. Time-Continuous Least Squares .... 79  
5.2 Jump Diffusion Process .................................... 80  
5.3 Time-Continuous Least Squares Estimation ............... 81  
5.4 Maximum Likelihood Estimation ........................... 83  
5.5 Consistency of Least Squares Estimator .................. 86  
5.6 Proof ..................................................... 87

## 6 Simulation
93

Bibliography ..................................................... 99
Introduction

This thesis is concerned with large sample theory in statistical inference for parametric time-inhomogeneous diffusion processes which are defined as solutions of stochastic differential equations of the form

\[ dX_t = \mu(t, X_t)dt + \sigma \, dL_t, \quad t \geq 0, \]

where \((L_t)_{t \geq 0}\) is a Lévy process or in particular a Brownian motion. Drawing a statistical conclusion on the parameter from historical observations after a parametric model has been chosen is one of the primary interests in practice. The outcome of such a conclusion might be a decision about making further analysis or an adjustment of the statistical model before its final implementation. This work concentrates on statistical inference for the parameter \(\theta\) of the drift \(\mu(t, X_t, \theta)\) which is linear in \(\theta\). Thereby, the diffusion coefficient \(\sigma\) is considered known and independent of \(\theta\). The incentive for this framework is stated below.

The study of inference problems for stochastic processes can be divided into two categories in terms of the observation scheme: the considered stochastic process may be assumed to be observed either in continuous time or at a discrete set of time points. Here, the focus of the theory lies on the continuous time framework providing a continuous time sample path \(\{X_t, 0 \leq t \leq T\}\). The asymptotic results which are the main purpose in large sample theory are obtained as \(T \to \infty\). This asymptotic concept complies with the classical setting of discrete observations \(X_1, \ldots, X_n, n \in \mathbb{N}\), where \(n \to \infty\). The reasons justifying this continuous time setting are the following: first, there exist many real-life stochastic systems that evolve in continuous time such that diffusion processes seem to form an adequate class to cover this nature. In order to obtain ‘optimal’ statistical inference procedures, the theory requires time-continuous observations which are in line with the original time-continuous model. Second, estimators and statistics derived from time-continuous observations of diffusion processes are natural objects to study since they often allow for a closed-form representation in terms of stochastic integrals such that techniques from stochastic calculus can be applied. And even though time-continuous samples are not observed in actual practice, due to an increased processing power of up-to-date computers, simulations and statistical experiments can provide discrete data which are sufficiently dense such that the discretization error may be neglected. In such a case the continuous time results are valid because the statistical error of the data prevails the discretization error. In the situation where the available data are not sufficiently dense it might be helpful to compare the discrete time statistics with the ‘optimal’ asymptotic properties of the continuous time statistics in order to get a better understanding of the discretization error, for example. The interesting problem of the accuracy of a discretization of a diffusion process and its estimators has been studied by many authors and is beyond the scope of this work. For a simple diffusion, namely for the classical Ornstein-Uhlenbeck process, the discretization error is briefly investigated in Section 2.5. However, Monte Carlo simulations which require an approximation of a time-continuous process by its time-discrete version and
of Riemann and Itô integrals by corresponding sums are presented. These simulations provide good results for quite fairly partitions of the time interval in which the process is observed.

Throughout this work, the diffusion parameter $\sigma$ is assumed to be known. This is a usual assumption in the theory of statistical inference for the drift of time-continuously observed diffusion processes. It can be justified by the fact that the quadratic variation of the process can be used to compute, rather than estimate, the volatility $\sigma$ when a time-continuous sample path is available. The challenge of the analysis of the asymptotics of drift estimators is in this setting the time-inhomogeneous property of the process such that it is neither stationary nor ergodic. That means that classical theorems like the ergodic theorem cannot be applied directly. The essential idea of the asymptotic study is the interpretation of the stochastic process as a sequence of random variables that take values in some function space. This method originates from probability theory on Banach spaces.

Outline

The present dissertation is organized as follows: After the Preliminaries which summarize some fundamental notions required for the purpose of this treatise, Chapter 2 gives an introduction to the large sample theory in the classical setting of a time-discretely sampled time-continuous stochastic process. In detail, two parameter estimation methods for the popular mean-reverting Ornstein-Uhlenbeck process with stochastic differential

$$dX_t = \alpha (L - X_t)dt + \sigma dB_t$$

are investigated. Consistency and asymptotic normality of the estimators of the parameter $(\alpha, L, \sigma)^t$ are proved by means of well-known results from the theory of autoregressive processes and Markov processes.

In Chapter 3 a generalized Ornstein-Uhlenbeck process defined as the solution of

$$dX_t = (L(t) - \alpha X_t)dt + \sigma dB_t, \quad t \geq 0, \quad X_0 = \zeta, \quad (1)$$

where the mean reversion function $L : \mathbb{R}_+ \to \mathbb{R}$ is of the form

$$L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t)$$

with known functions $\varphi_1(t), \ldots, \varphi_p(t)$, is proposed. The large sample behavior of the maximum likelihood estimator of the parameter $\theta = (\mu_1, \ldots, \mu_p, \alpha)^t$ based upon a time-continuous sample $X^T = \{X_t, 0 \leq t \leq T\}$ is studied. In doing so, the focus lies on a periodic mean reversion process of the form (1), that is

$$\varphi_i(t + \nu) = \varphi_i(t),$$

which is a meaningful model in practice, in particular for energy commodity and electricity data, since it captures important properties like mean reversion and seasonality. For this periodic Ornstein-Uhlenbeck process strong consistency and asymptotic normality of the maximum likelihood estimator are obtained. The challenging issue of the investigation is the time-inhomogeneity of the process, that is the time-dependence of $L(t)$ resulting in a process that is neither stationary nor ergodic in the classical sense.

Chapter 4 deals with the problem of detecting a possible change in the drift parameter $\theta = (\mu_1, \ldots, \mu_p, \alpha)^t$ of the periodic mean reversion process described above with a given sample
path $X^T = \{X_t, 0 \leq t \leq T\}$ at hand. Thereby, the process (1) is written as

$$dX_t = \left(S(t, X_t, \theta)1_{\{t \leq \tau\}} + S(t, X_t, \theta')1_{\{t > \tau\}}\right)\,dt + \sigma dB_t, \quad 0 \leq t \leq T,$$

where

$$S(t, X_t, \theta) = \sum_{i=1}^{p} \mu_i \varphi_i(t) - \alpha X_t,$$

and the likelihood ratio test is proposed as a hypothesis test that decides whether or not there exists some $\tau \in [0, T]$ of the form $\tau = sT$, $s \in (0, 1)$, such that $\theta \neq \theta'$. It is shown that the corresponding test statistic converges in distribution under the null hypothesis of no change to the quadratic form of a multi-dimensional Brownian bridge. This result may be used in practice to determine, via simulations, a test decision function. Change point analysis is in general a relevant issue in application since structural changes appear in many situations of experimental and statistical sciences. It was originally investigated in the classical i.i.d. case and then extended to time series but there are just a few results on diffusion processes with time-continuous sample paths.

Finally, Chapter 5 treats the case of a time-inhomogeneous jump diffusion model which is driven by a discontinuous Lévy process $(L_t)_{t \geq 0}$ and whose drift is linear in the parameter. In detail, this process is defined as the solution of

$$dX_t = f(t, X_t)\,dt + \sigma\,dL_t \tag{2}$$

where

$$f(t, x) = (f_1(t, x), \ldots, f_p(t, x)), \quad p \in \mathbb{N},$$

and where all real-valued functions $f_i(t, x)$ are assumed to be known. For this model a time-continuous least squares estimator of $\theta \in \mathbb{R}^p$, that is least squares based on time-continuous data, is derived and its advantage in application over the time-continuous maximum likelihood estimator is demonstrated. A specification of the drift function in (2) to

$$f(t, x) = (\varphi_1(t), \ldots, \varphi_p(t), -x), \quad p \in \mathbb{N},$$

with real-valued functions $\varphi_1(t), \ldots, \varphi_p(t)$ results in a mean-reverting jump diffusion process of Ornstein-Uhlenbeck type, compare (1). If the functions $\varphi_i$ are assumed to be periodic, consistency of the time-continuous least squares estimator of $\theta = (\theta_1, \ldots, \theta_p, \alpha)^t \in \mathbb{R}^p \times \mathbb{R}_+$ is proved under the condition that the Lévy process is $L^2$-integrable and centered. In the case of a continuous version of this Ornstein-Uhlenbeck type model, that is $L_t = B_t$, the time-continuous least squares estimator equals the maximum likelihood estimator.

The Chapters 3, 4 and 5 establish the main results of this thesis. These results constitute new findings in the theory of large samples for time-continuously observed diffusion processes and are connected with the following scientific papers:


Motivation

The class of time-continuous diffusion processes covers a large number of stochastic processes and is widespread in many applied problems of scientific fields such as physics, engineering, biology and financial mathematics, for example. Diffusion processes served originally as models for the physical process of diffusion, that is the thermal motion of particles of a solid, liquid or gas resulting in a gradual mixing of material. Briefly, a diffusion process can be mathematically described as a time-continuous Markov process with continuous sample paths.

The motivation of the investigation in this work is the application of diffusion models to financial mathematics, particularly as models for energy commodity prices. Modeling energy commodity markets has been gaining in importance over the last years due to recent developments of the European energy market: The big energy companies which previously monopolized the energy markets undergo nowadays more competition typically due to recent European government regulations. The political policy has created spot and forward energy trading markets that are growing in terms of a rapidly increasing market liquidity, that is the traded volume. Although it is not possible to talk about a world price for natural gas because the world market for natural gas is fragmented into different regional markets - in sharp contrast to crude oil - the progress of new transportation techniques, e.g. cargo ships transporting liquified natural gas, increases the interdependence of these regional markets. These developments boost the price competition and account for severe price fluctuations. Thereby, market fluctuations create exposure to risk which has to be managed by the energy firms. On the other hand, the energy companies have the opportunity to realize additional profits from market fluctuations and liquidity. Indeed, cash flow can be generated by sophisticated trading strategies on the spot and forward market as well as by a deliberate operation of available assets like volume-flexible supply contracts, storages and transport capacities. Consequently, risk management, hedging, option valuation and portfolio optimization have become essential for such companies in order to compete with other market participants and to benefit from the new market environment. Therefore, stochastic processes as models for both commodity and energy prices and supply volumes are required. In this context, energies are different from money markets which are mature markets while energy markets are the most recent to be transformed and replicated by quantitative analysis. Energies respond to underlying price drivers that differ dramatically from well-developed interest rates, for example, and exhibit more complicated features requiring different modeling issues. Hence, diffusion processes that are commonly applied in mathematical finance like the geometric Brownian motion which is the underlying process of the famous Black-Scholes formula or the classical Ornstein-Uhlenbeck process are not adequate models for energies and commodities. For instance, a mean-reverting behavior to a time-dependent, particularly to a periodic, mean function, and abrupt jumps are important phenomena frequently observed in data from different fields of the energy industry.
Chapter 1

Preliminaries

This chapter gives an introduction into some essential concepts of stochastic analysis and statistics such as different classes of stochastic processes, stochastic calculus including both the classical Itô theory and stochastic integrals with discontinuous processes as integrators and the maximum likelihood method for diffusions. The last section discusses briefly some notions and results from the theory of linear operators which are applied in Section 3.4.

1.1 Stochastic Processes

The notion of stochastic processes is one of the most important mathematical developments of the twentieth century since stochastic processes are not only rich from the mathematical point of view: they also have a wide range of applications in many fields like physics, engineering, economics and ecology, for example. The probabilistic concept of stochastic processes and its fundamental comprehension are due to the Russian mathematician Andrey Kolmogorov.

In observing a stochastic system or performing a random experiment, one might be interested in various numerical quantities resulting from the outcomes of the experiment or observation. A randomly driven outcome is usually viewed as a random variable $X$. That is a measurable function $X : \Omega \rightarrow S$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Thereby, $\Omega$ is a non-empty space whose subsets, contained in the $\sigma$-algebra $\mathcal{F}$, are measured with the probability measure $P$. The space $S$ is the state space and is equipped with some $\sigma$-algebra $\mathcal{S}$. Measurable means that $X^{-1}(U) = \{\omega \in \Omega : X(\omega) \in U\} \in \mathcal{F}$ for all sets $U \in \mathcal{S}$. In practice, the state space is usually $\mathbb{R}^d$ equipped with the natural Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^d)$ generated by the open sets in $\mathbb{R}^d$.

Intuitively, the concept of stochastic processes aims to model the interaction of randomness with time: In a stochastic system, the evolution over time of the phenomena might be of interest. This can be viewed as a collection of random variables. The general definition is given now:

**Definition 1.1.** A stochastic process $(X_t)_{t \in T}$ is a family of $S$-valued random variables defined on some probability space $(\Omega, \mathcal{F}, P)$ indexed by $t \in T$ where $T$ is some non-empty index set. It can be thus formulated as a function $T \times \Omega \rightarrow S$ such that $X_t(\cdot)$ is measurable for each $t \in T$.

For each $\omega \in \Omega$ the function $t \mapsto X_t(\omega)$ is called a realization, a (sample) path or a trajectory of the process.

Usually, $\mathbb{R}^d$-valued, continuous-time stochastic processes are considered, that is $S = \mathbb{R}^d$ and $T \subseteq \mathbb{R}_+$ such that the index $t$ is interpreted as time. Each $X_t(\omega)$ should be thought of
as a random observation made at time $t$. However, there is considerable interest in $S$ being a Hilbert or Banach space or a manifold. If $\mathcal{T}$ is countable the stochastic process is nothing but a sequence of random variables. For $\mathcal{T} = \mathbb{Z}$ it is referred to as time series.

At the most general level no restrictions on the relationship between the random variables $X_t$ at different time instants $t$ are imposed. In practice, it is useful to do so in order to adapt the quite general concept of stochastic processes to real-world phenomena. That is why both theory and applications distinguish various classes of stochastic processes according to their specific temporal dependencies.

In the following, some of the most popular classes of stochastic processes are introduced.

Lévy Processes

Lévy processes form a class of a quite simple stochastic processes including a number of important processes as special cases, like the Brownian motion and the Poisson process, for example. On the other hand, every Lévy process belongs to the class of Markov processes. Lévy processes are named after the French mathematician Paul Lévy who first studied them in the 1930s. For notational simplification, only real-valued processes and the index set $\mathcal{T} = \mathbb{R}_+$ are considered henceforth.

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

**Definition 1.2.** A stochastic process $(L_t)_{t \geq 0}$ defined on $(\Omega, \mathcal{F}, P)$ with $L_0 = 0$ is called a Lévy process, if for all $0 \leq s < t < \infty$

- it has independent increments, i.e. $L_t - L_s$ is independent of $L_s$
- it has stationary increments, i.e. $L_t - L_s \overset{D}{=} L_{t-s}$, $0 \leq s < t < \infty$
- $L_t$ is continuous in probability, i.e. for all $a > 0 \lim_{t \to s} P(|L_t - L_s| > a) = 0$.

**Example 1.3.** The (standard, one-dimensional) Brownian motion $(B_t)_{t \geq 0}$ is a Lévy process such that

- $B_t \sim \mathcal{N}(0, t)$ for each $t \geq 0$
- the paths $t \mapsto B_t$ are continuous almost surely.

The Brownian motion is the most popular Lévy process and is among the simplest continuous-time stochastic processes. It was originally introduced by the Scottish botanist Robert Brown and served originally as a model in particle theory. Today it is used in many fields, often to construct more complicated processes.

Note that it can be shown that a Lévy process $(L_t)_{t \geq 0}$ has continuous paths if and only if it is a Brownian motion with drift, i.e. it is of the form $bt + cB_t$.

**Example 1.4.** The Poisson process $(N_t)_{t \geq 0}$ with intensity $\lambda$ is an $\mathbb{N}$-valued Lévy process such that $N_t$ is Poisson-distributed with parameter $\lambda t$ for all $t \geq 0$, i.e.

$$P(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$
Example 1.5. Let \((Z_n)_{n \in \mathbb{N}}\) be a sequence of i.i.d. \(\mathbb{R}\)-valued random variables and let \((N_t)_{t \geq 0}\) be a Poisson process with intensity \(\lambda\) that is independent of all \(Z_n\). Then the process \((Y_t)_{t \geq 0}\) defined by

\[
Y_t = \sum_{i=1}^{N_t} Z_i, \quad t \geq 0,
\]

is called compound Poisson process. It can be verified that \((Y_t)_{t \geq 0}\) belongs to the Lévy class. By conditioning and the i.i.d. property it holds that

\[
E(Y_t) = E(\mathbb{E}(Y_t|\sigma\{N_s, s \leq t\})) = E(N_t Z_1) = \lambda t E(Z_1)
\]

where \(\sigma\{N_s, s \leq t\}\) is the \(\sigma\)-algebra generated by the set \(\{N_s, s \leq t\}\).

For every Lévy process \((L_t)_{t \geq 0}\) there exists a modified process \((L'_t)_{t \geq 0}\), modification in short, with càdlàg paths (right-continuous with left limits) which is again a Lévy process satisfying \(P(L_t = L'_t) = 1\) for each \(t\). Hence, for the remainder of this thesis, the càdlàg version of the Lévy process is always considered.

The discontinuities of a Lévy process, also called jumps, play a key role in the dynamics of the process. Therefore, define the jump process \(\Delta L_t := L_t - L_t^-\) where \(L_t^- = \lim_{\varepsilon \to 0^-} L_{t-}\). Instead of analyzing the process \((\Delta L_t)_{t \geq 0}\) itself it is more reasonable to count jumps of specified size (greater than zero) that occur up to a time instant \(t\). For a Borel set \(A \in \mathbb{R}\) that is bounded away from zero, i.e. the closure \(\bar{A}\) does not contain zero, the counting process \((q(t, A))_{t \geq 0}\) associated with \((L_t)_{t \geq 0}\) is defined as

\[
q(t, A) = \#\{0 \leq s \leq t : \Delta L_s \in A\} = \sum_{0 \leq s \leq t} 1_A(\Delta L_s)
\]

where \(1_A\) denotes the indicator function of the set \(A\). It can be shown that \((q(t, A))_{t \geq 0}\) is a Poisson process with intensity \(\lambda = E(q(1, A))\). For each \(t \geq 0\) \(q(t, \cdot)\) is a random counting measure on \(\mathbb{R} \setminus \{0\}\). The object \(q\{\cdot, \cdot\}\) is called Poisson random measure on \(\mathbb{R}_+ \times (\mathbb{R} \setminus \{0\})\).

Two main results in the elementary theory of Lévy processes are given in the following. Both results are closely connected to each other helping to understand the fundamental structure of a Lévy process. The proofs can be found in many books, see Protter [37] (Theorem I.42 and I.43, p. 32) for instance.

Theorem 1.1 (Lévy-Khintchine Formula). The characteristic function \(\varphi_{L_t}(s) := E(e^{isL_t})\) of a Lévy process \((L_t)_{t \geq 0}\) is of the form \(\varphi_{L_t}(s) = e^{\eta(s)}\) where

\[
\eta(s) = ibs - \frac{1}{2} c^2 s^2 + \int_{\mathbb{R} \setminus \{0\}} (e^{ixs} - 1 - isx1_{|x|<1}(x)) \nu(dx)
\]

where \(b \in \mathbb{R}\), \(c > 0\) and \(\nu\) is a finite Borel measure on \(\mathbb{R} \setminus \{0\}\), also known as Lévy measure, satisfying

\[
\int_{\mathbb{R} \setminus \{0\}} (x^2 \wedge 1) \nu(dx) < \infty.
\]

The map \(\eta : \mathbb{R} \to \mathbb{C}\) is called characteristic exponent, Lévy exponent or Lévy symbol. The triple \((b, c, \nu)\) is referred to as characteristic of the Lévy process.
The form of the characteristic exponent shows that the distribution of a Lévy process may be viewed as an interplay of a normal and a Poisson measure.

Note that a standard Brownian motion has characteristic $(0, 1, 0)$ and $(b, 1, 0)$ if it has a drift. For a Poisson process with intensity $\lambda$ the characteristic is $(0, 0, \lambda \delta_1)$, $\delta_1$ being the dirac measure with mass on $1$, and for a compound Poisson process it is $(0, 0, \lambda \mu_Z)$ where $\mu_Z$ denotes the common distribution of the $Z_i$’s used in the construction of the process.

The Lévy-Itô decomposition of a Lévy path into continuous and jump parts is a famous result in that theory.

**Theorem 1.2** (Lévy-Itô decomposition). For a Lévy process $(L_t)_{t \geq 0}$ there exist some $b \in \mathbb{R}$, $c > 0$ and a Poisson random measure $q$ on $\mathbb{R}_+ \times (\mathbb{R} \setminus \{0\})$ such that for all $t \geq 0$

$$L_t = bt + cB_t + \int_0^t \int_{|x| < 1} x \tilde{q}(dt, dx) + \int_0^t \int_{|x| \geq 1} x q(dt, dx)$$

where $\tilde{q}$ is the so-called compensated Poisson random measure given by $\tilde{q}(dt, dx) = q(dt, dx) - dt \nu(dx)$ while $\nu$ is the Lévy measure satisfying $\int_{\mathbb{R} \setminus \{0\}} (x^2 \wedge 1) \nu(dx) < \infty$. It holds that $b = E(L_1) - \int_{|x| \geq 1} x \nu(dx)$.

The last summand in the Lévy-Itô decomposition describes the ‘large’ jumps of the Lévy process, here jumps on $[0, t]$ with absolute jump size greater than or equal to $1$. Of course the threshold $1$ could be replaced by any number $\delta > 0$. Since the càdlàg modification of the process is considered the number of such big jumps on $[0, t]$ is finite and it holds

$$\int_0^t \int_{|x| \geq 1} x q(dt, dx) = \sum_{s \leq t} \Delta L_s 1_{\{\Delta L_s \geq 1\}}(\Delta L_s).$$

The process $\left(\int_0^t \int_{|x| \geq 1} x q(dt, dx)\right)_{t \geq 0}$ is a compound Poisson process.

In contrast to the sum of big jumps, the sum of small jumps does not converge in general since there might be too many small jumps such that the sum of those can be infinite. Further, it may be difficult to distinguish very small jumps from continuous fluctuations. This problem can be solved by compensating, that is subtracting the expected increase of the process on $[0, t]$ which can be expressed as

$$E\left(\int_0^t \int_{|x| < 1} x q(dt, dx)\right) = t \int_{|x| < 1} x \nu(dx)$$

where $0 < \delta < 1$. The crucial result is that

$$\int_0^t \int_{|x| < 1} x (q(dt, dx) - dt \nu(dx)) \to \int_0^t \int_{|x| < 1} x (q(dt, dx) - dt \nu(dx)) \quad (\text{in } L^2)$$

as $n \to \infty$. The limit process $\left(\int_0^t \int_{|x| < 1} x \tilde{q}(dt, dx)\right)_{t \geq 0}$ is an $L^2$-martingale, see the next subsection, and captures the ‘small’ jumps.

**Martingales**

The concept of martingale is a generalization of the sequence of partial sums arising from a sequence of i.i.d. random variables.
1.2. STOCHASTIC CALCULUS

Let \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)\) be a filtered probability space where \(\mathcal{F}_0\) contains all \(P\)-null sets of \(\mathcal{F}\). Filtered means that \((\mathcal{F}_t)_{t \geq 0}\) is a family of sub \(\sigma\)-algebras such that \(\mathcal{F}_s \subseteq \mathcal{F}_t\) whenever \(s \leq t\). A stochastic process is called \(\mathcal{F}_t\)-adapted if the random variable \(X_t\) is \(\mathcal{F}_t\)-measurable, i.e. \(X_t^{-1}(U) = \{\omega \in \Omega : X_t(\omega) \in U\} \in \mathcal{F}_t\) for all sets \(U \in \mathcal{B}(\mathbb{R})\).

**Definition 1.6.** An \(\mathcal{F}_t\)-adapted stochastic process \((M_t)_{t \geq 0}\) is called martingale if \(E|M_t| < \infty\) for all \(t\) and

\[
E(M_t|\mathcal{F}_s) = M_s
\]

almost surely, for all \(s \leq t\).

Submartingales and supermartingales are defined by replacing the equality in the previous definition with \(E(M_t|\mathcal{F}_s) \geq M_s\) for submartingale and \(E(M_t|\mathcal{F}_s) \leq M_s\) for supermartingale.

**Example 1.7.** Define the partial sum \(S_n = \sum_{i=1}^{n} X_i\) for a sequence of i.i.d. random variables \(X_1, \ldots, X_n\) with \(E|X_1| < \infty\). Let \((\mathcal{F}_n)_{n \in \mathbb{N}}\) be the natural filtration, i.e. \(\mathcal{F}_n = \sigma\{X_k, k \leq n\}\). If \(E(X_1) \leq 0\) then \((S_n)_{n \in \mathbb{N}}\) is a supermartingale, if \(E(X_1) \geq 0\) then \((S_n)_{n \in \mathbb{N}}\) is a submartingale. For \(E(X_1) = 0\), the partial sum sequence \((S_n)_{n \in \mathbb{N}}\) forms a martingale.

**Example 1.8.** A standard Brownian \((B_t)_{t \geq 0}\) motion is a martingale with respect to its natural filtration \(\mathcal{F}_t = \sigma\{B_s, s \leq t\}\) since

\[
E(B_t|\mathcal{F}_s) = E(B_t - B_s|\mathcal{F}_s) = E(B_s|\mathcal{F}_s) = B_s
\]

because the increment \(B_t - B_s\) is independent of \(\mathcal{F}_s\) and \(B_s\) is obviously \(\mathcal{F}_s\)-measurable.

Similarly to Lévy processes, for a martingale \((M_t)_{t \geq 0}\) there exists a càdlàg modification \((M'_t)_{t \geq 0}\) if and only if the function \(t \mapsto E(M_t)\) is right-continuous. The modification \((M'_t)_{t \geq 0}\) is again a martingale and satisfies \(P(M_t = M'_t) = 1\) for every \(t\).

1.2 Stochastic Calculus

Diffusion processes are usually defined in terms of stochastic differential equations. Loosely speaking, a stochastic differential equation can be viewed as a combination of an ordinary differential \(dx_t/dt = a(t, x_t)\) with an additional random term \(\sigma(t, x_t)dZ_t\) where \(Z_t\) is a stochastic process resulting in

\[
dX_t = a(t, X_t)dt + \sigma(t, X_t)dZ_t.
\]

The solution \(X_t\), if it exists, is then a stochastic process. A naive interpretation of this stochastic differential equation indicates that the change \(X_{t+dt} - X_t\) is caused by a change \(dt\) of time with factor \(a(t, X_t)\) in combination with a change \(Z_{t+dt} - Z_t\) with factor \(\sigma(t, X_t)\). Actually, the stochastic differential equation given above is a short form of the stochastic integral equation

\[
X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t \sigma(s, X_s)dZ_s.
\]

The Japanese Kiyoshi Itô was the first to define the stochastic integral \(\int_0^t \sigma(s, \omega)dB_s\) in a reasonably mathematical manner whereby \(B_s\) is a Brownian motion. Since with probability one the function \(t \mapsto B_t\) is nowhere differentiable, this integral cannot be defined in the ordinary way. However, the stochastic properties of Brownian motion admit of a definition of the stochastic integral for a quite large class of integrands.

It is possible to replace Brownian motion by another class of stochastic processes. This is done in the subsequent sections.
Itô Integrals

The concept of the definition of the Itô integral is natural: First one defines the integral for a simple class of functions, like in the ordinary Riemann calculus for example, and then this integral is extended to a larger class of functions by approximation.

Let \((B_t)_{t \geq 0}\) be a standard Brownian motion on \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)\).

**Definition 1.9.** Let \(V = \mathcal{V}(T), \ T > 0\), be the class of real-measurable, \(\mathcal{F}_t\)-adapted functions \(f(t, \omega) : [0, \infty] \times \Omega \to \mathbb{R}\) such that

\[
\|f\|_V := \mathbb{E}\left(\int_0^T f^2(t, \omega)dt\right)^{\frac{1}{2}} < \infty.
\]

Note that \(V\) is complete with respect to the semimetric induced by \(\|\cdot\|_V\).

**Definition 1.10.** Let \(V_0 \subset V\) be the sub class of elementary functions, i.e. each \(f \in V_0\) has the representation

\[
f(t, \omega) = \sum_{j=1}^n f_j(\omega) \mathbf{1}_{(t_j, t_{j+1}]}(t)
\]

for a partition \(0 = t_0 < t_1 < \ldots < t_n = T\) where each \(f_j\) is \(\mathcal{F}_{t_j}\)-measurable and \(\sup_j f_j < \infty\).

The next result is a crucial building block in the construction of the stochastic integral.

**Lemma 1.11.** \(V_0\) is dense in \(V\) with respect to \(\|\cdot\|_V\).

*Sketch of Proof.* For \(f \in V\) define \(g_n := f(t, \omega)\mathbf{1}_{[-n, n]}(f(t, \omega))\). Obviously \(g_n \in V\) and \(\|f - g_n\|_V \to 0\) as \(n \to \infty\) by the dominated convergence theorem. Thus it is enough to show that for every bounded \(f \in V\) there exists a sequence \((f_n)_n\) such that \(f_n \in V_0\) for each \(n\) and \(\|f - f_n\|_V \to 0\). The required sequence is then of the form \(f_n(t, \omega) = f\left(\frac{k}{2^n}, \omega\right)\) for \(t \in \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right], \ k = 0, 1, \ldots\), and \(f_n(0, \omega) = f(0, \omega)\). A complete proof can be found in Kuo [29].

Let \(L^2 = L^2(\Omega)\) be the space of all square integrable, real-valued random variables \(X\) on \((\Omega, \mathcal{F}, P)\) with inner product \((X, Y)_P := E(XY)\). Bear in mind that \(L^2\) equipped with the norm \(\|\cdot\|_2\) induced by the inner product, i.e. \(\|X\|_2 = ((X, X)_P)^{1/2} = E(X^2)^{1/2} < \infty\), is a Hilbert space.

Define for an elementary function \(f \in V_0\) the stochastic integral \(\mathcal{I}_f\) on \([0, T]\) with regard to the Brownian motion \((B_t)_{t \geq 0}\) by

\[
\mathcal{I}_f = \mathcal{I}_f(\omega) = \int_0^T f(t, \omega)dB_t(\omega) := \sum_{j=1}^n f_j(\omega)(B_{t_{j+1}}(\omega) - B_{t_j}(\omega)).
\]

By making use of the independence of the Brownian increments the following observation, called Itô isometry, can be made: The mapping \(\mathcal{I} : (V_0, \|\cdot\|_V) \to (L^2, \|\cdot\|_2)\) is a linear isometry, i.e.

\[
\|\mathcal{I}_f\|_2 = \|f\|_V
\]  (1.1)

and \(\mathcal{I}_{f+g} = \mathcal{I}_f + \mathcal{I}_g\).

Now, if \(f \in V\) then, by Lemma 1.11, there exists a sequence \((f_n)_{n \geq 1}\) of elementary functions such that \(\|f - f_n\|_V \to 0\) as \(n \to \infty\). Since

\[
\|\mathcal{I}_{f_m} - \mathcal{I}_{f_n}\|_2^2 = \|f_m - f_n\|_V^2 \leq 2\|f_m - f\|_V^2 + 2\|f - f_n\|_V^2 \to 0
\]
as \(m, n \to \infty\), the sequence \((\mathcal{I}_{f_n})_{n \geq 1}\) is a Cauchy sequence in \(L^2\). Therefore, the following definition is meaningful.
1.2. STOCHASTIC CALCULUS

Definition 1.12. For $f \in \mathcal{V}$ the stochastic integral $I_f$ is defined as

$$I_f := \lim_{n \to \infty} I_{f_n}$$

where the limit is obtained in $L^2$ and where $(f_n)_{n \geq 1}$ is a sequence in $\mathcal{V}_0$ such that $\|f - f_n\|_{\mathcal{V}} \to 0$ as $n \to \infty$. The usual notation is $I_f = \int_0^T f(t, \omega) dB_t$.

The Itô isometry and the linearity of the mapping $I : \mathcal{V} \to L^2$, $f \mapsto I_f$, can be used to show that the stochastic integral is well-defined, i.e. that the limit used in the definition is independent of the choice of the sequence $(f_n)_{n \geq 1}$.

The following implication of the Itô isometry is useful.

Corollary 1.13. It holds for $f, g \in \mathcal{V}$ that

$$E(I_f I_g) = E\left(\int_0^T f(t, \omega) g(t, \omega) dt\right).$$

Proof. By linearity and (1.1) it holds on the one hand that

$$E\left(\left(I_f + I_g\right)^2\right) = E\left(\left(I_{f+g}\right)^2\right) = E\left(\int_0^T (f(t, \omega) + g(t, \omega))^2 dt\right)$$

and on the other hand

$$E\left(\left(I_f + I_g\right)^2\right) = E\left(I_f^2 + 2I_f I_g + I_g^2\right) = E\left(\int_0^T f^2(t, \omega) dt + \int_0^T g^2(t, \omega) dt + 2I_f I_g\right).$$

\Box

Stochastic Integrals with respect to Martingales

Let $(M_t)_{t \geq 0}$ be a martingale with càdlàg paths. In order to define the stochastic integral $\int_0^T f(t, \omega) dM_t$ one has to take care of the discontinuities of the martingale. Therefore, the concept of predictable processes is needed. It can be intuitively motivated by the following consideration: If $h$ is a right-continuous step function on $[0, T]$ of the form $h(t) = 1_{[0,a)}(t) + 1_{[a,T]}(t)$, $0 < a < T$, then for a bounded function $f$ the Riemann-Stieltjes integral $\int_0^T f dh$ exists if and only if $f$ is left-continuous at $a$. Predictability is thus a slight extension of left-continuity of sample paths of the process.

Let the filtration $(\mathcal{F}_t)_{t \geq 0}$ be right continuous, i.e. $\mathcal{F}_t = \cap_{u \geq t} \mathcal{F}_u$. Let $\mathcal{P}$ be the $\sigma$-field on $[0, \infty) \times \Omega$ with respect to which all $\mathcal{F}_t$-adapted stochastic processes $(X_t)_{t \geq 0}$ with left-continuous paths are measurable.

Definition 1.14. If $(X_t)_{t \geq 0}$ is a stochastic process such that $(t, \omega) \mapsto X_t(\omega)$ is $\mathcal{P}$-measurable, then the process $(X_t)_{t \geq 0}$ is called predictable.

Example 1.15. If

$$f(t, \omega) = \sum_{j=1}^n f_j(\omega) 1_{(t_j, t_{j+1})}(t)$$

where each $f_j$ is $\mathcal{F}_{t_j}$-measurable and $\sup_j f_j < \infty$ then $(f(t, \omega))_{t \geq 0}$ is predictable.
Example 1.16. Let \((X_t)_{t \geq 0}\) be an \(\mathcal{F}_t\)-adapted stochastic process with càdlàg paths. Then \((X_{t-})_{t \geq 0}\), where \(X_{t-} = \lim_{s \to t^-} X_s\), is predictable.

The theorem stated now is a corollary to the Doob-Meyer decomposition and plays a crucial role in the construction.

**Theorem 1.3** (Doob-Meyer decomposition). Let \((M_t)_{t \geq 0}\) be a martingale with càdlàg paths satisfying \(E(M_t^2) < \infty\) for all \(0 \leq t < \infty\). Then there exists a uniquely determined, predictable process \((A_t)_{t \geq 0}\) with paths \(t \mapsto A_t\) that are non-decreasing and right-continuous almost surely, \(A_0 = 0\) and \(E(A_t) < \infty\) for all \(0 \leq t < \infty\) such that \((M_t^2 - A_t)_{t \geq 0}\) is a martingale with càdlàg paths.

The process \((A_t)_{t \geq 0}\) in the previous theorem is referred to as compensator or bracket process and will be denoted henceforth by \((\langle M \rangle_t)_{t \geq 0}\).

The class of integrands is specified in the next definition.

**Definition 1.17.** Let \(\mathcal{W} = \mathcal{W}(T), T > 0\), be the class of predictable processes \((f(t, \omega))_{t \geq 0}\) such that

\[
\|f\|_\mathcal{W} := E \left( \int_0^T f^2(t, \omega) d\langle M \rangle_t \right)^{1/2} < \infty.
\]

Since almost all paths \(t \mapsto \langle M \rangle_t\) are non-decreasing, almost all of them have finite variation such that the integral in the definition above is viewed in the Riemann-Stieltjes sense.

Note that \((\mathcal{W}, \| \cdot \|_\mathcal{W})\) is a Hilbert space.

In the case of Brownian motion it holds that \(\langle B \rangle_t = t\) and predictability is the same as being adapted, hence \(\mathcal{W} = \mathcal{V}\) in that case.

The subclass \(\mathcal{W}_0 \subset \mathcal{W}\) of elementary functions is analogously defined as in the previous section, i.e. it consists of functions of the form

\[
f(t, \omega) = \sum_{j=1}^n f_j(\omega) 1_{(t_j, t_{j+1}]}(t)
\]

where \(0 = t_0 < t_1 < \ldots < t_n = T\), every \(f_j\) is \(\mathcal{F}_{t_j}\)-measurable and \(\sup_j f_j < \infty\).

The key result is stated below and can be proved in exactly the same way as Lemma 1.11.

**Lemma 1.18.** \(\mathcal{W}_0\) is dense in \(\mathcal{W}\) with respect to \(\| \cdot \|_\mathcal{W}\).

Now the same procedure as in the case of Brownian motion will be followed: First, if \(f \in \mathcal{W}_0\), the integral is set to be

\[
\mathcal{I}_f^M = \mathcal{I}_f^M(\omega) = \int_0^T f(t, \omega) dM_t(\omega) := \sum_{j=1}^n f_j(\omega)(M_{t_j+1}(\omega) - M_{t_j}(\omega)).
\]

Note that this definition gives an intuitive understanding of the need of the notion of predictability: The integrand \(f_j\) at time instant \(t_j\) is \(\mathcal{F}_{t_j}\)-measurable while \((M_{t_j+1} - M_{t_j})\) is independent of \(\mathcal{F}_{t_j}\) and is ‘attached’ to the ‘future’ from \((t_j, t_{j+1}]\). Hence, the present \(t_j\) and the ‘future’ \((t_j, t_{j+1}]\) should not overlap forcing the elementary function to be left-continuous since the integrand \(t \mapsto M_t\) is right-continuous almost surely.

Again, the mapping \(\mathcal{I}_f^M : (\mathcal{W}_0, \| \cdot \|_\mathcal{W}) \to (L^2, \| \cdot \|_2)\) is an isometry, i.e.

\[
\|\mathcal{I}_f^M\|_2 = \|f\|_\mathcal{W}.
\]
This can be verified by means of the relation

\[ E \left( (I^M_f)^2 \right) = \sum_{j=1}^{n} E \left( f_j^2(\omega)(M_{t_{j+1}} - M_{t_j})^2 \right) = \sum_{j=1}^{n} E \left( f_j^2(\omega)E \left( (M_{t_{j+1}} - M_{t_j})^2 | \mathcal{F}_{t_j} \right) \right), \]

which is obtained by conditioning, and a conclusion to Theorem 1.3, namely

\[ E \left( (M_t - M_s)^2 | \mathcal{F}_s \right) = E \left( (\langle M \rangle_t - \langle M \rangle_s)^2 | \mathcal{F}_s \right), \quad s < t. \]

Lemma 1.18 justifies for any \( f \in \mathcal{W} \) the existence of a sequence \( (f_n)_{n \geq 1} \) in \( \mathcal{W}_0 \) such that \( \|f - f_n\|_W \to 0 \). Due to the isometry property given in (1.2) it holds that \( (\mathcal{I}_f^n)_{n \geq 1} \) is a Cauchy sequence in \( L^2 \) and thus converges to a limit in \( L^2 \).

**Definition 1.19.** For \( f \in \mathcal{W} \) the stochastic integral \( \mathcal{I}_f \) with respect to a martingale \( (M_t)_{t \geq 0} \) with \( E(M^2_t) < \infty \) for all \( 0 \leq t < \infty \) is defined by

\[ \mathcal{I}_f^M := \lim_{n \to \infty} \mathcal{I}_f^n \quad (\text{in } L^2) \]

where \( \|f - f_n\|_W \to 0 \) as \( n \to \infty \) and each \( f_n \in \mathcal{W}_0 \). The integral is denoted as \( \mathcal{I}_f^M = \int_0^T f(t, \omega) dM_t. \)

Again, this definition can be checked to be well-defined.

Note that there exists an extension of the class of permitted integrands in the literature: The condition \( E \left( \int_0^T f^2(t, \omega) d\langle M \rangle_t \right) < \infty \) is loosened to the requirement \( P \left( \int_0^T f^2(t, \omega) d\langle M \rangle_t \right) < \infty \) = 1, see Section 6.6 in Kuo [29]. Therefore, the integrand \( f \) is truncated if the integral \( \mathcal{I}_f^M \) is greater than resulting in a sequence \( (\mathcal{I}_f^{M,n})_{n \geq 1} \) that converges in probability to \( \mathcal{I}_f^M \). Note that the limit in Definition 1.19 is obtained in probability in that case. A similar extension of the notion of stochastic integrals with respect to Brownian motion can be obtained where the finiteness of the expectation is substituted by the almost sure finiteness of the integral.

### Stochastic Integrals with respect to Point Processes

Let \( (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P) \) be given where the filtration is right-continuous.

First, expand the notion of predictability to mappings \( f(t, x, \omega) \) defined on \([0, \infty) \times \mathbb{R} \times \Omega \).

Define \( \bar{P} \) to be the smallest \( \sigma \)-field on \([0, \infty) \times \mathbb{R} \times \Omega \) with respect to which all mappings \( f \) satisfying the following two properties are measurable:

(i) for each \( t > 0 \) the mapping \( (x, \omega) \mapsto f(t, x, \omega) \) is \( (\mathcal{B}(\mathbb{R}) \times \mathcal{F}_t) \)-measurable

(ii) for each \( (x, \omega) \) the mapping \( t \mapsto f(t, x, \omega) \) is left-continuous.

If \( f \) is \( \bar{P} \)-measurable it is said to be predictable.

Let \( (\mathcal{U}, \mathcal{B}(\mathcal{U})) \) be a measurable space. A point function \( p \) on \( \mathcal{U} \) is a mapping \( p : D \to \mathcal{U} \) where \( D \subset [0, \infty) \) is countable. A point function \( p \) defines a counting measure \( N \) on \([0, \infty) \times \mathcal{U} \) equipped with \( \mathcal{B}([0, \infty)) \times \mathcal{B}(\mathcal{U}) \) by

\[ N(t, U) = N_p(t, U) = \# \{ s \in ([0, \infty) \cap D) : p(s) \in U \} = \sum_{s \in ([0, \infty) \cap D)} 1_{U}(p(s)), \]
where $U \in \mathcal{B}(\mathcal{U})$ and $t \geq 0$.

A point process $(p_t)_{t \geq 0}$ is a randomization of the notion of point functions and can be seen as a mapping $p : \Omega \times D \to \mathcal{U}$. Equivalently, a point process on $\mathcal{U}$ can be viewed as an $\mathcal{I}_\mathcal{U}$-valued random variable, where $\mathcal{I}_\mathcal{U}$ is the space of all point functions on $\mathcal{U}$ equipped with the smallest $\sigma$-algebra with respect to which all mappings $p \mapsto N_p(t,U)$ for $t \geq 0$, $U \in \mathcal{U}$ are measurable.

Consider an $\mathcal{F}_t$-adapted, $\sigma$-finite point process $(p_t)_{t \geq 0}$, that means $(N_p(t,U))_{t \geq 0}$ is $\mathcal{F}_t$-adapted for every $U \in \mathcal{B}(\mathcal{U})$ and there exist $U_n \in \mathcal{B}(\mathcal{U})$, $n = 1, 2, \ldots$, such that $U_n \uparrow \mathcal{U}$ and $E(N_p(t,U_n)) < \infty$ for all $t \geq 0$ and $n \geq 1$. Define $V_p := \{ U \in \mathcal{B}(\mathcal{U}) : E(N_p(t,U)) < \infty \}$.

If $U \in V_p$, then the paths $t \mapsto N(t,U)$ are non-decreasing and right-continuous almost surely. A reversion of the assertion in Theorem 1.3 justifies that there exists a process $(\hat{N}(t,U))_{t \geq 0}$ such that $(\hat{N}(t,U))_{t \geq 0}$ defined by

$$\hat{N}(t,U) = N(t,U) - \tilde{N}(t,U)$$

is an $\mathcal{F}_t$-martingale. Note that $\tilde{N}(t,U)$ defines a random measure on $([0,\infty) \times \mathcal{U})$ and is called compensator. It is reasonable at least for the purposes of this work to assume henceforth that for $U \in V_p$, the mapping $t \mapsto \hat{N}(t,U)$ is continuous and that for each $t \geq 0$ and almost all $\omega$, $U \mapsto \hat{N}(t,U)$ is a $\sigma$-finite measure on $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$. These assumptions can be justified by the following example which is the most important application for this work.

**Example 1.20.** Let $(L_t)_{t \geq 0}$ be a Lévy process. Then $(\Delta L_t)_{t \geq 0}$ is a (Poisson) point process on $\mathbb{R} \setminus \{0\}$ defining a counting measure $\tilde{q}$ on $([0,\infty) \times \mathbb{R} \setminus \{0\})$ by

$$q(t,A) = \#\{ 0 \leq s \leq t : \Delta L_s \in A \} = \sum_{0 \leq s \leq t} 1_A(\Delta L_s)$$

where $A \in \mathcal{B}(\mathbb{R} \setminus \{0\})$. The compensated random measure is given by $\tilde{q}(t,A) = q(t,A) - t\nu(A)$ where $\nu$ is the Lévy measure. The process $(\tilde{q}(t,A))_{t \geq 0}$ is a martingale satisfying $E(\tilde{q}(t,A)^2) = t\nu(A)$.

The definition of the notion of stochastic integrals with point processes as integrators goes in line with the constructions in the previous sections:

**Definition 1.21.** Let $\mathcal{H} = \mathcal{H}(\mathcal{T},\mathcal{U})$, $T > 0$, be the class of predictable processes $(f(t,x,\omega))_{t \geq 0}$ such that

$$\|f\|_{\mathcal{H}} := E\left( \int_0^T \int_{\mathcal{U}} f^2(t,x,\omega)d\hat{N}(dt,dx) \right)^{\frac{1}{2}} < \infty.$$  

$(\mathcal{H}, \| \cdot \|_{\mathcal{H}})$ is a Hilbert space.

The space of elementary functions $\mathcal{H}_0 \subset \mathcal{H}$ consists of functions $f$ of the form

$$f(t,\omega) = \sum_{j=1}^n \sum_{k=1}^m f_j(\omega)1_{(t_j,t_{j+1}]}(t)1_{U_k}(x)$$

where each $f_j$ is $\mathcal{F}_{t_j}$-measurable, $0 = t_0 < t_1 < \ldots < t_n = T$ and and disjoint Borel sets $U_1,\ldots,U_m$.

For $f \in \mathcal{H}_0$ the stochastic integral with respect to a random measure is defined as

$$\mathcal{T}_f^\nu = \int_0^T \int_{U_k} f(t,x,\omega)d\tilde{N}(dt,dx) := \sum_{j=1}^n \sum_{k=1}^m f_j(\omega) \left( \hat{N}(t_{j+1},U_k) - \hat{N}(t_j,U_k) \right).$$
As in the other stochastic integral notions, $\mathcal{H}_0$ is dense in $\mathcal{H}$ with respect to $\|\cdot\|_\mathcal{H}$. The mapping $T^p : (\mathcal{H}_0, \|\cdot\|_\mathcal{H}) \to (L^2, \|\cdot\|_2)$ is a linear isometry preserving the norm, that is
\[ \|f\|_\mathcal{H} = \|T^p f\|_2. \]
It thus extends to an isometric embedding of $\mathcal{H}$ into $L^2$. It follows that for every $f \in \mathcal{H}$ there exists a sequence $f_n \in \mathcal{H}_0$ such that $\|f - f_n\|_\mathcal{H} \to 0$ as $n \to \infty$ and
\[ T^p f_n := \lim_{n \to \infty} T^p f_n \quad \text{(in } L^2) \]

Stochastic Integrals with respect to Lévy Processes

Finally, the notion of a stochastic integral of the form
\[ \int_0^T f(s, \omega) dL_s, \]
where $(L_t)_{t \geq 0}$ is a Lévy process, can be explained. Due to the Lévy-Itô decomposition, cf. Theorem 1.2, this integral can be expressed as
\[ \int_0^T bf(s, \omega) ds + \int_0^T cf(s, \omega) dB_s + \int_0^T \int_{|x| < 1} f(s, \omega) x \tilde{q}(ds, dx) + \int_0^T \int_{|x| \geq 1} f(s, \omega) x q(ds, dx). \]
The first integral in this representation has to be regarded in the Riemann sense, the second one is an Itô integral. The last integral is equal to the finite random sum
\[ \sum_{0 \leq s \leq t} f(s, \omega) \Delta L_s 1_{\{|\Delta L_s| > 1\}}(\Delta L_s). \]
The third integral exists as well if $E(\int_0^T f(s, \omega)^2 ds) < \infty$. This is due to Theorem 4.3.4 in Applebaum [2] (p. 207): Let $(\varepsilon_n)_{n \geq 2}$ be a sequence such that $\varepsilon_1 = 1$, $\varepsilon_n \downarrow 0$ and $\varepsilon_{n+1} < \varepsilon_n$. Set $A_n := \{ x : \varepsilon_n < |x| < 1 \}$. Then it holds that
\[ \int_0^T \int_{A_n} f(s, \omega) x \tilde{q}(ds, dx) \to \int_0^T \int_{|x| < 1} \int_0^T f(s, \omega) x \tilde{q}(ds, dx) \quad \text{a.s. as } n \to \infty. \]

Stochastic Differential Equations

The objective of this section is the elaboration of the term of stochastic differential equations. In detail, the Itô lemma which is an Itô integral version of the chain rule is presented. Further, conditions required for the existence and uniqueness of solutions to stochastic differential equations are given. Thereby, Itô processes are considered, that is continuous processes driven by Brownian motion. The more general case allowing for discontinuous process, e.g. processes driven by Lévy processes, is mentioned in between. For the sake of notational simplicity, only one-dimensional processes are considered.

**Definition 1.22.** Let $(B_t)_{t \geq 0}$ be a (standard, 1-dimensional) Brownian motion on $(\Omega, \mathcal{F}, P)$. A stochastic process $(X_t)_{t \geq 0}$ on $(\Omega, \mathcal{F}, P)$ is called an (1-dimensional) Itô process if it has a representation of the form
\[ X_t = X_0 + \int_0^t u(s, \omega) ds + \int_0^t v(s, \omega) dB_s, \tag{1.3} \]
where $u, v : [0, \infty) \times \Omega \to \mathbb{R}$ are stochastic processes with the following properties:
Let \( v \in \mathcal{V}_\infty := \bigcap_{T>0} \mathcal{V}(T) \)

(ii) \( u \) is \( \mathcal{F}_t \)-adapted and \( \int_0^t |u(s, \omega)| \, ds < \infty \) for all \( t \geq 0 \) almost surely.

If \((X_t)_{t \geq 0}\) is an Itô process then equation (1.3) is usually written in the shorter differential form

\[
dX_t = u \, dt + v \, dB_t.
\]

**Theorem 1.4.** Let \((X_t)_{t \geq 0}\) be an Itô process given by (1.4) and \( g(t, x) : [0, \infty) \times \mathbb{R} \to \mathbb{R} \) a twice continuously differentiable function. Then \( Y_t := g(t, X_t) \) is again an Itô process and

\[
dY_t = \frac{\partial g}{\partial t}(t, X_t) \, dt + \frac{\partial g}{\partial x}(t, X_t) \, dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t) (dX_t)^2
\]

where \((dX_t)^2 = (dX_t) \cdot (dX_t)\) is computed according to the rules

\[
dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0, \quad dB_t \cdot dB_t = dt.
\]

**Sketch of proof.** It is possible substitute (1.4) in (1.5) and use the integral notation so that

\[
Y_t = Y_0 + \int_0^t \left( g_t(s, X_s) + u(s, \omega) g_x(s, X_s) + \frac{1}{2} v(s, \omega)^2 g_{xx}(s, X_s) \right) \, ds
\]

\[
+ \int_0^t v(s, \omega) g_x(s, X_s) dB_s(\omega)
\]

where the subscripts to \( g \) denote partial derivatives, e.g. \( g_s(s, X_s) = \frac{\partial g}{\partial s}(s, X_s) \).

Assume that all partial derivatives (i.e. \( g_t, g_x, g_{xx}, \ldots \)) are bounded since the general case is obtained by approximating by \( C^2 \)-functions \( g_n \) such that all partial derivatives of first and second order are bounded for each \( n \) and converge uniformly on compact subsets of \([0, \infty) \times \mathbb{R}\) to \( g, g_t, g_x \) and \( g_{xx} \), respectively. Moreover, due to the definition of the Itô integral, it is assumed that \( u \) and \( v \) are elementary functions. Consider a sequence of partitions \((\pi_n)_n\) of \([0, t]\)

\[
\pi_n = \left\{ 0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_{m(n)}^{(n)} = t \right\}
\]

which improve the partitions of \( u \) and \( v \), respectively, i.e. \( \|\pi_n\| = \max_{1 \leq j \leq m(n)} (t_j^{(n)} - t_{j-1}^{(n)}) \to 0 \) as \( n \to \infty \). It holds that

\[
Y_t - Y_0 = g(t, X_t) - g(0, X_0) = \sum_{j=0}^{m(n)-1} (g(t_{j+1}^{(n)}, X_{j+1}) - g(t_j^{(n)}, X_j))
\]

\[
+ \frac{1}{2} g_t(t_j^{(n)}, X_j)(t_{j+1}^{(n)} - t_j^{(n)})^2 + g_x(t_j^{(n)}, X_j)(t_{j+1}^{(n)} - t_j^{(n)})(X_{j+1} - X_j)
\]

\[
+ \frac{1}{2} g_{xx}(t_j^{(n)}, X_j)(X_{j+1} - X_j)^2 + R_j
\]

where \( X_j = X_{t_j^{(n)}} \). Taylor’s theorem gives for \( j = 0, 1, \ldots, m(n) - 1 \)

\[
g(t_{j+1}^{(n)}, X_{j+1}) - g(t_j^{(n)}, X_j) = g_t(t_j^{(n)}, X_j)(t_{j+1}^{(n)} - t_j^{(n)}) + g_x(t_j^{(n)}, X_j)(X_{j+1} - X_j)
\]

\[
+ \frac{1}{2} g_t(t_j^{(n)}, X_j)(t_{j+1}^{(n)} - t_j^{(n)})^2 + g_x(t_j^{(n)}, X_j)(t_{j+1}^{(n)} - t_j^{(n)})(X_{j+1} - X_j)
\]

\[
+ \frac{1}{2} g_{xx}(t_j^{(n)}, X_j)(X_{j+1} - X_j)^2 + R_j
\]
where $R_j = o\left((t^{(n)}_{j+1} - t^{(n)}_j)^2 + (X_{j+1} - X_j)^2\right)$. Further, it is

$$X_{j+1} - X_j = u(t_j, \omega)(t^{(n)}_{j+1} - t^{(n)}_j) + v(t_j, \omega)(B_{j+1} - B_j).$$

After putting these representations into (1.7) the task is to study the asymptotic behavior of every sum and to prove that the overall limit is equal to (1.6).

If one replaces the driving Brownian motion in (1.4) by a martingale $(M_t)_{t \geq 0}$ with jumps, then the integral version of (1.5) for $g$ as above becomes, see Protter [37] (Theorem II.32, p. 71) or Kuo [29] (Theorem 7.6.1 and 7.6.4, p. 111 and 113),

$$g(t, X_t) = g(0, X_0) + \int_0^t \frac{\partial g}{\partial t}(s, X_s)ds + \int_0^t \frac{\partial g}{\partial x}(s, X_{s-})dX_s + \frac{1}{2} \int_0^t \frac{\partial^2 g}{\partial x^2}(s, X_{s-})d[X]^c_s$$

$$+ \sum_{0 \leq s \leq t} \left( g(s, X_s) - g(s, X_{s-}) - \frac{\partial g}{\partial x}(s, X_{s-}) \Delta X_s \right)$$

where $\Delta X_t = X_t - X_{t-}$ and where $[X]^c_t$ is the continuous part of the quadratic variation process $[X]_t$ defined as

$$[X]_t := \lim_{\|\pi_n\| \to 0} \sum_{i=1}^{m(n)} (X_{t_i} - X_{t_{i-1}})^2$$

whereby $(\pi_n)_n$ is a sequence of partitions of the interval $[0, t]$ as in the proof above. Note that the last term in the Itô formula given above, that is the sum, takes the jumps of the process into account. In the case of a continuous martingale the resulting process is also continuous and all these summands are obviously zero.

In the following an existence and uniqueness theorem for stochastic differential equations driven by Brownian motion is presented. The proof is omitted and can be found in Kuo [29] (Theorem 10.3.5 on p. 192). The two conditions are usually referred to as the linear growth and the Lipschitz condition.

**Theorem 1.5.** Let $T > 0$ and $(B_t)_{t \geq 0}$ be an $\mathcal{F}_t$-adapted Brownian motion. Further, let $b, \sigma : [0, T] \times \mathbb{R} \to \mathbb{R}$ be measurable functions satisfying

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|), \quad x \in \mathbb{R}, \ t \in [0, T],$$

for some constant $C$ and

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|, \quad x, y \in \mathbb{R}, \ t \in [0, T],$$

for some constant $D$. Let $Z$ be a random variable which is independent of the $\sigma$-algebra generated by Brownian motion $(B_s)_{s \geq 0}$ and which satisfies $E|Z|^2 < \infty$. Then the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad t \in [0, T], \ X_0 = Z,$$  \hspace{1cm} (1.8)

has a unique solution $X_t(\omega)$ which is adapted to the filtration generated by $Z$ and $B_s, s \leq t$, almost all paths of that solution are continuous and $E \left( \int_0^T |X_t|^2 dt \right) < \infty$.

For discontinuous processes, analogous conditions are specified in Jacod and Shiryaev [25] (Theorem III.2.32 on p. 145).
1.3 Statistical Inference for Diffusion Processes

In this section, the application of the famous maximum likelihood concept to statistical inference problems for the important class of diffusion processes is demonstrated. Thereby, the likelihood ratio, or likelihood function, is defined for a given sample path as the Radon-Nikodym derivative of two parameterized measures induced by diffusions. The results in that theory rely on the Girsanov theorem which is a fundamental result in stochastic analysis.

Construction of Diffusion Processes

Itô’s original motivation to develop the theory of stochastic calculus was to construct diffusion processes by solving stochastic differential equations. There are basically two ways to define the class of diffusion processes which are special cases of Markov processes with continuous sample paths: One can either construct them in terms of conditions on the transition probabilities or one can study the state $X_t$ itself and its variation with respect to time, see below. It is well known that both ideas provide essentially the same class of processes, cf. Arnold [3] (Theorem 9.3.1, p. 164). In the following, the latter definition is given.

The incentive of the next definition emerged in physics: Let $X_t$ be the position of a particle suspended in a fluid at time instant $t$. Further, let the velocity of the fluid at position $x$ and time instant $t$ be equal to $a(t,x)$ and assume that a sufficiently small fluctuation can be approximated by a random term whose distribution depends upon $x$, $t$ and the length $\Delta t$ of the time interval during which the motion is observed. Hence, the evolution of the motion can be written as

$$X_{t+\Delta t} - X_t = a(t,X_t)\Delta t + \sigma(t,X_t)\zeta_{t,\Delta t}$$

where $E(\zeta_{t,\Delta t}) = 0$ and $E(\zeta_{t,\Delta t}^2) = \Delta t$. If the increments $\zeta_{t,\Delta t}$ and $\zeta_{s,\Delta t}$ are required to be independent and continuous, it turns out that Brownian motion is the driving process in that model. Going from the infinitesimal representation to a differential form yields a stochastic differential equation.

**Definition 1.23.** A stochastic process $(X_t)_{t \geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is called a (1-dimensional, time-inhomogeneous) diffusion process if it satisfies the stochastic differential

$$dX_t = a(t,X_t) \, dt + \sigma(t,X_t) \, dB_t, \quad t \geq 0, \ X_0 = \xi,$$

where $(B_t)_{t \geq 0}$ denotes standard Brownian motion, $E(\xi^2) < \infty$ and $a, \sigma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are functions that satisfy the linear growth and the Lipschitz condition, see Theorem 1.5, and that are further continuous in $t$.

The function $a$ is usually called drift function, or drift for short, and $\sigma$ is referred to as diffusion coefficient.

In the case of time-independent functions $a, \sigma : \mathbb{R} \to \mathbb{R}$ the process is termed time-homogeneous diffusion process.

It is a well known fact that a diffusion process is a Markov process, see Theorem 10.6.1 on p. 204 in Kuo [29]. The definition of a Markov process is given below.

**Definition 1.24.** A stochastic process $(X_t)_{t \geq 0}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is called Markov process with state space $\mathbb{R}$ if for any $t_1 < t_2 < \ldots < t_n$, $n \in \mathbb{N}$, and $B \in \mathcal{B}(\mathbb{R})$ the relation

$$P(X_t \in B|X_{t_1}, X_{t_2}, \ldots, X_{t_n}) = P(X_t \in B|X_{t_n})$$

(1.9)

holds.
1.3. STATISTICAL INFERENCE FOR DIFFUSION PROCESSES

Loosely speaking, the Markov property (1.9) means that given the present and the past, the future state of the process depends only on the present state. In other words, the future, given the present, is independent of the past.

A diffusion process may be alternatively defined in terms of conditions on the transition probabilities \( P(X_s \in \cdot \) \( |X_t = x \), \( s \leq t \), \( x \in \mathbb{R} \), which determine, together with the initial distribution, i.e. the distribution of \( \xi \), the properties of the Markov process. See e.g. Definition IX.2.1. in Basawa and Prakasa Rao [5] (p. 203) for a definition of a diffusion based on the transition probabilities. The crucial point thereby is that the drift \( a \) and the diffusion coefficient \( \sigma \) uniquely determine the transition probabilities and are given by

\[
a(t, x) = \lim_{\varepsilon \to 0} E \left( X_{t+\varepsilon} - x \big| X_t = x \right),
\sigma(t, x) = \lim_{\varepsilon \to 0} E \left( (X_{t+\varepsilon} - x)^2 \big| X_t = x \right).
\]

Maximum Likelihood

In many statistical procedures, in particular in maximum likelihood estimation, the likelihood ratio, or likelihood function, plays a central role. Technically, it is defined as a Radon-Nikodym derivative.

For the sake of comprehension, recall first the classical Radon-Nikodym theorem: Let \( \mu \) and \( \nu \) be \( \sigma \)-finite positive measures on a measurable space \( (\Omega, \mathcal{F}) \). The Radon-Nikodym theorem states that if \( \nu \ll \mu \), i.e. \( \nu \) is absolutely continuous with respect to \( \mu \), then there exists a \( \mu \)-integrable and \( \mu \)-almost surely uniquely determined function \( f: \Omega \to \mathbb{R}_+ \) such that

\[
\nu(A) = \int_A f(\omega) \mu(\mathrm{d}\omega)
\]

for all \( A \in \mathcal{F} \). Absolute continuity \( \nu \ll \mu \) means that \( \mu(A) = 0 \) implies \( \nu(A) = 0 \). The function \( f \) is often denoted as

\[
f = \frac{d\nu}{d\mu}
\]

and is referred to as Radon-Nikodym derivative. In addition to some useful properties like linearity, the Radon-Nikodym derivative \( f \) exhibits the change of measure rule: If \( \nu \ll \mu \) and \( g \) is \( \nu \)-integrable then \( \int g \, d\nu = \int g f \, d\mu \). Moreover, a sort of chain rule can be shown to hold: If \( \nu \ll \mu \ll \lambda \) then

\[
\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \quad (1.10)
\]

**Example 1.25.** Let \( f \) and \( g \) be densities of two probability measures \( P \) and \( Q \) on \( \mathbb{R}^d \). Then \( Q \ll P \) if and only if \( \{ x : f(x) > 0 \} \subseteq \{ x : g(x) > 0 \} \) and in this case \( \frac{dQ}{dP}(x) = \frac{g(x)}{f(x)} \), \( x \in \mathbb{R}^d \).

**Definition 1.26.** Let \((Y, \mathcal{B}(Y), \{P_\theta : \theta \in \Theta\})\) be a statistical model, that is a measurable space \((Y, \mathcal{B}(Y))\) provided with a family of probability measures \( \{P_\theta : \theta \in \Theta\} \). Let further \( \theta_0 \in \Theta \) satisfy \( P_\theta \ll P_{\theta_0} \) for all \( \theta \in \Theta \). Then the maximum likelihood ratio or function at \( y \in Y \) is defined by

\[
L(\theta) := \frac{dP_\theta}{dP_{\theta_0}}(y)
\]

as a function of \( \theta \). The maximum likelihood estimator \( \hat{\theta} \) is defined to be the solution to

\[
L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta).
\]
If this equation has more than one solution, then the estimator has to be chosen by means of further constraints.

In many situations the set of probability measures \( \{P_\theta : \theta \in \Theta\} \) is chosen in such a way that all measures in that set are equivalent, that is \( P_{\theta_0} \ll P_{\theta_1} \) and \( P_{\theta_1} \ll P_{\theta_0} \) for all \( \theta_0, \theta_1 \in \Theta \). This is denoted by \( P_{\theta_0} \sim P_{\theta_1} \).

**Remark 1.27.** Note that the definition of the maximum likelihood estimator does not depend on the dominating reference measure: If there exists some \( \theta_1 \in \Theta \) such that \( P_{\theta} \ll P_{\theta_0} \ll P_{\theta_1} \) for all \( \theta_0, \theta_1 \in \Theta \), then by (1.10)

\[
\frac{dP_{\theta}}{dP_{\theta_0}} = \frac{dP_{\theta}}{dP_{\theta_1}}.
\]

The maximum likelihood ratio \( \mathcal{L}(\theta) \) is defined for a given \( y \in \mathcal{Y} \) as a function of \( \theta \). When considering \( \frac{dP_{\theta_0}}{dP_{\theta}}(y) \) as a function of \( \theta \) it is constant for a fixed \( y \in \mathcal{Y} \). Consequently, if

\[
\sup_{\theta \in \Theta} \frac{dP_{\theta}}{dP_{\theta_0}}(y) = \frac{dP_{\theta}}{dP_{\theta_0}}(y)
\]

then

\[
\sup_{\theta \in \Theta} \frac{dP_{\theta}}{dP_{\theta_0}}(y) = \frac{dP_{\theta}}{dP_{\theta_1}}(y).
\]

For stochastic processes Definition 1.26 is translated as follows: Let \( C[0,T] \) be the space of continuous functions \( \omega : [0,T] \to \mathbb{R} \) and \( \{P_\theta : \theta \in \Theta\} \) a family of probability measures on \( C[0,T] \). Assume that the stochastic process \( (X_t)_{0 \leq t \leq T} \) defined on a probability space \( (\Omega, \mathcal{F}, P) \) with sample paths in \( C[0,T] \) evolves according to one of these measures. Denote by \( X^T = \{X_t, 0 \leq t \leq T\} \) a time-continuous sample path. Fix some \( \theta_0 \) that obeys \( P_\theta \ll P_{\theta_0} \) for all \( \theta \in \Theta \). For a given sample \( X^T \in C[0,T] \) the likelihood ratio is then

\[
\mathcal{L}(\theta) = \frac{dP_\theta}{dP_{\theta_0}}(X^T).
\]

If a particular class of stochastic processes provided with a specified family of probability measures \( \{P_\theta : \theta \in \Theta\} \) is considered, the important and challenging problem is to find conditions for the absolute continuity of the measures with respect to some dominating measure or for the equivalence of all measures and to find their Radon-Nikodym derivatives. However, the theory of absolute continuity of measures corresponding to diffusion processes parameterized in the drift is well developed.

The following chapters are concerned with diffusion processes possessing the stochastic differential

\[
dX_t = a(t, X_t, \theta) \, dt + \sigma \, dB_t, \quad t \geq 0, \, X_0 = \xi,
\]

where \( \theta \in \Theta \subseteq \mathbb{R}^k \) is unknown and the function \( a \) and \( \sigma > 0 \) are known. Denote by \( P_\theta \) the measure induced by the process satisfying (1.12) for some unknown value of \( \theta \). The following theorem is probably the main result in the maximum likelihood theory of diffusion processes. It is proved in Lipster and Shiryaev [33] (Theorem 7.19 on p. 277).

**Theorem 1.6.** Let \( X^T = \{X_t, 0 \leq t \leq T\} \) be a sample path of the process given in (1.12) inducing a family of measures \( \{P_\theta : \theta \in \Theta\} \) on \( C[0,T] \). It holds for some \( \theta_0 \) that \( P_\theta \sim P_{\theta_0} \) for all \( \theta \) if and only if

\[
P\left( \int_0^T (a(t, X_t, \theta) - a(t, X_t, \theta_0))^2 \, dt < \infty \right) = 1
\]
1.3. STATISTICAL INFERENCE FOR DIFFUSION PROCESSES

for all $\theta$. In this case, the Radon-Nikodym derivative $dP_\theta/dP_{\theta_0}$ is given by

$$
\frac{dP_\theta}{dP_{\theta_0}}(X_T) = \exp \left( \frac{1}{\sigma^2} \int_0^T a(t, X_t, \theta) \, dt - \frac{1}{2\sigma^2} \int_0^T a(t, X_t, \theta)^2 \, dt \right),
$$

almost surely.

Note that integrals of the form of the first integral in the expression of the Radon-Nikodym derivative above are understood in accordance to (1.12) in the sense

$$
\int_0^T a(t, X_t, \theta) \, dt = \int_0^T a(t, X_t, \theta)^2 \, dt + \sigma \int_0^T a(t, X_t, \theta) \, dB_t.
$$

In the case of diffusion processes it is common and reasonable to take the law $P_B$ on $C[0, T]$ induced by the driving process which is a Brownian motion $(B_t)_{0 \leq t \leq T}$ to be the dominating measure such that the likelihood ratio is then defined as $dP_\theta/dP_B$. This is due to the possible interpretation of the diffusion process (1.12) as basically a Brownian motion that is transformed by adding a time-depending and stochastic drift function $a$. The change of measure density $dP_\theta/dP_B$ contains the whole information on the drift term required for the estimation of the drift parameter. The condition guaranteeing $P_\theta \ll P_B$ for all $\theta \in \Theta$ is well-known and, due to relation (1.10), the maximum likelihood estimator defined by $\sup_{\theta \in \Theta} dP_\theta/dP_B = dP_{\hat{\theta}}/dP_B$ is the same as the one obtained by maximizing (1.11).

See Theorem 7.6 on p. 246 Lipster and Shiryayev [33] for a proof of the next theorem.

**Theorem 1.7.** Under the framework of Theorem 1.6 it holds that $P_\theta \ll P_B$ if and only if

$$
P \left( \int_0^T a(t, X_t, \theta)^2 \, dt < \infty \right) = 1
$$

for all $\theta$ and then the Radon-Nikodym derivative is

$$
\frac{dP_\theta}{dP_B}(X_T) = \exp \left( \frac{1}{\sigma^2} \int_0^T a(t, X_t, \theta) \, dt - \frac{1}{2\sigma^2} \int_0^T a(t, X_t, \theta)^2 \, dt \right),
$$

almost surely.

**Remark 1.28.** In the case of a jump diffusion process, that is the process solving the stochastic differential equation

$$
dX_t = a(t, X_t, \theta) \, dt + \sigma \, dL_t,
$$

where $(L_t)_{t \geq 0}$ is a homogenous Lévy process, the corresponding Radon-Nikodym derivative $P_\theta/P_L$ takes the form

$$
\exp \left( \frac{1}{\sigma^2} \int_0^T a(t, X_t-, \theta) \, dX_t^c - \frac{1}{2\sigma^2} \int_0^T a(t, X_t, \theta)^2 \, dt \right)
$$

where $X_{t-} = \lim_{\epsilon \to 0} X_{t-\epsilon}$ and where $X^c_t$ denote the continuous part of the discontinuous sample. For more information on that, see Sørensen [39] (Theorem 2.1).
Example 1.29. Define the probability measure $P_a$ on $C[0,T]$ to be the law of $(a + B_t)_{0 \leq t \leq T}$, where $a \in \mathbb{R}$ and $B_t$ is standard Brownian motion. Obviously, it holds $P_a \ll P_0$ and $\frac{dP_a}{dP_0}(B^T) = \exp(aB_T - \frac{1}{2}a^2T)$ where $B^T \in C[0,T]$.

The next example sheds light on the general relation between the Radon-Nikodym derivative and the likelihood ratio since it describes the maximum likelihood methodology in probably the most familiar statistical situation.

Example 1.30. Let $X^n = \{X_1, \ldots, X_n\}$ be realizations of a discrete time stochastic process $(X_k)_{k \in \mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in the measurable space $(\mathbb{R}, B(\mathbb{R}))$. Assume that the random vector $(X_1, \ldots, X_n)$ which induces a probability measure $P_X$ on $(\mathbb{R}^n, B(\mathbb{R}^n))$ has a density $f_X$ with respect to a $\sigma$-finite, positive product measure $\mu$ on $\mathbb{R}^n$, and that $P_X \ll \mu$, i.e.

$$P_X(A) = P((X_1, \ldots, X_n) \in A) = \int_A f_X(y)\mu(dy)$$

for all Borel sets $A$ in $\mathbb{R}^n$. Then the density $f_X$ is the Radon-Nikodym derivative

$$f_X(y) = \frac{dP_X}{d\mu}(y).$$

Let $\theta \in \Theta$ be an unknown parameter yielding a family of densities $\{f_X(y, \theta) : \theta \in \Theta\}$. The parameter is estimated by regarding the Radon-Nikodym derivative $f_X(X^n, \theta)$ for a given sample $X^n = \{X_1, \ldots, X_n\}$ as a function of $\theta$ and by choosing the estimator to be the value of $\theta$ that is the best one, in some sense, to explain the observations $X^n$. The likelihood function is defined as

$$\mathcal{L}(\theta) := f_X(X^n, \theta) = \frac{dP_X^{(\theta)}}{d\mu}(X^n)$$

for a fixed sample $X^n$ and a likelihood estimator $\hat{\theta}$ is defined as any solution to

$$\mathcal{L}(\hat{\theta}) = \sup_{\theta \in \Theta} \mathcal{L}(\theta).$$

The proofs of Theorems 1.6 and 1.7 rely on the Girsanov theorem. Applied to a diffusion process it essentially states that if the drift of a given diffusion process is changed, then the law of the process does not change dramatically. In detail, the law of the new process is essentially states that if the drift of a given diffusion process is changed, then the law of the process does not change dramatically. In detail, the law of the new process is

$$Q$$

is well-defined, whereby $(B_t)_{0 \leq t \leq T}$ is standard Brownian motion. If $(Z_t)_{0 \leq t \leq T}$ is a martingale, then there exists a unique measure $Q$ on $C[0,T]$ such that $Q \ll P_B$ and $Z_t$ is a version of its Radon-Nikodym derivative when restricted to $\mathcal{F}_t$, i.e.

$$\frac{dQ}{dP_0}|_{\mathcal{F}_t} = Z_t \quad P_B - \text{almost surely}.$$
Further, the process \( \left( B_t - \int_0^t X_s \, ds \right) \) \( 0 \leq t \leq T \) is a standard Brownian motion under the measure \( Q \).

For a diffusion of the form (1.12), the Girsanov theorem is applied to the drift \((a(t, X_t, \theta))_{0 \leq t \leq T}\) in order to prove Theorem 1.6.

Note that an equivalent condition for \((Z_t)_{0 \leq t \leq T}\) to be a martingale is the requirement \( E(Z_t) = 1 \) for all \( t \in [0, T] \). This condition is usually hard to verify such that it is often replaced by the stronger condition

\[
E \left( \exp \left( \frac{1}{2} \int_0^T X_t^2 \, dt \right) \right) < \infty
\]

which is easier to verify in many cases. It is known as Novikov condition in literature.

### 1.4 Linear Operators on Hilbert Spaces

This section collects some material on integral operators which belong to the class of compact, symmetric and linear operators on Hilbert spaces. For the scope of this work it is not necessary to consider linear operators on the most general level. The notions and results that are explained here are based on Jörgens [26] and Taylor [41] and are used in the proof of the asymptotic normality of the maximum likelihood estimator in Section 3.4.

Denote by \( L^2[a, b] \) the Hilbert space of all Borel-measurable functions \( f : [a, b] \to \mathbb{R} \), where \([a, b]\) is an interval, such that \( \int_a^b f^2(t) \, dt < \infty \). The inner product \((\cdot, \cdot)\) of \( f, g \in L^2[a, b] \) is defined by

\[
(f, g) := \int_a^b f(t)g(t) \, dt
\]

and the norm \( \| \cdot \| \) of \( f \in L^2[a, b] \) as

\[
\|f\| := \left( \int_a^b f^2(t) \, dt \right)^{1/2}.
\]

The Cauchy-Schwarz inequality states that \( |(f, g)| \leq \|f\| \|g\| \) for \( f, g \in L^2[a, b] \).

A sequence \((f_n)_{n \geq 1}\) in \( L^2[a, b] \), i.e. \( f_n \in L^2[a, b] \) for all \( n \), converges to \( f \in L^2[a, b] \), in symbols \( f_n \to f \), if \( \|f_n - f\| \to 0 \) as \( n \to \infty \). The sequence \((f_n)_{n \geq 1}\) is called bounded if \( \|f_n\| \leq K < \infty \) for all \( n \).

**Definition 1.31.** A map \( T : L^2[a, b] \to L^2[a, b] \) is referred to as linear operator if \( T(\alpha f + \beta g) = \alpha T(f) + \beta T(g) \) for all \( f, g \in L^2[a, b] \) and \( \alpha, \beta \in \mathbb{R} \). Write \( Tf := T(f) \).

A linear operator \( T \) on \( L^2[a, b] \) is said to be

- **continuous** if \( f_n \to f \) implies \( T(f_n) \to T(f) \) for every convergent sequence \((f_n)_{n \geq 1}\) in \( L^2[a, b] \);
- **compact** if for every bounded sequence \((f_n)_{n \geq 1}\) in \( L^2[a, b] \) the sequence \((Tf_n)_{n \geq 1}\) has a convergent subsequent;
- **self-adjoint** or symmetric if \( (Tf, g) = (f, Tg) \) for all \( f, g \in L^2[a, b] \);
• **bounded** if $\|T\| := \sup_{f \in L^2[a,b], \|f\| = 1} \|Tf\| < \infty$.

**Proposition 1.32.** The linear operator $T$ on $L^2[a,b]$ is continuous if and only if it is bounded.

It can be readily shown by means of the previous proposition that a compact linear operator is continuous.

A real number $\lambda$ is called eigenvalue of the linear operator $T$ on $L^2[a,b]$ if $Te = \lambda e$ for some non-zero $e \in L^2[a,b]$. Any $e$ satisfying this equation is referred to as eigenfunction associated with the eigenvalue $\lambda$.

**Proposition 1.33.** The eigenfunctions associated with distinct eigenvalues of the linear symmetric operator $T$ are orthogonal, i.e. $Te = \lambda e$ and $Tf = \tau f$ with $\lambda \neq \tau$ implies $(e, f) = 0$. In addition, if $T$ is compact, then the set of all eigenvalues is countable.

**Proposition 1.34.** Let $\lambda_1, \lambda_2, \ldots$ be the non-zero eigenvalues of the compact and self-adjoint linear operator $T$ on $L^2[a,b]$. Then

$$\max_i |\lambda_i| = \|T\|.$$  

**Integral Operators**

Now integral operators on $L^2[a,b]$ where $[a,b]$ is a finite interval are introduced. The interest in integral operators was originally one of the impulses to develop functional analysis.

Define the integral operator $T_K : L^2[a,b] \to L^2[a,b]$ by

$$T_K f := \int_a^b K(s,t)f(t)\,dt, \quad f \in L^2[a,b],$$  \hspace{1cm} (1.13)

for some function $K : [a,b] \times [a,b] \to \mathbb{R}$ satisfying $\int_a^b \int_a^b K(s,t)\,ds\,dt < \infty$. If $K$ is continuous in both variables the last integrability is immediate. The function $K$ is usually referred to as kernel of the operator $T_K$.

Obviously, the operator $T_K$ is linear. It is well-known that $T_K$ defined by (1.13) is in fact a compact, and thus continuous, operator. If the kernel $K$ is symmetric, i.e. $K(s,t) = K(t,s)$ for all $s, t$, then the operator defined by (1.13) is self-adjoint.

Note that, if $e$ is an eigenfunction associated with the eigenvalue $\lambda$, then the eigenfunction satisfies the relation $\int_a^b K(s,t)e(t)\,dt = \lambda e(s)$. Consequently, continuity of the kernel implies continuity of the eigenfunction.

**Definition 1.35.** A continuous kernel $K : [a,b] \times [a,b] \to \mathbb{R}$ is said to be positive semidefinite if

$$\sum_{i=1}^n \sum_{j=1}^n K(t_i,t_j)c_ic_j \geq 0$$

for all possible choices $t_1, \ldots, t_n$ in $[a,b]$ and all possible real-numbers $c_1, \ldots, c_n$. The kernel $K$ is called positive definite if $\geq$ in the condition above is replaced by $>$.  

**Proposition 1.36.** If $K$ is continuous, symmetric and positive semidefinite then all possible real-valued eigenvalues of the operator (1.13) are non-negative.
The next theorem is the so-called Mercer’s Theorem which is an important tool in the theory of integral equations and of stochastic processes on Hilbert spaces where Mercer’s Theorem is used to prove the Karhunen-Loève theorem which provides a representation of stochastic process as a linear combination of orthogonal functions.

Mercer’s Theorem characterizes a kernel as a bilinear expansion in terms of its eigenvalues and eigenfunctions:

**Theorem 1.9 (Mercer’s Theorem).** Let $K$ be a continuous, symmetric and positive semidefinite kernel on $[a, b]^2$. Denote by $e_i$ the eigenfunction associated with the eigenvalue $\lambda_i$ of the operator (1.13). Then it holds that

$$K(s, t) = \sum_{i=1}^{\infty} \lambda_i e_i(s)e_i(t)$$

where the convergence is absolute and uniform on $[a, b]^2$. 

Chapter 2

Time-Discrete Observations

The present chapter may be viewed as an introduction to the large sample theory in the classical framework of time-discrete observations of a stochastic process. Two estimation methods for the popular mean-reverting Ornstein-Uhlenbeck process which is a simple diffusion are studied. Thereby, the explicit representation of the time-continuous process is used to obtain a time-discrete version of the process. Consistency and asymptotic normality of the estimators are obtained as the number of observations \( n \) tends to infinity whereby the discretization acuteness \( \Delta t \) is considered to be fixed. The proofs rely on classical techniques from the theory of time series and Markov processes. At the end of the chapter, the bias of the least squares estimator applied to the discretized stochastic differential equation is briefly discussed.

2.1 The Classical Ornstein-Uhlenbeck Process

The classical (arithmetic) mean reversion process, also known as the Ornstein-Uhlenbeck process, see Ornstein and Uhlenbeck [36], is defined as the solution of the stochastic differential equation

\[
    dX_t = \alpha(L - X_t)dt + \sigma dB_t, \quad t \geq 0, \quad X_0 = \xi, \quad (2.1)
\]

for some positive constants \( \alpha, L \) and \( \sigma \) and standard Brownian motion \((B_t)_{t \geq 0}\).

In this model, the process \((X_t)_{t \geq 0}\) fluctuates randomly but tends to revert to the equilibrium level \( L \). The so-called mean reversion rate \( \alpha \) determines the strength with which the long-term mean \( L \) attracts the process \( X_t \).

In this simple model, an explicit solution of the stochastic differential equation can be easily obtained by means of the Itô formula:

**Proposition 2.1.** The solution of the stochastic differential equation (2.1) can be explicitly represented as

\[
    X_t = X_0 e^{-\alpha t} + L(1 - e^{-\alpha t}) + \sigma \int_0^t e^{\alpha(s-t)} dB_s. \quad (2.2)
\]

**Proof.** It holds by the Itô lemma for \( Y_t = g(t, X_t) \) that

\[
    dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2
\]

which reduces for \( g(t, x) = e^{\alpha t}x \) to

\[
    dY_t = \alpha e^{\alpha t} X_t dt + e^{\alpha t} dX_t.
\]
Making then use of (2.1) yields
\[ dY_t = e^{\alpha t}(\alpha L dt + \sigma dB_t) \]
such that integrating from 0 to \( t \) and multiplying by \( e^{-\alpha t} \) provides (2.2).

Due to the deterministic integrand in (2.2) this Itô integral is normally distributed.

**Proposition 2.2.** The conditional distribution of \( X_t \) given \( X_0 \) from representation (2.2) is normal. It holds that
\[ E(X_t|X_0) = X_0 e^{-\alpha t} - L(1 - e^{-\alpha t}) \]
and
\[ \text{Var}(X_t|X_0) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}). \]

**Proof.** The Itô integral in (2.2) is normally distributed with mean zero and and variance
\[
E \left( \left( \sigma \int_0^t e^{\alpha(s-t)} dB_s \right)^2 \right) = E \left( \sigma^2 \int_0^t e^{2\alpha(s-t)} ds \right) = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t})
\]
whereby the Itô isometry is used.

**Remark 2.3.** Another popular model, the geometric mean reversion model, aims at resembling the famous geometric Brownian motion while introducing mean reversion to a long-term value \( \theta \) in the drift term: it is defined by
\[ \frac{dX_t}{X_t} = \alpha(\theta - \ln X_t) dt + \sigma dB_t, \quad t \geq 0, \ X_0 = \xi, \quad (2.3) \]
where \( \alpha \) is again the strength of mean reversion. This process is intuitively similar to the process defined in (2.1). Indeed, if the variables are changed, say \( Z_t := \ln X_t \), then again the Itô formula gives
\[ dZ_t = \alpha \left( \theta - \frac{\sigma^2}{2\alpha} - Z_t \right) dt + \sigma dB_t. \]
A comparison of this representation to (2.1) makes clear that in the geometric mean reversion model the logarithmic process follows an arithmetic mean reversion model where \( \theta - \frac{\sigma^2}{2\alpha} \) is the mean reversion level of the logarithms. Thus, \( X_t \) conditional on \( X_0 \) follows a log-normal distribution for each \( t \).

In the following sections the aim is to derive and analyze estimators of the unknown parameters \( \alpha, L \) and \( \sigma \) of the process defined in (2.1) based on discrete time realizations \( X_{t_1}, \ldots, X_{t_n} \) observed on an equidistant time mesh with refinement \( \Delta t = t_i - t_{i-1} \). A common estimation methodology in diffusion models is based on the discretization of the time-continuous process. Thereby, Euler’s discretization is applied to the stochastic differential equation converting the initial time-continuous form of the stochastic differential equation (2.2) into the discrete difference equation
\[ \Delta X_{t_i} = \alpha_0 + \alpha_1 X_{t_{i-1}} + \sigma \epsilon_{t_i}, \quad i = 1, 2 \ldots, \quad (2.4) \]
where $\Delta X_t = X_t - X_{t-1}$, $\alpha_0 = \alpha L \Delta t$ and $\alpha_1 = -\alpha \Delta t$.

Equation (2.4) induces the consideration of given observations $X_{t_1}, \ldots, X_{t_n}$ of the process defined in (2.2) as observations of a linear relationship between $\Delta X_t$ and $X_{t-1}$ in the presence of a normally distributed error term $\sigma \varepsilon_t$. Hence, one might regress historical observations of $\Delta X_t$ against $X_{t-1}$ so that estimates of $\alpha_0$ and $\alpha_1$ are obtained as the estimates of the intercept and slope of the linear dependence. Since $\Delta t$ is known, the estimates of $\alpha$ and $L$ can be easily computed. The parameter $\sigma$ is then estimated by the empirical variance of the residuals.

The drawback of the method described above is that the accuracy of the discretization in (2.4) requires very small $\Delta t$. So in practice the problem may arise that this discrete time format of the process introduces discretization errors into the estimation, see Section 2.5 for more details.

### 2.2 Method of Moments: Yule-Walker Estimator

The parameter estimation for the Ornstein-Uhlenbeck process may be based on the exact, i.e. valid for large $\Delta t$, discrete time expression of the mean-reverting process which can be obtained by means of the explicit solution given in Proposition 2.1. Then data on a set of discrete time points may be interpreted as realizations of an autoregressive process. The Yule-Walker estimation which is based upon the idea of the classical method of moments is proposed. Properties of the well-studied autoregressive process are used to obtain asymptotic results for the estimators.

The same procedure as in the proof of Proposition 2.1 where the last integration has to be done on an interval $[s, t]$ instead of $[0, t]$ yields the following corollary.

**Corollary 2.4.** The solution of the stochastic differential equation (2.1) can be represented for $0 \leq s < t$ as

$$X_t = X_s e^{-\alpha (t-s)} + L (1 - e^{-\alpha (t-s)}) + \sigma \int_s^t e^{\alpha (u-t)} dB_u.$$  

Note that this representation is valid for arbitrary time points $0 \leq s < t$. Equivalently, the process can be written for two time points $t_{i-1} < t_i$ as

$$X_{t_i} = X_{t_{i-1}} e^{-\alpha \Delta t} + L (1 - e^{-\alpha \Delta t}) + Z_{t_i},$$

where $Z_{t_i}$ is normally distributed with zero-mean and $\text{Var}(Z_{t_i}) = \frac{\sigma^2}{2} (1 - e^{-2\alpha \Delta t}) = \sigma^2 Z$.

$$\Delta t = t_i - t_{i-1}, \text{ see Proposition 2.2.}$$

Due to this representation, a given time-discrete set of observations $X_{t_1}, \ldots, X_{t_n}$ on a mesh of constant time steps $\Delta t = t_{i+1} - t_i$ may be viewed as a sample of the process satisfying the relation

$$X_i = \phi_0 + \phi_1 X_{i-1} + Z_i$$  

(2.5)

by making use of the simplifying notation $X_i = X_{t_i}$, where $\phi_0 = L (1 - e^{-\alpha \Delta t})$, $\phi_1 = e^{-\alpha \Delta t}$ and $Z_i$ is normally distributed with zero-mean and $\text{Var}(Z_i) = \sigma^2 Z$.

Representation (2.5) can be understood as a first-order autoregressive process, AR(1) process for short, which is probably the most popular model in time series analysis. Hence, the goal of this section is the study of an estimation method for AR(1) processes which can be used for the estimation of the mean reversion parameters.
Definition 2.5. Let \((\varepsilon_i)_{i \in \mathbb{Z}}\) be a white noise, that is \(E(\varepsilon_i) = 0, \ Var(\varepsilon_i) = \delta^2\) and \(\text{Cov}(\varepsilon_i, \varepsilon_{i+h}) = 0\) for all \(i, h \in \mathbb{Z}\). Then the (weakly) stationary process \((Y_i)_{i \in \mathbb{Z}}\) is said to be an autoregressive process of order 1 if

\[
Y_i = \phi_0 + \phi_1 Y_{i-1} + \varepsilon_i, \quad i \in \mathbb{Z}. \tag{2.6}
\]

If \(|\phi_1| < 1\) then the AR(1) process is called causal and it can be shown in that case that the process satisfying the autoregressive relation (2.6) is weakly stationary: By iterating this relation one obtains

\[
Y_i = \frac{\phi_0}{1 - \phi_1} + \sum_{k=0}^{\infty} \phi_1^k \varepsilon_{i-k} \tag{2.7}
\]

where the series converges absolutely with probability one, see Proposition 3.1.1 (p. 83) in Brockwell and Davis [11]. Consequently, it can be seen that

\[
E(Y_i) = \frac{\phi_0}{1 - \phi_1} \tag{2.8}
\]

and

\[
\text{Cov}(Y_i, Y_{i+h}) = \lim_{n \to \infty} E \left( \left( \sum_{k=0}^{n} \phi_1^k \varepsilon_{i-k} \right) \left( \sum_{k=0}^{n} \phi_1^k \varepsilon_{i+h-k} \right) \right) = \delta^2 \phi_1^{|h|} \sum_{k=0}^{\infty} \phi_1^{2k} = \phi_1^{|h|} \frac{\delta^2}{1 - \phi_1^2}.
\]

Now an estimator of the parameters which is based on the so-called Yule-Walker equations of an autoregressive model is introduced. Therefore, a causal and weakly stationary AR(1) process is considered which satisfies equation (2.6). Adjusting the process by subtraction of the mean gives

\[
Y_i^* = \phi_1 Y_{i-1}^* + \varepsilon_i
\]

where \(Y_i^* = Y_i - E(Y_i)\). Multiplying both sides by \(Y_{i-m}^*\) and taking expectations yields the Yule-Walker equations for \(m \geq 1\)

\[
\gamma(m) = \phi_1 \gamma(m-1)
\]

where

\[
\gamma(m) = \text{Cov}(Y_i^*, Y_{i-m}^*) = E(Y_i^* Y_{i-m}^*) = \text{Cov}(Y_i, Y_{i-m}).
\]

For \(m = 0\) the white noise term is not independent of \(Y_i\) so that

\[
E(\varepsilon_i Y_i) = E(\varepsilon_i^2) = \delta^2
\]

implying the relation

\[
\gamma(0) = \phi_1 \gamma(1) + \delta^2.
\]

Hence, the Yule-Walker equations provide for \(m = 0, 1\) the following representations of the unknown parameters of the AR(1) process:

\[
\phi_1 = \frac{\gamma(1)}{\gamma(0)}, \quad \delta^2 = \gamma(0) - \phi_1 \gamma(1). \tag{2.9}
\]


2.2. METHOD OF MOMENTS: YULE-WALKER ESTIMATOR

The essential idea is to make use of the generalized method of moments: Take the sample covariance function of observations $Y_1^*, \ldots, Y_n^*$ which is generally defined as

$$
\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-|h|} Y_j^* Y_{j+|h|}^*, \quad h \in \mathbb{N}_0, h < n,
$$

and replace the covariance $\gamma(1)$ and the variance $\gamma(0)$ in (2.9) by the sample counterparts $\hat{\gamma}(1)$ and $\hat{\gamma}(0)$ yielding the Yule-Walker estimators $\hat{\phi}_1$ and $\hat{\delta}^2$ given by

$$
\hat{\phi}_1 = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}, \quad \delta^2 = \hat{\gamma}(0) - \hat{\phi}_1 \hat{\gamma}(1). \tag{2.10}
$$

The mean $E(Y_i)$ may be estimated by the sample mean $\bar{Y} = \frac{1}{n} \sum_{k=1}^{n} Y_k$ such that, by making use of (2.8), the Yule-Walker estimator of $\phi_0$ is defined as

$$
\hat{\phi}_0 = \bar{Y} (1 - \hat{\phi}_1).
$$

**Proposition 2.6.** Let $Y_1, \ldots, Y_n$ be realizations of a causal AR(1) process as in Definition 2.5 provided with a normally distributed white noise $(\epsilon_i)_{i \in \mathbb{Z}}$, i.e. $\epsilon_i \sim N(0, \delta^2)$ for all $i$. Then the Yule-Walker estimator $\hat{\theta}_n$ of $\theta = (\phi_0, \phi_1, \delta^2)^t$ introduced above is consistent, i.e.

$$
\hat{\theta}_n \rightarrow \theta, \text{ almost surely,}
$$

as $n \rightarrow \infty$.

**Proof.** Due to representation (2.7) the process $(Y_i)_{i \in \mathbb{Z}}$ is ergodic as it can be represented as a linear combination of an i.i.d. sequence for each $i$. Further, the normality assumption on the innovations $(\epsilon_i)_{i \in \mathbb{Z}}$ justifies that the process $(Y_i)_{i \in \mathbb{Z}}$ is Gaussian and thus strictly stationary since it is weakly stationary by construction. The ergodic theorem implies that $\frac{1}{n} \sum_{k=1}^{n} Y_k \rightarrow E(Y_0)$ as $n \rightarrow \infty$, almost surely. Obviously, the process $(Y_i^*)_{i \in \mathbb{Z}}$, where $Y_i^* = Y_i - E(Y_0)$, is strongly stationary and ergodic, too. So is $(Y_i^* Y_{i+h})_{i \in \mathbb{Z}}, h \geq 0$, because $Y_i Y_{i+h} = g^h(Y_i, Y_{i+1}, \ldots)$ where $g^h$ is a measurable function for each $h$. It follows again by the ergodic theorem that $\frac{1}{n} \sum_{k=1}^{n} Y_k^* Y_{k+h}^* \rightarrow E(Y_0^* Y_0^*)$ as $n \rightarrow \infty$, almost surely, and thus $\hat{\gamma}(h) \rightarrow \gamma(h)$. Since $\hat{\phi}_1 = e(\hat{\gamma}(0), \hat{\gamma}(1))$, $\hat{\delta}^2 = f(\hat{\gamma}(0), \hat{\gamma}(1))$ and $\hat{\phi}_0 = g(\bar{Y}, \hat{\gamma}(0), \hat{\gamma}(1))$ where $e$, $f$ and $g$ are continuous functions, the assertion is immediate.

The asymptotic normality of the Yule-Walker estimator is given in the next proposition.

**Proposition 2.7.** In the same setting as in Proposition 2.6 the Yule-Walker estimators of the single parameters are asymptotically normal, i.e.

$$
\sqrt{n} (\hat{\phi}_1 - \phi_1) \overset{D}{\rightarrow} N(0, 1 - \phi_1^2),
$$

$$
\sqrt{n} (\bar{Y} - \mu) \overset{D}{\rightarrow} N(0, \tau)
$$

where $\mu = E(Y_0)$, $\tau = \frac{\delta^2}{(1 - \phi_1^2)}$ and

$$
\sqrt{n} (\hat{\delta}^2 - \delta^2) \overset{D}{\rightarrow} N(0, \rho)
$$

where $\rho = \frac{2\delta^2}{(1 - \phi_1^2)^2} (1 + \phi_1^2)^3 + 2\phi_1^2 + 8\phi_1^4 + 2\phi_1^6)$, as $n \rightarrow \infty$, respectively.
Proof. The first convergence stated above is a direct application of Theorem 8.1.1 in Brockwell and Davis [11] (p. 240) which gives the asymptotic distribution of the Yule-Walker estimator for an AR(p) process. By Theorem 7.1.2 in that work (p. 219) it holds that

$$\sqrt{n} (\bar{Y} - \mu) \xrightarrow{D} N(0, \nu)$$

where \( \nu = \sum_{h=0}^{\infty} \gamma(h) \). Here, it is \( \gamma(h) = \phi_1^{|h|} \frac{\delta^2}{1 - \phi_1^2} \) such that the expression for \( \tau \) follows. Further, Proposition 7.3.4 in Brockwell and Davis [11] (p. 229) states that

$$\sqrt{n} \left( \left( \hat{\gamma}(0) \right) - \left( \gamma(0) \right) \right) \xrightarrow{D} N(0, W')$$

where the \((2 \times 2)\)-matrix is given by

$$W = \begin{pmatrix} (\nu - 3)\gamma(l)\gamma(m) + \sum_{k=-\infty}^{\infty} (\gamma(k)\gamma(k - l + m) + \gamma(k + m)\gamma(k - l)) \\ l,m=0,1 \end{pmatrix}$$

whereby \( E(\varepsilon_i^4) = \nu \delta^4 < \infty \) is required. Due to the normality of the innovations it holds that \( \nu = 3 \). Note that in Proposition 7.3.4 in [11] the assertion (2.11) holds for processes of the form \( Y_i = \sum_{k=-\infty}^{\infty} \varepsilon_i - k \). However, it can be seen by means of Proposition 7.3.1 in [11] (p. 226) and by its proof that the convergence (2.11) with the given representation of \( W \) is valid for the process considered here. Making use of \( \gamma(h) = \phi_1^{|h|} \frac{\delta^2}{1 - \phi_1^2} \) implies that

$$W = \frac{\delta^2}{(1 - \phi_1^2)^2} \begin{pmatrix} 2(1 + \phi_1^2) & 4\phi_1 \\ 4\phi_1 & 1 + 4\phi_1^2 + \phi_1^4 \end{pmatrix}.$$ 

Taking into account the relation

$$\delta^2 = \hat{\gamma}(0) - \hat{\phi}_1 \hat{\gamma}(1) = \hat{\gamma}(0) - \frac{\hat{\gamma}(1)^2}{\hat{\gamma}(0)}$$

which can be written as

$$\frac{\delta^2}{2} = g(\hat{\gamma}(0), \hat{\gamma}(1))$$

where \( g: \mathbb{R}^2 \to \mathbb{R} \) is defined as \( g(x, y) = x - \frac{y^2}{2} \) with partial derivatives \( \frac{\partial g}{\partial x}(x, y) = 1 + \frac{y^2}{2x} \) and \( \frac{\partial g}{\partial y}(x, y) = -\frac{2y}{2} \) that are continuous in a neighborhood of \( (\gamma(0), \gamma(1)) \) results in

$$\sqrt{n}(\delta^2 - \delta^2) \xrightarrow{D} N(0, DW'D)$$

whereby

$$D = \begin{pmatrix} \frac{\partial g}{\partial x}(\gamma(0), \gamma(1)) & \frac{\partial g}{\partial y}(\gamma(0), \gamma(1)) \end{pmatrix}.$$

It can be computed that \( DW'D = \rho \).

Next, the previous results on the asymptotics of the Yule-Walker estimator for an AR(1) process are translated into the case of the mean reversion process that is the model of interest. Due to representation (2.5) the discretely sampled mean-reverting process can be interpreted as an AR(1) process. The only discrepancy between the discretized mean reversion process
and the AR(1) process is the fact that the latter is defined on \( Z \) which is necessary to obtain ergodicity and stationary. Otherwise, if the AR(1) process was defined on \( \mathbb{N} \) one would have to make assumptions on the distribution of the initial state \( X_0 \) in order to obtain stationarity and ergodicity.

In order to solve this problem and to apply the results for the AR(1) process it will be shown in the following that the asymptotic theory (as \( n \to \infty \)) of the process (2.6) defined for all \( i \in \mathbb{Z} \) is the the same as of a process that is defined on \( i = 0, 1, \ldots \) and that obeys (2.6) for \( i = 0, 1, \ldots \) with a random or constant initial state \( X_0 = \xi \).

**Lemma 2.8.** Let \((Y_i)_{i \in \mathbb{Z}}\) be a causal AR(1) process as defined in Definition 2.5. Let \((\tilde{Y}_i)_{i \in \mathbb{N}}\) be defined analogously by the relation

\[
\tilde{Y}_i = \phi_0 + \phi_1 \tilde{Y}_{i-1} + \varepsilon_i, \quad i \in \mathbb{N}, \quad \tilde{Y}_0 = \xi, \tag{2.12}
\]

where the parameter \(\phi_0, \phi_1, \delta^2\) and the innovations are the same as in \((Y_i)_{i \in \mathbb{Z}}\). Let \(\xi\) be independent of the innovations and satisfy \(E(\xi^2) < \infty\). Then it holds that

\[
|Y_k - \tilde{Y}_k| \to 0, \quad \text{almost surely, as } k \to \infty.
\]

*Proof.* Iterating relation (2.12) yields

\[
\tilde{Y}_k = \phi_0 \frac{\phi_1^k - 1}{\phi_1 - 1} + \sum_{j=0}^{k-1} \phi_1^j \varepsilon_{k-j} + \phi_1^k \xi
\]

such that in combination with (2.7) it is

\[
\limsup_{k \to \infty} |Y_k - \tilde{Y}_k| \leq \frac{1}{1 - \phi_1} |\phi_0 \phi_1^k| + |\phi_1^k \xi| + |\phi_1^k \sum_{j=1}^{\infty} \phi_1^j \varepsilon_j| = 0
\]

since the series converges almost surely by Proposition 3.1.1 (p. 83) in Brockwell and Davis [11] and \(|\phi_1| < 1\).

The asymptotic equivalence of the second-order sample moments from both processes is proved in the following corollary.

**Corollary 2.9.** In the same setting as in Lemma 2.8 it holds that

\[
\left| \frac{1}{n} \sum_{k=1}^{n} Y_k - \frac{1}{n} \sum_{k=1}^{n} \tilde{Y}_k \right| \to 0
\]

and

\[
\left| \frac{1}{n} \sum_{k=1}^{n} Y_k Y_{k+h} - \frac{1}{n} \sum_{k=1}^{n} \tilde{Y}_k \tilde{Y}_{k+h} \right| \to 0
\]

, almost surely as, \( n \to \infty \), respectively.
Proof. The first assertion follows directly from Lemma 2.8. For the second one, note that
\[
\frac{1}{n} \sum_{k=1}^{n} \left( \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{k-j} \right) \rightarrow E \left( \sum_{j=0}^{\infty} \phi_1^j \varepsilon_j \right) = 0 \quad \text{a.s.}
\]
by the ergodic theorem since \( \left( \sum_{j=0}^{\infty} \phi_1^j \varepsilon_{k-j} \right)_{k \in \mathbb{N}} \) is stationary and ergodic. Moreover, it is
\[
\frac{1}{n} \sum_{k=1}^{n} \left( \sum_{j=0}^{k-1} \phi_1^j \varepsilon_{k-j} - \sum_{i=0}^{k-1} \phi_1^i \varepsilon_{k-i} \right) = \frac{1}{n} \sum_{k=1}^{n} \left( \phi_0 \phi_1^k - 1 + \sum_{j=0}^{k-1} \phi_1^j \varepsilon_{k-j} + \phi_1^k \xi_k \right) \rightarrow 0 \quad \text{a.s.}
\]
such that
\[
\limsup_{k \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = \limsup_{k \to \infty} \frac{1}{n} \sum_{k=1}^{n} \left( \phi_0 \phi_1^k - 1 + \sum_{j=0}^{k-1} \phi_1^j \varepsilon_{k-j} + \phi_1^k \xi_k \right) < \infty.
\]
This fact, the ergodic theorem and the relations
\[
Y_k Y_{k+h} - \tilde{Y}_k \tilde{Y}_{k+h} = (Y_k - \tilde{Y}_k)(Y_{k+h} + \tilde{Y}_{k+h}) - Y_k \tilde{Y}_{k+h} + \tilde{Y}_k Y_{k+h}
\]
and
\[
Y_k \tilde{Y}_{k+h} = (Y_k - \tilde{Y}_k)\tilde{Y}_k + \tilde{Y}_k \tilde{Y}_{k+h}
\]
combined with Lemma 2.8 justify the convergence of \( \frac{1}{n} \sum_{k=1}^{n} (Y_k Y_{k+h} - \tilde{Y}_k \tilde{Y}_{k+h}) \) toward zero. \( \square \)

By the previous corollary, the asymptotic properties of the sample mean and the Yule-Walker estimators based on a discrete sample \( X_{t_1}, \ldots, X_{t_n} \) of the mean reversion process defined in (2.1), \( \Delta t = t_i - t_{i-1} \) fixed, which exhibits an autoregressive structure, see (2.5), are equal to those given in Proposition 2.6 and 2.7. Hence, it is reasonable to estimate the parameters \( \alpha, L \) and \( \sigma \) from a time-discretely observed mean reversion process by means of the sample mean and the Yule-Walker estimates for the equivalent AR(1) process given in (2.5). Due to the equation \( \phi_1 = e^{-\alpha \Delta t} \) the mean reversion rate \( \alpha \) may be estimated by
\[
\hat{\alpha} = -\frac{\ln \hat{\phi}_1}{\Delta t} \quad \text{(2.13)}
\]
where \( \hat{\phi}_1 \) terms the Yule-Walker estimator. The relation \( \sigma_Z^2 = \delta^2 = \frac{\sigma^2}{2\alpha \Delta t} (1 - e^{-2\alpha \Delta t}) \) gives reason to estimate the diffusion coefficient \( \sigma \) by
\[
\hat{\sigma} = f(\hat{\delta}^2) \quad \text{(2.14)}
\]
where \( f : (0, \infty) \to (0, \infty) \) is a continuous mapping defined by \( f(x) = \left( \frac{2x}{1 - e^{-2x \Delta t}} \right)^{1/2} \cdot \sqrt{x} \) and where \( \hat{\delta}^2 \) denotes the Yule-Walker estimator. The representation \( \phi_0 = L(1 - e^{-\alpha \Delta t}) = L(1 - \phi_1) \) yields
\[
L = \frac{\phi_0}{1 - \phi_1} = E(Y_0)
\]
such that
\[
\hat{L} = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} X_{t_i}.
\]
Due to Corollary 2.9 the asymptotic distribution of the sample mean is obtained directly from Proposition 2.7, consistency follows by the ergodic theorem applied to the sample mean of the ergodic and stationary AR(1) process, compare the proof of Proposition 2.6.

**Corollary 2.10.** Let \( \hat{\alpha} \) and \( \hat{\sigma} \) be the estimators given in (2.13) and (2.14) computed from a time-discrete sample \( X_{t_1}, \ldots, X_{t_n}, \Delta t = t_i - t_{i-1} \) fixed, of the mean reversion process defined in (2.1) which may be written in discrete time as in (2.5). Then

\[
\hat{\alpha} \to \alpha, \quad \hat{\sigma} \to \sigma,
\]

almost surely, as \( n \to \infty \), respectively, and

\[
\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{D} N \left( 0, \frac{(e^{-2\alpha \Delta t} - 1)}{(\Delta t)^2} \right)
\]

and

\[
\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{D} N \left( 0, \frac{\sigma^2 \rho}{4(1 - e^{-2\alpha \Delta t})^2} \right)
\]

where \( \rho \) is given in Proposition 2.7.

**Proof.** These statements follow immediately from Proposition 2.6 and 2.7: it holds that \( \hat{\alpha} = h(\hat{\phi}_1) \) where \( h(x) = -\frac{\ln(x)}{\Delta t} \) is a continuous function which is differentiable at \( \phi_1 = e^{-\alpha \Delta t} \). The asymptotic variance is computed according to

\[
\left( \frac{\partial h}{\partial x}(\phi_1) \right)^2 (1 - \phi_1^2) = (\Delta t \phi_1)^{-2}(1 - \phi_1^2) = \frac{e^{-2\alpha \Delta t} - 1}{(\Delta t)^2}.
\]

An analogous argument can be applied to \( \hat{\sigma} = f(\hat{\delta}) \) where \( f : (0, \infty) \to (0, \infty) \) is a continuous mapping defined by \( f(x) = \left( \frac{2x}{1 - e^{-2\alpha \Delta t}} \right)^{1/2} \sqrt{x} \) and \( \frac{\partial f}{\partial x}(x) = \left( \frac{2x}{(1 - e^{-2\alpha \Delta t})^2} \right)^{1/2} x^{-1/2} \). Plugging in \( \delta^2 = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha \Delta t}) \) yields the asymptotic variance. \( \Box \)

## 2.3 Maximum Likelihood Estimation

The method of maximum likelihood is widespread in mathematical statistics because it provides in many cases estimators which exhibit under quite weak regularity conditions preferable characteristics, like consistency, asymptotic normality and asymptotic efficiency.

Consider time-discrete realizations \( X_n^{n+1} = \{X_0, X_{t_1}, \ldots, X_{t_n}\} \) from the diffusion model specified in (2.1). For notational simplicity, assume that the observations are equidistant such that \( \Delta t = t_i - t_{i-1} \) is constant and let \( X_i = X_{t_i}, i \geq 1 \). Due to Corollary 2.4 it holds that

\[
X_i = X_{i-1}e^{-\alpha \Delta t} + L(1 - e^{-\alpha \Delta t}) + \sigma \int_{t_{i-1}}^{t_i} e^{\alpha(u-t_i)} dB_u \tag{2.15}
\]

and it follows that the conditional distribution of \( X_i \), given \( X_{i-1} \), is normal with

\[
E(X_i|X_{i-1}) = X_{i-1}e^{-\alpha \Delta t} + L(1 - e^{-\alpha \Delta t})
\]

and

\[
\text{Var}(X_i|X_{i-1}) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha \Delta t}).
\]
Let \( \theta = (\alpha, L, \sigma)^t \in \Theta = \{(x, y, z) : x > 0, z > 0\} \) be the unknown parameter. Denote by \( P_\theta \) the measure on \( \mathbb{R}^{n+1} \) induced by the random vector \((X_0, X_1, \ldots, X_n)\) satisfying equation (2.15), that is
\[
P_\theta(A) = P((X_0, X_1, \ldots, X_n) \in A)
\]
for all Borel sets \( A \subset \mathbb{R}^{n+1} \). Further, let \( \mu \) be the Lebesgue measure on \( \mathbb{R} \) and \( \mu^{(n+1)} \) its \((n+1)\)-fold product measure. Then \( P_\theta \ll \mu^{(n+1)} \) and the Radon-Nikodym derivative \( \frac{dP_\theta}{d\mu^{(n+1)}} \) evaluated at \( x = (x_1, \ldots, x_{n+1})^t \in \mathbb{R}^{n+1} \) is given by
\[
\frac{dP_\theta}{d\mu^{(n+1)}}(x) = f_0(x_1, \theta) \prod_{i=1}^{n+1} f(x_i, x_{i-1}, \theta)
\]
where
\[
f(x_i, x_{i-1}, \theta) = \left(\frac{\pi \sigma^2}{\alpha} (1 - e^{-2\alpha \Delta t})\right)^{-\frac{1}{2}} \exp \left(-\frac{(x_i - e^{-\alpha \Delta t} x_{i-1} - L(1 - e^{-\alpha \Delta t}))^2}{\alpha^2 \sigma^2 (1 - e^{-2\alpha \Delta t})}\right)
\]
and where \( f_0(x_1, \theta) \) denotes the probability density of \( X_0 \) with respect to \( \mu \). Since the information on \( X_0 \) in \( X^{n+1} = \{X_0, X_1, \ldots, X_n\} \) does not increase as \( n \to \infty \), the distribution of \( X_0 \) does not play an important role in the large sample theory studied in the sequel. Hence, the log-likelihood ratio \( \mathcal{L} \) for a given sample \( X^{n+1} \) may be defined as the logarithm of the Radon-Nikodym derivative \( \frac{dP_\theta}{d\mu^{(n+1)}} \) divided by \( f_0(x_1, \theta) \) and evaluated at \( X^{n+1} \), namely
\[
\mathcal{L}(\theta) = \sum_{i=1}^{n} \ln (f(X_i, X_{i-1}, \theta)).
\]
The maximum likelihood estimator \( \hat{\theta} \) is defined as the maximum of \( \mathcal{L}(\theta) \) and might be obtained in practice as the solution to the system of equations
\[
\frac{\partial}{\partial \theta_i} \mathcal{L}(\hat{\theta}) = 0, \quad i = 1, 2, 3.
\]
There may arise the problem of several solutions to these equations, just like in the case of independent realizations. In such a case, further constraints have to be made. In the subsequent treatment, it is assumed that there exists one solution to the system in a small neighborhood of the true parameter \( \theta_0 \).

Note that the Ornstein-Uhlenbeck process defined in (2.1) is obviously a diffusion process and thus a Markov process. So is its time-discrete counterpart satisfying relation (2.15) which obviously demonstrates the Markov property. The Markov characteristic is used to obtain asymptotic results for the maximum likelihood estimator introduced above. The following investigation is based upon large sample results for homogeneous Markov processes obtained by Billingsley [6].

Let \( f_k = \frac{\partial}{\partial \theta_0} f \), \( f_{kl} = \frac{\partial^2}{\partial \theta_k \partial \theta_l} f \) etc. be the abbreviatory notation for the partial derivatives of the density \( f \) given in (2.16) and let \( g \) be defined as \( g(x, y, \theta) := \ln(f(x, y, \theta)) \). In order to apply the findings in [6], some regularity conditions which are analogous to those needed in the instance of independence are required. In a nutshell, these conditions include the continuity of the partial derivatives of \( f \) such that a Taylor expansion of \( g \) is justified. Furthermore, they insure the differentiability under the integral and the existence of a non-singular fisher information matrix. Note that these requirements hold true for the process given in (2.15), see next lemma.
Lemma 2.11. Let \((X_1, \ldots, X_n)\) be a random vector from the process (2.15) and \(f(x, y, \theta)\) the conditional density specified in (2.16). Then, the partial derivatives \(f_k(x, y, \theta)\), \(f_{kl}(x, y, \theta)\) and \(f_{klm}(x, y, \theta)\) exist and are continuous in \(\Theta\) for all \(x\) and \(y\). Further, there exists for all \(\theta' \in \Theta\) a neighborhood \(N(\theta')\) such that

\[
\int \sup_{\theta \in N(\theta')} |f_k(x, y, \theta)|dx < \infty, \quad \int \sup_{\theta \in N(\theta')} |f_{kl}(x, y, \theta)|dx < \infty
\]

and

\[
E_\theta \left( \sup_{\theta \in N(\theta')} |g_{klm}(X_2, X_1, \theta)| \right) < \infty.
\]

In addition, there exists some \(\eta > 0\) such that

\[
E_\theta \left( (g_k(X_2, X_1, \theta))^{2+\eta} \right) < \infty, \quad k = 1, 2, 3.
\]

Proof. The assertion relies on Example 1.1. in Billingsley [6] (p. 8) where these properties are shown to hold true for a first-order autoregressive process of the form

\[
Y_i = a + b Y_{i-1} + \sqrt{c} \varepsilon_i, \quad i \geq 1,
\]

where \(|b| < 1\), \(c > 0\) and each \(\varepsilon_i\) is standard normal. Note that the process in (2.15) is also an autoregressive process admitting the representation

\[
X_i = \phi_0(\theta_1, \theta_2) + \phi_1(\theta_1)X_{i-1} + \gamma^{1/2}(\theta_1, \theta_3)Z_i, \quad i \geq 1,
\]

where each \(Z_i\) is standard normal and the functions \(\phi_0, \phi_1\) and \(\gamma\) have continuous derivatives of all orders.

Note that the stationary transition measures given by

\[
P_\theta(B, z) = P_\theta(X_i \in B | X_{i-1} = z) = \int_B f(x, z, \theta)dx
\]

for all \(B \in \mathcal{B}\). In detail, \(Q_\theta\) is the normal distribution with mean \(L\) and variance \(\sigma^2/(2\alpha)\) and it holds that \(P_\theta(\cdot, z) \ll Q_\theta(\cdot)\) for all \(\theta\) and \(z\). This absolute continuity combined with Lemma 2.11 justifies the next lemma which corresponds to Theorem 1.1 and 1.2 in Billingsley [6] (p. 6). Therefor, it is essentially required that the \((3 \times 3)\)-matrix \(I(\theta) = (I_{kl}(\theta))_{1 \leq k, l \leq 3}\) defined by

\[
I_{kl}(\theta) = E_\theta \left( g_k(X_2, X_1, \theta)g_l(X_2, X_1, \theta) \right)
\]

is non-singular. This presumption excludes redundancy of the parameters \(\theta_1, \theta_2\) and \(\theta_3\).

The next lemma provides a law of large numbers and a central limit theorem for the Marlov process \((X_i)_{i \geq 1}\) given in (2.15). These results are essential for the proofs of consistency and asymptotic normality of the maximum likelihood estimator. Note that these proofs presented in the following are modeled on Billingsley’s [6] techniques. The line of arguments is similar to the corresponding results for the i.i.d. case.
Lemma 2.12. Under the framework of Lemma 2.11 and in the case of a non-singular fisher information matrix \(I(\theta_0)\), it holds for all \((B \times B)\)-measurable functions \(\varphi\) satisfying \(E_\theta(|\varphi(X_1, X_0)|) < \infty\) that

\[
\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i, X_{i-1}) \to E_\theta(\varphi(X_1, X_0)),
\]

almost surely, and

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g(X_i, X_{i-1}, \theta) \overset{D}{\to} N(0, I(\theta))
\]
as \(n \to \infty\), respectively. Thereby, \(\nabla g\) denotes the gradient of \(g\) with respect to \(\theta\), \(g = \ln f\).

Proposition 2.13. Under the condition of Lemma 2.12 there exists, with probability one, a consistent sequence of solutions \((\hat{\theta}_n)_{n \geq 1}\) to the system (2.17), that is

\[
\hat{\theta}_n \to \theta_0, \text{ almost surely,}
\]
as \(n \to \infty\), where \(\theta_0\) is the true value of the parameter.

Proof. For \(\theta\) in the neighborhood \(N(\theta_0)\) specified in Lemma 2.11 a Taylor expansion of \(g_k\) about the point \(\theta_0\) provides

\[
g_k(X_i, X_{i-1}, \theta) = g_k(X_i, X_{i-1}, \theta_0) + \sum_{j=1}^{3} (\theta_j - \theta_{0,j}) g_{kj}(X_i, X_{i-1}, \theta_0) \\
+ \frac{1}{2} \sum_{j,m=1}^{3} (\theta_j - \theta_{0,j})(\theta_m - \theta_{0,m}) g_{km}(X_i, X_{i-1}, \theta')
\]

where \(\theta'\) lies on the line that joins the points \(\theta\) and \(\theta_0\). Hence, it holds that

\[
g_k(X_i, X_{i-1}, \theta) \leq g_k(X_i, X_{i-1}, \theta_0) + \sum_{j=1}^{3} (\theta_j - \theta_{0,j}) g_{kj}(X_i, X_{i-1}, \theta_0) \\
+ \frac{1}{2} \sum_{j,m=1}^{3} |\theta_j - \theta_{0,j}| |\theta_m - \theta_{0,m}| \sup_{\theta \in N(\theta_0)} |g_{klm}(X_i, X_{i-1}, \theta)|
\]

\[
\leq g_k(X_i, X_{i-1}, \theta_0) + \sum_{j=1}^{3} (\theta_j - \theta_{0,j}) g_{kj}(X_i, X_{i-1}, \theta_0) + \frac{9}{2} \|\theta - \theta_0\|^2 H(X_i, X_{i-1})
\]

where \(H(X_i, X_{i-1}) = \sup_{l,m} \sup_{\theta \in N(\theta_0)} |g_{klm}(X_i, X_{i-1}, \theta)|\) and where \(\|\cdot\|\) denotes the Euclidean norm. These relations implicate the lower bound

\[
g_k(X_i, X_{i-1}, \theta) \geq g_k(X_i, X_{i-1}, \theta_0) + \sum_{j=1}^{3} (\theta_j - \theta_{0,j}) g_{kj}(X_i, X_{i-1}, \theta_0) - \frac{9}{2} \|\theta - \theta_0\|^2 H(X_i, X_{i-1}).
\]

Consequently, by the intermediate value theorem, there exists a \(\tau\) satisfying \(|\tau| < 9/2\) such that

\[
g_k(X_i, X_{i-1}, \theta) = g_k(X_i, X_{i-1}, \theta_0) + \sum_{j=1}^{d} (\theta_j - \theta_{0,j}) g_{kj}(X_i, X_{i-1}, \theta_0) + \tau \|\theta - \theta_0\|^2 H(X_i, X_{i-1}).
\]
By Lemma 2.11 it holds for the partial derivative \( f \) 
\[
\int f(x, y, \theta) dx = \frac{\partial}{\partial \theta_k} \int f(x, y, \theta) dx = 0
\]
and thus
\[
\int f_{kl}(x, y, \theta) dx = 0.
\]

Consequently
\[
E_\theta(g_k(X_i, X_{i-1}, \theta)|X_{i-1}) = \int g_k(x, X_{i-1}, \theta)f(x, X_{i-1}, \theta)dx
\]
\[
= \int \frac{1}{f(x, X_{i-1}, \theta)} \frac{\partial f(x, X_{i-1}, \theta)}{\partial \theta_k} f(x, X_{i-1}, \theta)dx
\]
\[
= 0.
\]
Hence, it is \( E_\theta(g_k(X_i, X_{i-1}; \theta)) = 0 \) such that Lemma 2.12 yields
\[
\frac{1}{n} \sum_{i=1}^{n} A_i(\theta_0) \rightarrow 0,
\]
almost surely. Furthermore, for the second partial derivative it can be seen that
\[
E_\theta(g_{kl}(X_i, X_{i-1}, \theta)|X_{i-1}) = \int \frac{\partial^2 \ln f(x, X_{i-1}, \theta)}{\partial \theta_k \partial \theta_l} f(x, X_{i-1}, \theta)dx
\]
\[
= \int \frac{\partial}{\partial \theta_k} \left( \frac{1}{f(x, X_{i-1}, \theta)} \frac{\partial \ln f(x, X_{i-1}, \theta)}{\partial \theta_l} \right) f(x, X_{i-1}, \theta)dx
\]
\[
= \int -\frac{1}{f^2(x, X_{i-1}, \theta)} \frac{\partial f(x, X_{i-1}, \theta)}{\partial \theta_k} \frac{\partial f(x, X_{i-1}, \theta)}{\partial \theta_l} f(x, X_{i-1}, \theta)dx
\]
\[
= -E_\theta(g_k(X_i, X_{i-1}; \theta)g_l(X_i, X_{i-1}; \theta)|X_{i-1})
\]
which, by taking expectation, results in
\[
E_\theta(g_{kl}(X_i, X_{i-1}, \theta)) = -E_\theta(g_k(X_i, X_{i-1}, \theta)g_l(X_i, X_{i-1}, \theta)) = -I_{kl}(\theta).
\]

Again, by Lemma 2.12, it can be concluded that
\[
\frac{1}{n} \sum_{i=1}^{n} B_{ij}(\theta_0) \rightarrow -I_{kj}(\theta_0),
\]
and
\[
\frac{1}{n} \sum_{i=1}^{n} C_i(\theta_0) \to E_\theta(H(X_1, X_0)),
\] (2.21)
almost surely, respectively.

It will be demonstrated in the sequel that for any arbitrary \( \varepsilon > 0 \) there exists an index \( n_\varepsilon \) such that for all \( n \geq n_\varepsilon \) there is with probability greater than \( 1 - \varepsilon \) a solution \( \hat{\theta} \) to the likelihood equations (2.17)
\[
\sum_{i=1}^{n} g_k(X_i, X_{i-1}, \hat{\theta}) = 0, \quad k = 1, 2, 3,
\]
and \( \|\hat{\theta} - \theta_0\| < \varepsilon \). In order to do so, fix some \( \varepsilon > 0 \) and choose \( 0 < \delta < \varepsilon \) such that \( \|\theta - \theta_0\| \leq \delta \) implies \( \theta \in \mathcal{N}(\theta_0) \) and further such that \( \delta < \frac{\varepsilon}{81(K+1)} \). Thereby, \( E_\theta(H(X_1, X_0)) =: K < \infty \) and \( \eta > 0 \) is defined by the relation \( x^t I(\theta_0) x \geq \eta \) for all \( x \in \mathbb{R}^3 \) with \( \|x\| = 1 \). Note that such a lower bound \( \eta \) exists because \( I(\theta_0) \) is assumed to be non-singular and thus positive-definite.

Now take \( n \geq n_\varepsilon \) large enough such that by the convergence properties stated in (2.19), (2.20) and (2.21) it holds with probability greater than \( 1 - \varepsilon \) that
\[
\left| \frac{1}{n} \sum_{i=1}^{n} A_i(\theta_0) \right| < \delta^2,
\]
\[
\left| \frac{1}{n} \sum_{i=1}^{n} B_{ij}(\theta_0) + I_{kj}(\theta_0) \right| < \delta
\]
and
\[
0 \leq -\frac{1}{n} \sum_{i=1}^{n} C_i(\theta_0) < K + 1.
\]
These bounds and representation (2.18) implicate for \( \theta \) with \( \|\theta - \theta_0\| \leq \delta \) that
\[
\left| \frac{1}{n} \frac{\partial}{\partial \theta_k} \mathcal{L}(\theta) + \sum_{j=1}^{3} (\theta_j - \theta_{0,j}) I_{kj}(\theta_0) \right| \leq \delta^2 + 3\delta \|\theta - \theta_0\| + 9 \|\theta - \theta_0\|^2(K + 1)/2
\]
\[
\leq 27\delta^2(K + 1).
\]
This inequality applied in both directions and multiplied by \( (\theta_l - \theta_{0,l}) \) yields
\[
\frac{1}{n} \frac{\partial}{\partial \theta_k} \mathcal{L}(\theta_l - \theta_{0,l}) \leq -\sum_{j=1}^{3} (\theta_j - \theta_{0,j}) I_{kj}(\theta_0)(\theta_l - \theta_{0,l}) + 27\delta^2(K + 1) |\theta_l - \theta_{0,l}|.
\] (2.22)
Denote by \( \nabla \mathcal{L} \) the gradient with respect to \( \theta \) and note that the conditions of Lemma 2.14 are fulfilled: Inequality (2.22) in combination with the fact that \( I(\theta_0) \) is positive-definite implicate for \( \theta \) with \( \|\theta - \theta_0\| = \delta < \frac{\varepsilon}{81(K+1)} \) that
\[
\frac{1}{n} (\theta - \theta_0)^t \nabla \mathcal{L}(\theta) \leq - (\theta - \theta_0)^t I(\theta_0)(\theta - \theta_0) + 81\delta^3(K + 1)
\]
\[
\leq -\eta \|\theta - \theta_0\|^2 + 81\delta^3(K + 1) = \delta^2 (81(K + 1)\delta - \eta) < 0.
\]
So, by Lemma 2.14 there exists some \( \hat{\theta} \) with \( \|\hat{\theta} - \theta_0\| \leq \delta < \varepsilon \) such that \( \frac{\partial}{\partial \theta_k} \mathcal{L}(\hat{\theta}) = 0 \) for \( k = 1, 2, 3 \). \( \square \)
2.3. MAXIMUM LIKELIHOOD ESTIMATION

The following lemma is a slightly generalized version of Lemma 2 in Aitchison and Silvey [1].

**Lemma 2.14.** Let \( h : \mathbb{R}^d \rightarrow \mathbb{R}^d \) be continuous. If \( x^t h(x) < 0 \) for all \( x \in \mathbb{R}^d \) with \( \|x\| = 1 \) then there exists an \( x' \in \mathbb{R}^d \) such that \( h(x') = 0 \) and \( \|x\| < 1 \).

**Proposition 2.15.** Under the condition of Lemma 2.12 the maximum likelihood estimator \( \hat{\theta}_n \) is asymptotically normal, in detail it is

\[
n^{1/2}(\hat{\theta}_n - \theta_0) \overset{D}{\rightarrow} N(0, I(\theta_0)^{-1})
\]
as \( n \rightarrow \infty \), where \( \theta_0 \) is the true value of the parameter.

**Proof.** Let \( \hat{\theta}_n \) be the maximum likelihood estimator such that

\[
\nabla \mathcal{L}(\hat{\theta}_n) = 0
\]

and \( \hat{\theta}_n \in N(\theta_0) \) with probability going to one as \( n \rightarrow \infty \), see Proposition 2.13. Taylor expansion for \( \hat{\theta}_n \), compare (2.18), yields

\[
0 = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta_k} \mathcal{L}(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i(\theta_0) + \frac{3}{n} \sum_{j=1}^{3} \sqrt{n}(\hat{\theta}_{n,j} - \theta_{0,j}) + \frac{1}{n} \sum_{i=1}^{n} B_{ij}(\theta_0)
\]

\[
+ \tau \|\hat{\theta}_n - \theta_0\| \cdot \|\sqrt{n}(\hat{\theta}_n - \theta_0)\| \cdot \frac{1}{n} \sum_{i=1}^{n} C_{i}(\theta_0)
\]

Since \( \|\hat{\theta}_n - \theta_0\| \rightarrow 0 \) by Proposition 2.13, it can be concluded from (2.20) and (2.21) that

\[
\|S_n - I(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)\| \leq \beta_n \|\sqrt{n}(\hat{\theta}_n - \theta_0)\|
\]

where \( S_n = (S_{n,1}, S_{n,2}, S_{n,3})^t \), \( S_{n,k} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_k(X_i, X_{i-1}, \theta_0) \) and where \( \beta_n \overset{P}{\rightarrow} 0 \). By Lemma 2.12, it is

\[
\frac{1}{\sqrt{n}} S_n \overset{D}{\rightarrow} N(0, I(\theta_0))
\]
such that the assertion follows from Lemma 2.16. \( \square \)

The next auxiliary result is well-known and may be proved quite easily. For a proof, see Theorem 10.1 in Billingsley [6] on p. 62.

**Lemma 2.16.** Let \( (Y_n)_{n \geq 1} \), \( (Z_n)_{n \geq 1} \) and \( Y \) be random vectors in \( \mathbb{R}^d \). If \( Y_n \overset{D}{\rightarrow} Y \) as \( n \rightarrow \infty \) and

\[
\|Y_n - Z_n\| \leq \beta_n \|Z_n\|
\]

where \( \beta_n \rightarrow 0 \) in probability, then \( Z_n \rightarrow Y \).
2.4 Maximum Likelihood in Practice: Numeric Search for Solutions

In this section an analytic approach which simplifies the numerical search for a solution of the first order conditions given in (2.17) is derived. The numerical problem is reduced to a one-dimensional optimization problem.

Let \( X^{n+1} = \{X_0, X_1, \ldots, X_n\} \) be observations at regularly space time points from the simple Ornstein-Uhlenbeck process satisfying (2.15). As introduced in the previous section, the log-maximum likelihood function is of the form

\[
L(\theta) = \sum_{i=1}^{n} \ln \left( f(X_i, X_{i-1}, \theta) \right)
\]

where the density \( f \) is given in (2.16). Recall that \( \theta = (\alpha, L, \sigma)^t \in \Theta = \{(x, y, z)^t : x > 0, z > 0\} \). The partial derivative of the log-likelihood function \( L \) with respect to \( L \) can be computed to be

\[
\frac{\partial}{\partial L} L(\theta) = 2\frac{\alpha}{\sigma^2} \sum_{i=1}^{n} \frac{X_i - e^{-\alpha \Delta t} X_{i-1} - L(1 - e^{-\alpha \Delta t})}{1 + e^{-\alpha \Delta t}}
\]

and the criterion (2.17) gives

\[
\hat{L} = \sum_{i=1}^{n} \frac{X_{i_1} - X_{i_{i-1}} e^{-\alpha \Delta t}}{1 + e^{-\alpha \Delta t}} \left( \frac{\sum_{i=1}^{n} 1 - e^{-\alpha \Delta t}}{1 + e^{-\alpha \Delta t}} \right)^{-1} =: s(\hat{\alpha}).
\]

That means that \( \hat{L} \) can be represented as a function \( s \) of the estimator \( \hat{\alpha} \). Moreover, setting the partial derivative with respect to \( \sigma \)

\[
\frac{\partial}{\partial \sigma} L(\theta) = -\frac{n}{\sigma} + 2\frac{\alpha}{\sigma^3} \sum_{i=1}^{n} \frac{(X_i - e^{-\alpha \Delta t} X_{i-1} - L(1 - e^{-\alpha \Delta t}))^2}{1 - e^{-2\alpha \Delta t}}
\]

equal zero implies together with (2.17) that

\[
\hat{\sigma} = \left( \frac{2\hat{\alpha}}{n} \sum_{i=1}^{n} \frac{(X_i - e^{-\alpha \Delta t} X_{i-1} - \hat{L}(1 - e^{-\alpha \Delta t}))^2}{1 - e^{-2\alpha \Delta t}} \right)^{\frac{1}{2}} := g(\hat{L}, \hat{\alpha}).
\]

Substituting the functions \( s(\hat{\alpha}) \) and \( g(\hat{L}, \hat{\alpha}) \) directly into the likelihood function results in a closed form solution for the maximum likelihood estimates, namely

\[
\hat{h}(\alpha) := \frac{n}{2} \ln \left( \frac{\pi g(s(\alpha), \alpha)^{\frac{2}{\alpha}}}{(1 - e^{-2\alpha \Delta t})} \right) - \sum_{i=1}^{n} \left( \frac{(X_i - e^{-\alpha \Delta t} X_{i-1} - s(\alpha)(1 - e^{-\alpha \Delta t}))^2}{g(s(\alpha), \alpha)^{\frac{2}{\alpha}}(1 - e^{-2\alpha \Delta t})} \right).
\]

Now the task becomes to solve

\[
\max_{\alpha} \hat{h}(\alpha)
\]

which is a one-dimensional problem yielding the maximum likelihood estimate \( \hat{\alpha} \) for the parameter \( \alpha \). The estimates for the remaining parameters \( \hat{L} \) and \( \hat{\sigma} \) can be computed by means of the functions \( s \) and \( g \).
2.5 Bias through Discretization

Employing an estimation approach to the discretized time-continuous stochastic differential equation is frequently used in statistical inference for diffusion processes. For example, Clewlow and Strickland [14] and Blanco and Soronow [9] make use of the linear regression method upon the discretized stochastic differential equation. Such an estimation procedure seems to be meaningful since the sample paths of a discrete version of a stochastic differential equation converge to those of the time-continuous stochastic differential. But a discretization approach can result in biased and inconsistent estimators. For maximum likelihood Lo [34] demonstrates that the estimators derived from a discretized stochastic differential need not be consistent in general. In the following, the effect of the discretization approach on the least squares estimator in the case of the mean-reverting Ornstein-Uhlenbeck process is studied.

Application of Euler’s discrete approximation to the time-continuous stochastic differential equation

\[ dX_t = \alpha(L - X_t)dt + \sigma dB_t, \quad t \geq 0, \]

results in the difference equation

\[ \Delta X_k = \alpha_0 + \alpha_1 X_k + \sigma \Delta B_k \]  

(2.23)

where \( X_k = X_{t_k}, \Delta X_k = X_{k+1} - X_k, \Delta B_k = B_{t_{k+1}} - B_{t_k}, \alpha_0 = \alpha L \Delta k, \alpha_1 = -\alpha \Delta t \) and \( \Delta t = t_{k+1} - t_k \). The classical least squares estimator \( \hat{\alpha}_1^{LS} \) of \( \alpha_1 \) in the linear regression model (2.23) is given by

\[ \hat{\alpha}_1^{LS} = \frac{\sum_{i=0}^{n-1} (X_i - \bar{X})(\Delta X_i - \overline{\Delta X})}{\sum_{i=0}^{n}(X_i - \bar{X})^2} \]  

(2.24)

whereby

\[ \bar{X} = (n+1)^{-1} \sum_{k=0}^{n} X_k, \quad \overline{\Delta X} = n^{-1} \sum_{k=0}^{n-1} \Delta X_k, \]

computed from the discrete sample \( X^{n+1} = \{X_0, X_1, \ldots, X_n\} \).

In the following the purpose is to demonstrate that the least squares estimator \( \hat{\alpha}_1^{LS} \) of the mean reversion rate \( \alpha \) given by \( \hat{\alpha}_1^{LS} = -\hat{\alpha}_1^{LS}/\Delta t \) is biased. Therefor, the asymptotic expectation of the estimator is computed.

It is assumed that the initial value \( X_0 \) follows the invariant distribution \( N(L, \sigma^2/2\alpha) \) such that the solution of (2.5) is strictly stationary.

Due to the propagation of uncertainty, the numerator and the denominator can be considered separately while calculating the expected value.

Expanding the brackets in the numerator in (2.24) gives

\[ E \left( \sum_{i=0}^{n-1} (X_i - \bar{X})(\Delta X_i - \overline{\Delta X}) \right) = \sum_{i=0}^{n-1} E \left( X_i \Delta X_i + X_i \overline{\Delta X} - \Delta X_i \bar{X} + \bar{X} \overline{\Delta X} \right). \]

Making use of the stationary moments given by \( E(X_t) = L \) and \( \gamma(h) = \text{Cov}(X_{t+h}, X_t) = \frac{\sigma^2}{2\alpha} e^{-\alpha|h|} \) yields

\[ E(X_i \Delta X_i) = E(X_i(X_{i+1} - X_i)) = \gamma(1) - \gamma(0) \]

and

\[ E(X_i \overline{\Delta X}) = \frac{1}{n} (\gamma(n - i) - \gamma(i)) \]
since $\overline{\Delta X} = \frac{1}{n} \sum_{k=0}^{n-1} \Delta X_k = \frac{1}{n} (X_n - X_0)$. Moreover, it can be easily seen that

$$E(\Delta X, \overline{X}) = \frac{1}{n+1} E \left( (X_{i+1} - X_i) \sum_{k=0}^{n} X_k \right)$$

$$= \frac{1}{n+1} \sum_{k=0}^{n} \gamma(i+1-k) - \gamma(i-k)$$

$$= \frac{1}{n+1} (\gamma(i+1) - \gamma(n-i))$$

and

$$E(\overline{S} \Delta \overline{S}) = \frac{1}{n(n+1)} E \left( (X_n - X_0) \sum_{k=0}^{n} X_k \right)$$

$$= \frac{1}{n(n+1)} \sum_{k=0}^{n} \gamma(n-k) - \gamma(k)$$

$$= 0.$$ 

These computations lead to the following representation of the numerator:

$$E \left( \sum_{i=0}^{n-1} (S_i - \overline{S})(\Delta S_i - \overline{\Delta S}) \right) = \frac{1}{n} \sum_{i=0}^{n-1} \{ \gamma(1) - \gamma(0) + n^{-1}(\gamma(n-i) - \gamma(i)) \}$$

$$- \frac{1}{n+1} (\gamma(i+1) - \gamma(n-i))$$

$$= n (\gamma(1) - \gamma(0)) + \frac{1}{n} (\gamma(n) - \gamma(0))$$

Note that $\frac{1}{n} \sum_{i=0}^{n} (X_i - \overline{X})^2$ is an unbiased estimator of the variance $\gamma(0)$ of the strict stationary process. So it follows that

$$E(\hat{\alpha}_L^S) \approx \frac{\gamma(1) - \gamma(0) + n^{-2}(\gamma(n) - \gamma(0))}{\gamma(0)}$$

which converges as $n \to \infty$ to

$$\frac{\gamma(1) - \gamma(0)}{\gamma(0)} = e^{-\alpha_1} - 1.$$ 

This implicates that for an arbitrary $\Delta t > 0$ the expected value $E(\hat{\alpha}_L^S)$ converges as $n \to \infty$ to

$$- \frac{e^{-n \Delta t} - 1}{\Delta t} = \alpha - \alpha^2 \frac{\Delta t}{2} + \alpha - \alpha^3 \frac{(\Delta t)^2}{6} + \ldots$$

where the right-hand side corresponds to the Taylor expansion of the exponential function. Hence, this equation highlights the bias of the least squares estimator of $\alpha$ which is due to the dependence of the estimator upon the accuracy $\Delta t$ of the discrete approximation. Even though this bias vanishes as $\Delta t \to 0$, one has to keep in mind in practice that a naive application of the discretization provides a biased estimator.
Chapter 3

A Maximum Likelihood Approach

In this chapter a generalized mean-reverting Ornstein-Uhlenbeck process of the form

\[ dX_t = (L(t) - \alpha X_t) \, dt + \sigma dB_t, \quad t \geq 0, \]

where \( L(t) \) is a parametric function which is linear in the parameter, is proposed. Application of maximum likelihood estimation based on time-continuous observations yields a time-continuous estimator of the parameter incorporated in \( L(t) \) and of \( \alpha \). The resulting estimator may be represented explicitly and large sample results like strong consistency and asymptotic normality are proved in the special case of a periodic mean reversion function \( L(t) \). In the asymptotic framework of these results the observed number of periods tends to infinity. The essential idea of the investigation of the asymptotic properties is the interpretation of the stochastic process as a sequence of random variables that take values in some function space.

3.1 Generalized Ornstein-Uhlenbeck Process

The ordinary Ornstein-Uhlenbeck process, see Section 2.1, is defined as solution to the stochastic differential equation

\[ dX_t = \alpha (\mu - X_t) \, dt + \sigma dB_t, \quad t \geq 0, \]

where \( \alpha \) and \( \sigma \) are positive constants, \( \mu \in \mathbb{R} \) and where \((B_t)_{t \geq 0}\) is standard Brownian motion. Originally introduced by Ornstein and Uhlenbeck [36] as a model for particle motion in a fluid, this process is now widely used in many areas of application. This is due to its main characteristic which is the tendency to return toward the long-term equilibrium \( \mu \), also known as mean reversion. This property is observed in many real-life processes, particularly in commodity and energy price processes, see e.g. Geman [22].

However, in many practical situations the assumption of a constant mean level is not adequate due to seasonality patterns or a long-term trend of the process. Thus a more general version of the process introduce above may be considered, namely the process satisfying the stochastic differential equation

\[ dX_t = (L(t) - \alpha X_t) \, dt + \sigma dB_t, \quad t \geq 0, \quad X_0 = \zeta, \tag{3.1} \]

where \( L(t) \) is a time-dependent mean reversion level and where \( \alpha, \sigma \) are positive constants. Let \( \zeta \) be a real-valued random variable that is independent of the Brownian motion \((B_t)_{t \geq 0}\) and that satisfies \( \mathbb{E}(\zeta^2) < \infty \).
 CHAPTER 3. A MAXIMUM LIKELIHOOD APPROACH

Note that model (3.1) differs from the original Ornstein-Uhlenbeck process in the position of $\alpha$ within the drift term. However, model (3.1) can easily be transformed to a process with drift term $\alpha(\tilde{L}(t) - X_t)dt$ where $\tilde{L}(t) = L(t)/\alpha$. The advantage of (3.1) compared to the process provided with the drift $\alpha(L(t) - X_t)dt$ is the simplification of the study of the estimators.

Drift parameter estimation for time-continuously observed diffusion processes is a well-established area of research, for which a variety of techniques has been proposed. For example, Kutoyants [31] investigates several estimation techniques for the drift term of ergodic, time-homogenous diffusion processes. The analysis of asymptotic properties of drift estimates for time-inhomogeneous diffusion models has been paid much less attention. Among some other authors, Bishwal [8] studies the maximum likelihood estimator for time-inhomogeneous diffusions, see Remark 3.6 for some comments.

In this work a parametric model for the mean reversion function $L(t)$ is made by assuming

$$L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t),$$

(3.2)

where the basis functions $\varphi_1(t), \ldots, \varphi_p(t)$ are known and $\mu_1, \ldots, \mu_p$ and $\alpha$ are unknown parameters. The diffusion parameter $\sigma$ is assumed to be known which is a common assumption in the field of drift parameter estimation for a time-continuous diffusion. The reason for this assumption is given by the singularity of the measures corresponding to different diffusion parameters so that $\sigma$ can be computed, rather than estimated, from a single continuous-time observation path.

The conditions on the drift coefficient, here

$$S(t, X_t, \theta) = \sum_{i=1}^{p} \mu_i \varphi_i(t) - \alpha X_t,$$

(3.3)

where $\theta = (\mu_1, \ldots, \mu_p, \alpha)^t \in \Theta = \{ x \in \mathbb{R}^{p+1} : x_i \in \mathbb{R}, i = 1, \ldots, p, \ x_{p+1} > 0 \}$, that ensure existence and uniqueness of a solution of equation (3.1) are well known, see Kuo [29] (Theorem 10.3.5, p. 192), for example. Due to the linear form of $S(\theta, t, \cdot)$ the global Lipschitz condition is satisfied in this setting such that there exists at most one solution of (5.6). If the basis functions $\varphi_1(t), \ldots, \varphi_p(t)$ are bounded on compact sets in $\mathbb{R}$, for instance, then the linear growth condition which implies the uniqueness of an existing solution is fulfilled.

### 3.2 Maximum Likelihood Estimation

Denote by $P_X$ the measure induced by the observable realizations $X^T = \{ X_t, 0 \leq t \leq T \}$ on the measurable space $(\mathcal{C}[0,T], \mathcal{B}[0,T])$, $\mathcal{C}[0,T]$ being the space of continuous, real-valued functions on $[0,T]$ and $\mathcal{B}[0,T]$ the associated Borel $\sigma$-field. Moreover, let $P_B$ be the measure generated by the Brownian motion on $(\mathcal{C}[0,T], \mathcal{B}[0,T])$. Then the likelihood function of observations $X^T$ of the process with stochastic differential (3.1) is defined by

$$\mathcal{L}(\theta, X^T) := \frac{dP_X}{dP_B}(X^T)$$
3.2. MAXIMUM LIKELIHOOD ESTIMATION

where \( dP_X/dP_B \) is the Radon-Nikodym derivative, cf. Section 1.3 for more details. The maximum likelihood estimator is defined as the maximum of the functional \( \theta \mapsto \mathcal{L}(\theta, X^T) \), i.e.

\[
\hat{\theta}_{ML} := \arg \max_{\theta} \mathcal{L}(\theta, X^T).
\]

A corollary to Girsanov’s theorem, see Theorem 7.6 on p. 246 in [33] by Lipster and Shiryayev, gives an explicit expression of the likelihood function of a diffusion process provided that

\[
P \left( \int_0^T S(t, X_t, \theta)^2 dt < \infty \right) = 1
\]

for all \( 0 \leq T < \infty \) and all \( \theta \). It is required that the distribution of the initial variable \( \zeta \) does not depend on \( \theta \) otherwise the Radon-Nikodym derivative given in (3.5) would contain an additional factor, see Kutoyants [31] (p. 37) for details.

**Proposition 3.1.** Let \( \mathcal{L}(\theta, X^T) \) denote the likelihood function of observations \( X^T = \{X_t, 0 \leq t \leq T\} \) of the process introduced in (3.1) provided with the mean reversion function (3.2). If the drift term given in (3.3) satisfies condition (3.4) then

\[
\hat{\theta}_{ML} = Q_T^{-1}P_T.
\]

The objects \( Q_T \in \mathbb{R}^{(p+1) \times (p+1)} \) and \( P_T \in \mathbb{R}^{p+1} \) are defined as

\[
Q_T = \begin{pmatrix} G_T & -a_T \\ -a_T^t & b_T \end{pmatrix},
\]

\[
P_T = \left( \int_0^T \varphi_1(t) dX_t, \ldots, \int_0^T \varphi_p(t) dX_t, -\int_0^T X_t dX_t \right)^t
\]

where \( G_T = (\int_0^T \varphi_j(t)^2 dt)_{1 \leq j,k \leq p} \in \mathbb{R}^{p \times p}, \ A_T = (\int_0^T \varphi_1(t)X_t dt, \ldots, \int_0^T \varphi_p(t)X_t dt)^t \) and \( b_T = \int_0^T X_t^2 dt \).

**Proof.** The likelihood function of a diffusion process of the form

\[
dX_t = S(t, X_t, \theta)dt + \sigma dB_t, \quad 0 \leq t \leq T,
\]

is given by

\[
\mathcal{L}(\theta, X^T) = \frac{dP_X}{dP_B}(X^T) = \exp \left( \frac{1}{\sigma^2} \int_0^T S(t, X_t, \theta) dX_t - \frac{1}{2\sigma^2} \int_0^T S(t, X_t, \theta)^2 dt \right)
\]

if condition (3.4) is fulfilled, see Lipster and Shiryayev [33] (Theorem 7.6, p. 246). The partial derivatives of the logarithm of this functional are

\[
\frac{\partial}{\partial \theta_i} \ln(\mathcal{L}(\theta, X^T)) = \frac{1}{\sigma^2} \int_0^T \frac{\partial}{\partial \theta_i} S(t, X_t, \theta) dX_t - \frac{1}{2\sigma^2} \int_0^T S(t, X_t, \theta) \frac{\partial}{\partial \theta_i} S(t, X_t, \theta) dt.
\]

The drift function of our mean reversion model is given in (3.3) and the derivatives are

\[
\frac{\partial}{\partial \theta_i} S(t, X_t, \theta) = \begin{cases} \varphi_i(t), & i = 1, \ldots, p; \\ -X_t, & i = p + 1. \end{cases}
\]

Setting the partial derivatives of the log-likelihood function in (5.12) equal zero gives a system of linear equations which yields the assertion.
CHAPTER 3. A MAXIMUM LIKELIHOOD APPROACH

Remark 3.2. Note that the matrix $Q_T$ introduced in the previous lemma is not a priori invertible. However, it will be shown later that it is invertible for $T$ large enough in the periodic mean reversion model, see Remark 3.5 for more details on that.

Remark 3.3. The maximum likelihood estimator introduced above can be motivated by an alternative derivation: Interpreting Euler’s discretization of the stochastic differential equation (3.1) as a linear model and applying the ordinary least squares estimation method to this discrete version provides an estimator $\hat{\theta}_{ML}^T$ containing Riemann and Itô sums. It can be then seen that $\hat{\theta}_{ML}^T \to \theta_{ML}$ as $\Delta t \to 0$. See Section 5.3 for more on that.

3.3 Consistency and Asymptotic Normality

In many applications, particularly in energy and commodity prices and temperature, the data display regular seasonal effects. This important phenomenon can be modeled by assuming that the mean reversion function $L(t)$ is periodic, i.e. that

$$L(t + \nu) = L(t)$$

where $\nu$ is the period observed in the data. The resulting stochastic process exhibits a cyclical evolution due to the periodicity of this mean reversion mechanism. Combining the assumption of periodicity with the parametric model (3.2) leads to the requirement

$$\varphi_j(t + \nu) = \varphi_j(t).$$

By applying Gram-Schmidt orthogonalization, one may assume without loss of generality that $\varphi_1(t), \ldots, \varphi_p(t)$ form an orthonormal system in $L^2([0, \nu], \frac{1}{\nu} d\lambda)$, i.e. that

$$\int_0^\nu \varphi_j(t)\varphi_k(t)dt = \begin{cases} \nu, & j = k \\ 0, & j \neq k. \end{cases}$$

In the rest of this paper it will be supposed that an integral multiple of the period length is observed, i.e.

$$T = N \nu,$$

for some integer $N$. Moreover, without loss of generality, it will be assumed that $\nu = 1$.

Under the above assumptions, the matrix $Q_T$, defined in Proposition 3.1, simplifies to

$$Q_T = \begin{pmatrix} I_p & -a_T \\ -a_T^T & b_T \end{pmatrix}$$

where $I_p$ denotes the $(p \times p)$-identity matrix. The inverse of a matrix of this special form can be explicitly computed by the following lemma.

Lemma 3.4. The inverse of the matrix $Q_T$, given in (3.10), is given by

$$Q_T^{-1} = \frac{1}{T} \begin{pmatrix} I_p + \gamma_T \Lambda_T \Lambda_T^T & -\gamma_T \Lambda_T \\ -\gamma_T \Lambda_T^T & \gamma_T \end{pmatrix}$$

where

$$\Lambda_{T,i} = \frac{1}{T} \int_0^T \varphi_i(t)X_t dt, \quad i = 1, \ldots, p$$

and

$$\gamma_T = \left(1 - \frac{1}{T} \int_0^T X_t^2 dt - \sum_{i=1}^p \Lambda_{T,i}^2 \right)^{-1}.$$
3.3. CONSISTENCY AND ASYMPTOTIC NORMALITY

Proof. The following formula for the inverse of a partitioned matrix which can be deduced from
the Frobenius matrix inversion formula, cf. Gantmacher [21], p. 73, is used. Alternatively
the formula can also be verified directly. It is for \( a \in \mathbb{R}^p, b \in \mathbb{R} \)

\[
\begin{pmatrix}
I_p & a \\
\ast & b
\end{pmatrix}^{-1} = \begin{pmatrix}
I_p + \frac{1}{\|a\|^2} aa^t & -\frac{1}{\|a\|^2} a \\
-\frac{1}{\|a\|^2} a^t & \frac{1}{\|a\|^2}
\end{pmatrix}
\]

(3.14)

where \( \| \cdot \| \) denotes the usual Euclidean norm on \( \mathbb{R}^p \). With the notation introduced above \( Q_T \)
can written as

\[
Q_T = T \begin{pmatrix}
I_p & -\Lambda_T \\
-\Lambda_t & \frac{1}{T} \int_0^T X_t^2 dt
\end{pmatrix}
\]

and application of the above formula yields the assertion. \( \Box \)

Remark 3.5. Note that the Frobenius matrix inversion formula holds if and only if the entries
of the matrix on the right hand side of (3.14) are well-defined. It will be demonstrated in the
proof of Proposition 3.11 that the limit of \( \frac{1}{T} Q_T^{-1} \) is well defined since it will be shown that the
limit of \( \gamma_T \) denoted by \( \gamma \) is greater than zero. Consequently, \( \frac{1}{T} Q_T^{-1} \) exists almost surely if \( T \) is
large enough.

The main results on the asymptotic behavior of the maximum likelihood estimator in the
periodic Ornstein-Uhlenbeck model are stated now.

Theorem 3.1. Let \( \{X_t, 0 \leq t \leq T \} \) be observations of the periodic mean reversion process as
introduced in (3.2), satisfying (3.8) and (3.9). Then the maximum likelihood estimator given
in Proposition 3.1 is consistent, i.e.

\[
\hat{\theta}_{ML} \to \theta, \text{ almost surely,}
\]
as \( T \to \infty \).

For the description of the asymptotic distribution of \( \hat{\theta}_{ML} \), the \((p + 1) \times (p + 1)\) matrix \( C \)
has to be introduced:

\[
C = \begin{pmatrix}
I_p + \gamma \Lambda \Lambda^t & -\gamma \Lambda \\
-\gamma \Lambda^t & \gamma
\end{pmatrix}
\]

(3.15)

where the entries are defined by

\[
\Lambda_i = \int_0^1 \varphi_i(t) \tilde{h}(t) dt, \quad i = 1, \ldots, p,
\]

(3.16)

and \( \Lambda = (\Lambda_1, \ldots, \Lambda_p)^t \). Here, the function \( \tilde{h} : [0, \infty) \to \mathbb{R} \) is defined by

\[
\tilde{h}(t) = e^{-\alpha t} \sum_{j=1}^p \mu_j \int_{-\infty}^t e^{\alpha s} \varphi_j(s) ds.
\]

(3.17)
Theorem 3.2. Let \( \{X_t, 0 \leq t \leq T\} \) be observations of the periodic mean reversion process as introduced in (3.2), satisfying (3.8) and (3.9). Then the maximum likelihood estimator given in Proposition 3.1 is asymptotically normal. More precisely,

\[
\sqrt{T}\sigma^{-1}(\hat{\theta}_{ML} - \theta) \xrightarrow{D} N(0, C)
\]

where \( C \) is defined as in (3.15).

The proofs of these theorems are postponed into the next section.

Remark 3.6. Bishwal [8] obtains large sample results on the maximum likelihood estimator for the drift of time-inhomogeneous diffusions provided with drift function \( f(\theta, t, x) \) defined on \( \Theta \times [0, T] \times C[0, T] \), where \( \Theta \) is the parameter space and \( C[0, T] \) is the space of all continuous, real-valued functions on \([0, T]\). Note that the model presented here belongs to this class of diffusion processes and that the results in this chapter comply with Bishwal’s findings on the asymptotic behavior of the estimator. However, in order to apply Bishwal’s results directly, one has to verify the required conditions, among others the convergence of

\[
\frac{1}{m_T} \int_0^T \left( \frac{\partial}{\partial \theta} f(\theta, t, X_t) \right)^2 dt \tag{3.18}
\]

where \( m_T \) is an increasing, non-random sequence, see condition (A7) on p. 64 in Bishwal [8]. The verification of this condition does not go without saying in the periodic Ornstein-Uhlenbeck model and requires some ideas and techniques. In contrast to Bishwal’s general setting, an explicit representation of the estimator which is used for the study of its asymptotic behavior is obtained here. By investigating this explicit representation, the convergence of the multi-dimensional version of (3.18) is shown, see Proposition 3.11 where \( \frac{1}{T} Q_T \) is the analog of the term (3.18).

3.4 Proofs

The proofs of Theorem 3.1 and Theorem 3.2 require a number of auxiliary results which are established in this section.

Proof of Consistency

The proof of both the consistency and the asymptotic normality makes use of the following representation of the maximum likelihood estimator.

Proposition 3.7. The maximum likelihood estimator \( \hat{\theta}_{ML} \), defined in Proposition 3.1, can be written as

\[
\hat{\theta}_{ML} = \theta + \sigma Q_T^{-1} R_T \tag{3.19}
\]

where

\[
R_T := \begin{pmatrix}
\int_0^T \varphi_1(t) dB_t \\
\vdots \\
\int_0^T \varphi_p(t) dB_t \\
- \int_0^T X_t dB_t
\end{pmatrix}
\]

and where \( Q_T \) is defined in Proposition 3.1.
3.4. PROOFS

Proof. By definition

$$\hat{\theta}_{ML} = Q_T^{-1} P_T$$

where $Q_T$ and $P_T$ are defined as in Proposition 3.1. Rewriting this by making use of (3.1), namely

$$dX_t = \left( \sum_{j=1}^{p} \mu_j \varphi_j(t) - \alpha X_t \right) dt + \sigma dB_t,$$

gives

$$\int_0^T \varphi_i(t) dX_t = \sum_{j=1}^{p} \mu_j \int_0^T \varphi_i(t) \varphi_j(t) dt - \alpha \int_0^T \varphi_i(t) X_t dt + \sigma \int_0^T \varphi_i(t) dB_t, \quad i = 1, \ldots, p,$$

$$\int_0^T X_t dX_t = \sum_{j=1}^{p} \mu_j \int_0^T X_t \varphi_j(t) dt - \alpha \int_0^T X_t^2 dt + \sigma \int_0^T X_t dB_t.$$

Hence, it follows that

$$P_T = \begin{pmatrix} \int_0^T \varphi_1(t) dX_t \\ \vdots \\ \int_0^T \varphi_p(t) dX_t \\
- \int_0^T X_t dX_t \end{pmatrix} = Q_T \theta + \sigma R_T$$

so that $\hat{\theta}_{ML} = \theta + \sigma Q_T^{-1} R_T$.

In what follows, it will be shown that $Q_T^{-1} R_T$ converges to zero almost surely, as $T \to \infty$. In order to do so, write

$$Q_T^{-1} R_T = (T Q_T^{-1}) \left( \frac{1}{T} R_T \right).$$

It detail, it will be established that $T Q_T^{-1}$ converges almost surely to a finite limit and that $\frac{1}{T} R_T$ converges almost surely to zero. Both of these results require some auxiliary results which will be proved first.

**Lemma 3.8.** The solution of the stochastic differential equation (3.1) has the explicit representation

$$X_t = e^{-\alpha t} X_0 + h(t) + Z_t,$$

(3.22)

where

$$h(t) = e^{-\alpha t} \int_0^t e^{\alpha s} L(s) ds = e^{-\alpha t} \sum_{i=1}^{p} \mu_i \int_0^t e^{\alpha s} \varphi_i(s) ds$$

and

$$Z_t = \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s.$$

Proof. The Itô lemma states for $Y_t = g(t, X_t)$ that

$$dY_t = \frac{\partial g}{\partial t}(t, X_t) dt + \frac{\partial g}{\partial x}(t, X_t) dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t) (dX_t)^2$$

$$= e^{-\alpha t} \int_0^t e^{\alpha s} L(s) ds$$

and

$$Z_t = \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s.$$
which yields for $g(t, x) = e^{\alpha t} x$

$$dY_t = \alpha e^{\alpha t} X_t dt + e^{\alpha t} dX_t.$$  

Plugging (3.1) in this equation gives

$$dY_t = e^{\alpha t} (L(t) dt + \sigma dB_t).$$

Integrating and multiplying by $e^{-\alpha t}$ finishes the proof of the lemma.

The process $(X_t)_{t \geq 0}$ is not stationary in the ordinary sense since $h$ and the distribution of $Z_t$ are time-dependent, see (3.22). Thus the ergodic theorem cannot be applied directly. In order to solve this problem, a stationary solution to the stochastic differential equation (5.6) for $t \in \mathbb{R}$ instead of $t \geq 0$ is introduced in the following. Define the process

$$\bar{X}_t = \bar{h}(t) + \bar{Z}_t$$  

(3.23)

where $\bar{h}(t)$ is defined in (3.17) and $\bar{Z}_t$ is defined as

$$\bar{Z}_t = \sigma e^{-\alpha t} \int_{-\infty}^{t} e^{\alpha s} d\tilde{B}_s$$  

(3.24)

where $(\tilde{B}_s)_{s \in \mathbb{R}}$ denotes a bilateral Brownian motion, i.e.

$$\tilde{B}_s := B_s 1_{\mathbb{R}^+}(s) + \bar{B}_s 1_{\mathbb{R}^-}(s)$$

with $(B_s)_{s \geq 0}$ and $(\bar{B}_s)_{s \geq 0}$ two independent standard Brownian motions. Thereby, $1_A$ denotes the indicator function of the set $A$.

In the following lemma the process introduced above is viewed as a sequence of function-valued random variables. This method originates from probability theory on Banach spaces.

**Lemma 3.9.** The sequence $(W_k)_{k \in \mathbb{N}}$ of $C[0, 1]$-valued random variables defined by

$$W_k(s) := \bar{X}_{k-1+s}, \quad 0 \leq s \leq 1, \quad k \in \mathbb{N}$$  

(3.25)

is stationary and ergodic.

**Proof.** Denote by $\tilde{h}_0$ the restriction of the function $\tilde{h}$ to $[0, 1]$. Since the function $\tilde{h}$ is periodic, one has the decomposition

$$W_k(t) = \tilde{h}(k - 1 + t) + \sigma e^{-\alpha (k-1+t)} \int_{-\infty}^{t} e^{\alpha s} d\tilde{B}_s$$

$$= \tilde{h}_0(t) + \sigma e^{-\alpha (k-1+t)} \int_{k-1}^{k-1+t} e^{\alpha s} d\tilde{B}_s + \sigma \sum_{l=-\infty}^{k-1} e^{-\alpha (k-1+t)} \int_{l-1}^{l} e^{\alpha s} d\tilde{B}_s.$$  

Making use of the time shifted Brownian motion

$$\tilde{B}_s^{(l)} := \tilde{B}_{s+l}$$
3.4. PROOFS

yields

\[ W_k(t) = \tilde{h}_0(t) + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} d\tilde{B}_s^{(k-1)} + \sigma \sum_{l=-\infty}^{k-1} e^{-\alpha (k-l+t)} \int_0^1 e^{\alpha s} d\tilde{B}_s^{(l-1)} \]

\[ = \tilde{h}_0(t) + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} d\tilde{B}_s^{(k-1)} + \sigma \sum_{j=-\infty}^{0} e^{-\alpha (1+t-j)} \int_0^1 e^{\alpha s} d\tilde{B}_s^{(j+k-2)} . \]

Consequently, this can be written as

\[ W_k(\cdot) = \tilde{h}_0(\cdot) + F_0(Y_{k-1}) + \sum_{j=-\infty}^{0} e^{\alpha(j-1)} F(Y_{j+k-2}) \]

where the almost surely defined functionals

\[ F_0 : C[0,1] \to C[0,1]; \omega \mapsto \left( t \mapsto \sigma e^{-\alpha t} \int_0^t e^{\alpha s} d\omega(s) \right), \]

\[ F : C[0,1] \to C[0,1]; \omega \mapsto \left( t \mapsto \sigma e^{-\alpha t} \int_0^1 e^{\alpha s} d\omega(s) \right) \]

and the \( C[0,1] \)-valued random variables

\[ Y_l = (s \mapsto \tilde{B}_s^{(l)} - \tilde{B}_0^{(l)}, 0 \leq s < 1) \]

were used. The sequence \( (Y_l)_{l \in \mathbb{Z}} \) consists of independent and identically distributed random variables. This implies that \((W_k)_{k \in \mathbb{N}}\) is stationary and ergodic since each element of this sequence can be represented as a measurable function \( G : (C[0,1])^\mathbb{N} \to C[0,1] \) of elements of the i.i.d. sequence \((Y_l)_{l \in \mathbb{Z}}\), i.e.

\[ W_k = G(Y_{k-1}, Y_{k-2}, \ldots) . \]

\[ \square \]

**Lemma 3.10.** As \( t \to \infty \) one has

\[ |\tilde{X}_t - X_t| \to 0, \text{ almost surely.} \]

**Proof.** It holds that

\[ |\tilde{X}_t - X_t| \leq e^{-\alpha t} |X_0| + |\tilde{h}(t) - h(t)| + |\tilde{Z}_t - Z_t| \]

\[ \leq e^{-\alpha t} |X_0| + e^{-\alpha t} \sum_{i=1}^{p} \mu_i \int_{-\infty}^0 e^{\alpha s} |\varphi_i(s)| ds + e^{-\alpha t} \int_{-\infty}^0 e^{\alpha s} d\tilde{B}_s| . \]

Obviously, the three terms on the right-hand side converge toward zero as \( t \to \infty \).

\[ \square \]

**Proposition 3.11.** As \( T \to \infty \), it is

\[ T Q_T^{-1} \to C, \text{ almost surely,} \]

where \( C \) is the matrix defined in (3.15).
CHAPTER 3. A MAXIMUM LIKELIHOOD APPROACH

Proof. Consider first the entries of the vector $\Lambda_T$, i.e. $\frac{1}{T} \int_0^T X_t \varphi_j(t) dt$. From Lemma 3.10 one may conclude that
\[
\frac{1}{T} \int_0^T \tilde{X}_t \varphi_j(t) dt - \frac{1}{T} \int_0^T X_t \varphi_j(t) dt \to 0,
\]
almost surely. Moreover, since $(\tilde{X}_{k-1+s})_{k \in \mathbb{N}}$ is stationary and ergodic by Lemma 3.9, the ergodic theorem gives
\[
\frac{1}{T} \int_0^T \tilde{X}_t \varphi_j(t) dt = \frac{1}{T} \sum_{k=1}^{T} \int_{k-1}^k \tilde{X}_t \varphi_j(t) dt \to E \left( \int_0^1 \tilde{X}_t \varphi_j(t) dt \right) = \int_0^1 \tilde{h}(t) \varphi_j(t) dt.
\]
Thus the convergence of $\Lambda_{T,j}$, $1 \leq j \leq p$ has been established. In order to determine the limit of $\gamma_T$, it suffices to consider the term $\frac{1}{T} \int_0^T X_t^2 dt$. It holds that
\[
\left| \frac{1}{T} \int_0^T (Z_t - \tilde{Z}_t) dt \right| \leq \frac{1}{T} \int_0^T |Z_t - \tilde{Z}_t| dt = \frac{1}{T} \int_0^T e^{-\alpha t} \left| \int_{-\infty}^0 e^{\alpha s} d\tilde{B}_s \right| dt \to 0,
\]
almost surely, as $T \to \infty$. The ergodic theorem gives
\[
\frac{1}{T} \int_0^T \tilde{Z}_t dt \to E(\tilde{Z}_0) = 0,
\]
compare the proof of Lemma 3.9. Thus
\[
\limsup_{T \to \infty} \frac{1}{T} \int_0^T Z_t dt < \infty.
\]
Observe that the function
\[
h(t) = e^{-\alpha t} \sum_{i=1}^{p} \mu_i \int_0^t e^{\alpha s} \varphi_i(s) ds
\]
is bounded and $X_0 < \infty$ almost surely. These facts justify
\[
\limsup_{T \to \infty} \frac{1}{T} \int_0^T X_t dt = \limsup_{T \to \infty} \frac{1}{T} \int_0^T (e^{-\alpha t} X_0 + h(t) + Z_t) dt < \infty, \tag{3.26}
\]
almost surely. It follows from (3.26) and Lemma 3.10 that
\[
\frac{1}{T} \int_0^T \tilde{X}_t^2 dt - \frac{1}{T} \int_0^T X_t^2 dt = \frac{1}{T} \int_0^T (\tilde{X}_t + X_t)(\tilde{X}_t - X_t) dt \to 0.
\]
Consequently, again by the ergodic theorem, one gets
\[
\frac{1}{T} \int_0^T \tilde{X}_t^2 dt = \frac{1}{T} \sum_{k=1}^{T} \int_{k-1}^k \tilde{X}_t^2 dt
\]
\[
\to E \left( \int_0^1 \tilde{h}(t)^2 dt \right)
\]
\[
= \int_0^1 (\tilde{h}(t))^2 dt + 2 \int_0^1 \tilde{h}(t) \tilde{Z}_t dt + \int_0^1 \tilde{Z}_t^2 dt
\]
\[
= \int_0^1 (\tilde{h}(t))^2 dt + \int_0^1 (\tilde{h}(t))^2 dt + \frac{\sigma^2}{2\alpha}.
\]
By Bessel’s inequality, it holds
\[ \sum_{i=1}^{p} \Lambda_i^2 \leq \int_0^1 (\tilde{h}(t))^2 \, dt \]
and thus \((\int_0^1 (\tilde{h}(t))^2 \, dt + E(\tilde{Z}_0)^2) - \sum_{i=1}^{p} \Lambda_i^2) \geq E(\tilde{Z}_0)^2 = \frac{\sigma^2}{2\alpha} > 0\). This proves the assertion of the proposition.

**Lemma 3.12.** The sequence \(\frac{1}{\sqrt{T}} R_T\) is bounded in \(L^2\).

**Proof.** Note that
\[
\frac{1}{\sqrt{T}} \int_0^T \varphi_i(t) dB_t
\]
is \(L^2\)-bounded because
\[
\text{Var} \left( \frac{1}{\sqrt{T}} \int_0^T \varphi_i(t) dB_t \right) = E \left( \frac{1}{\sqrt{T}} \int_0^T \varphi_i(t) dB_t \right)^2 = \frac{1}{T} \int_0^T \varphi_i^2(t) \, dt = 1. \quad (3.27)
\]
For the last entry of \(\frac{1}{\sqrt{T}} R_T\) one has to prove the boundedness of
\[
\text{Var} \left( \frac{1}{\sqrt{T}} \int_0^T X_t dB_t \right) = \frac{1}{T} E \left( \int_0^T X_t^2 \, dt \right) = \frac{1}{T} E \left( \int_0^T 2e^{-\alpha t} X_0 h(t) + 2e^{-\alpha t} X_0 Z_t + e^{-2\alpha t} X_0^2 + 2h(t) Z_t + h(t)^2 + Z_t^2 \, dt \right).
\]
Since \(Z_t\) is a zero-mean random variable the expectation of the second and fourth term is zero. Moreover, the variance \(E(Z_t^2) = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t})\) is bounded and justifies
\[ \sup_{T \geq 0} \frac{1}{T} E \left( \int_0^T Z_t^2 \, dt \right) < \infty. \]
Moreover, the boundedness of \(h(t)\) and \(E(X_0^2) < \infty\) give
\[ \sup_{T \geq 0} \frac{1}{T} E \left( \int_0^T e^{-\alpha t} X_0 h(t) \, dt \right) < \infty \]
and
\[ \sup_{T \geq 0} \frac{1}{T} \int_0^T h(t)^2 \, dt < \infty. \]
This finishes the proof of the \(L^2\)-boundedness of \(\frac{1}{\sqrt{T}} R_T\).

**Proposition 3.13.** As \(T \to \infty\), it is
\[ \lim_{T \to \infty} \frac{1}{T} R_T = 0, \text{ almost surely.} \quad (3.28) \]
Proof. Observe that $R_T$ is a martingale; thus it holds by using Doob’s maximal inequality for submartingales that for any $\epsilon > 0$

$$P\left( \sup_{2^k \leq T \leq 2^{k+1}} \frac{1}{T} |R_T| \geq \epsilon \right) \leq P\left( \sup_{2^k \leq T \leq 2^{k+1}} |R_T| \geq \epsilon 2^k \right) \leq \frac{4}{\epsilon^2 2^{2k}} \mathbb{E} |R_{2^{k+1}}|^2 = O(2^{-k}).$$

Applying the Borel-Cantelli theorem, one obtains $\limsup_{T \to \infty} \frac{1}{T} |R_T| \leq \epsilon$, almost surely, and thus $R_T / T \to 0$.

Proof of Theorem 3.1. This follows directly from Proposition 3.11 and Proposition 3.13. □

Proof of Asymptotic Normality

Proof of Theorem 3.2. The representation (3.19) is used again, i.e. $\hat{\theta}_{ML} - \theta = \sigma Q_T^{-1} R_T$, which can be written as

$$\sqrt{T} \left( \frac{\hat{\theta}_{ML} - \theta}{\sigma} \right) = \sqrt{T} Q_T^{-1} R_T = (TQ_T^{-1}) \frac{1}{\sqrt{T}} R_T.$$

By Proposition 3.11, $T Q_T^{-1}$ converges almost surely to the matrix $C$. Then, by Slutsky’s theorem, Theorem 3.2 is an immediate corollary of Proposition 3.14. Note that $\Sigma_0^{-1} = C$ by applying the same formula as in the proof of Lemma 3.4 to $\Sigma_0$.

Proposition 3.14. Under the assumptions of Theorem 3.2, it holds, as $T \to \infty$, that

$$\frac{1}{\sqrt{T}} R_T \xrightarrow{D} N(0, \Sigma_0),$$

where the $(p + 1) \times (p + 1)$ matrix $\Sigma_0$ is defined as

$$\Sigma_0 = \left( \begin{array}{cc} I_p & \Lambda^t \\ \Lambda^t & \omega \end{array} \right),$$

where $\omega = \int_0^1 \hat{h}(t)^2 dt + \frac{\sigma^2}{2\alpha}$. The entries of the vector $\Lambda$ are specified in (3.16).

The remaining part of this section is devoted to the proof of this proposition. Recall that

$$\frac{1}{\sqrt{T}} R_T = \left( \begin{array}{c} \frac{1}{\sqrt{T}} \int_0^T \varphi_1(t) dB_t \\ \vdots \\ \frac{1}{\sqrt{T}} \int_0^T \varphi_p(t) dB_t \\ -\frac{1}{\sqrt{T}} \int_0^T X_t dB_t \end{array} \right).$$

Since the basis functions $\varphi_1, \ldots, \varphi_p$ are orthonormal, the first $p$ entries of the vector $\frac{1}{\sqrt{T}} R_T$ are independent, normally distributed random variables with mean zero and variance 1. Thus it remains to investigate the asymptotic distribution of the last entry

$$\frac{1}{\sqrt{T}} \int_0^T X_t dB_t,$$

and its joint distribution with the first $p$ components.
3.4. PROOFS

By Lemma 3.8, the process \((X_t)_{t \geq 0}\) can be expressed as
\[
X_t = e^{-\alpha t} X_0 + h(t) + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s,
\]
and thus
\[
\frac{1}{\sqrt{T}} \int_0^T X_t dB_t = \frac{X_0}{\sqrt{T}} \int_0^T e^{-\alpha t} dB_t + \frac{1}{\sqrt{T}} \int_0^T h(t) dB_t + \sigma \frac{1}{\sqrt{T}} \int_0^T \int_0^t e^{\alpha(t-s)} dB_s dB_t.
\]
(3.29)
The first term on the right hand side converges to 0 in probability, as
\[
\text{Var} \left( \frac{1}{\sqrt{T}} \int_0^T e^{-\alpha t} dB_t \right) = \frac{1}{T} \int_0^T e^{-2\alpha t} dt \to 0.
\]
The second term is normally distributed with mean zero and variance
\[
\frac{1}{T} \int_0^T (h(t))^2 dt \to \int_0^1 (\tilde{h}(t))^2 dt.
\]
The asymptotic distribution of the third term, as well as its joint distribution with any stochastic integral \(\int_0^T \phi(t) dB_t\), will be evaluated next. The notions from the theory of integral operators which are used in the subsequent proofs are briefly explained in Section 1.4.

**Proposition 3.15.** Let \(\phi : [0, \infty) \to \mathbb{R}\) be an \(L^2\)-function for which
\[
\sigma^2_\phi := \lim_{T \to \infty} \frac{1}{T} \int_0^T (\phi(t))^2 dt
\]
exists. Then, as \(T \to \infty\),
\[
\frac{1}{\sqrt{T}} \int_0^T \int_0^t e^{\alpha(t-s)} dB_s dB_t = \int_0^1 \int_0^t e^{\alpha(T(s-t))} dB_s dB_t(T) \quad \overset{D}{\to} \quad N(0, \begin{pmatrix} \frac{1}{2\alpha} & 0 \\ 0 & \sigma^2_\phi \end{pmatrix}),
\]
where \(N(0, A)\) denotes a bivariate normal distribution with mean vector 0 and covariance matrix \(A\).

**Proof.** Application of the time change formula for stochastic integrals twice, cf. Øksendal [35] (Theorem 8.5.7, p. 156), for \(g(\tau) := T \tau, \ g'(\tau) = T\), results in
\[
\frac{1}{\sqrt{T}} \int_0^T \int_0^t e^{\alpha(t-s)} dB_s dB_t = \int_0^1 \int_0^t e^{\alpha(T(s-t))} dB_s dB_t(T)
\]
\[
= \sqrt{T} \int_0^1 \int_0^t e^{\alpha T(s-t)} dB_s dB_t(T)
\]
where \(B_t^{(T)} = \frac{1}{\sqrt{T}} B_{Tt}\). Therefore, it is sufficient to study the asymptotic distribution of
\[
\sqrt{T} \int_0^1 \int_0^t e^{\alpha T(s-t)} dB_s dB_t
\]
where \((W_t)_{t \geq 0}\) denotes a Brownian motion with the same distribution as \((B_t^{(T)})_{t \geq 0}\). The symmetrization theorem for double Wiener integrals, cf. Kuo [29] (Theorem 9.2.8, p. 154), provides the identity
\[
\sqrt{T} \int_0^1 \int_0^t e^{\alpha T(s-t)} dB_s dB_t = \frac{\sqrt{T}}{2} \int_0^1 \int_0^t e^{-\alpha T|s-t|} dB_s dB_t.
\]
By Lemma 3.16 and Remark 3.17 one obtains
\[
\sqrt{T} \int_0^1 \int_0^1 e^{-\alpha T|s-t|} dW_s dW_t = \sum_{j=1}^{\infty} \lambda_{T,j} (\xi_{T,j}^2 - 1)
\]  
(3.31)

where \((\lambda_{T,j})_{j\in\mathbb{N}}\) is the set of eigenvalues of the integral operator with kernel \(f_T(s, t) = \sqrt{T} e^{-\alpha T|s-t|}\) and where \(\xi_{T,j} = \int_0^1 e_{T,j}(t) dW_t\). Here \(e_{T,j}(t)\) denotes the eigenfunction associated with the eigenvalue \(\lambda_{T,j}\).

By Lemma 3.18 the eigenvalues have the properties
\[
\lim_{T \to \infty} \sum_{j=1}^{\infty} \lambda_{T,j}^2 = \frac{1}{\alpha}
\]
and
\[
\lim_{T \to \infty} \max_{j \geq 1} |\lambda_{T,j}| = 0.
\]

Define \(\xi_T := \frac{1}{\sqrt{T}} \int_0^T \varphi(t) dB_t\). Note that \(\xi_T, \xi_{T,j}, j \geq 1\) are jointly normally distributed and that \((\xi_{T,j})_{j \geq 1}\) are i.i.d. standard normally distributed random variables. Projecting \(\xi_T\) onto the space spanned by the random variables \((\xi_{T,j})_{j \geq 1}\), one can write
\[
\xi_T = \xi_{T,0} + \sum_{j=1}^{\infty} \alpha_{T,j} \xi_{T,j}
\]
(3.33)

where \(\xi_{T,0}\) is independent of \((\xi_{T,j})_{j \geq 1}\). Define \(\sigma_T^2 := \text{Var}(\xi_T) = \frac{1}{T} \int_0^T \varphi^2(t) dt\) and \(\sigma_{T,0}^2 := \text{Var}(\xi_{T,0})\) and note that
\[
\sigma_T^2 = \sigma_{T,0}^2 + \sum_{j=1}^{\infty} \alpha_{T,j}^2 \sigma_{\varphi}^2 \rightarrow \sigma_{\varphi}^2.
\]
(3.34)

The Cramér-Wold device will be applied now to prove convergence of the joint distribution of \(\xi_T\) and \(\sum_{j=1}^{\infty} \lambda_{T,j} (\xi_{T,j}^2 - 1)\). Let \(\mu_1, \mu_2 \in \mathbb{R}\); the aim is to prove that
\[
\mu_1 \xi_T + \mu_2 \sum_{j=1}^{\infty} \lambda_{T,j} (\xi_{T,j}^2 - 1) \xrightarrow{D} N(0, \mu_1^2 \sigma_{\varphi}^2 + 2 \mu_2^2 \frac{1}{\alpha}).
\]

In order to do so, compute the characteristic function of the left hand side and note that by plugging in (3.33) the relation
\[
\mu_1 \xi_T + \mu_2 \sum_{j=1}^{\infty} \lambda_{T,j} (\xi_{T,j}^2 - 1) = \mu_1 \xi_{T,0} + \sum_{j=1}^{\infty} (\mu_1 \alpha_{T,j} \xi_{T,j} + \mu_2 \lambda_{T,j} (\xi_{T,j}^2 - 1))
\]
is obtained. If \(Z\) is standard normally distributed, the characteristic function of \(aZ + b(Z^2 - 1)\) is given by
\[
\psi(t) = (1 - 2ibt)^{-1/2} \exp \left( -ibt - \frac{a^2 t^2}{2(1 - 2ibt)} \right).
\]
Thus the characteristic function of \(\mu_1 \xi_{T,0} + \sum_{j=1}^{\infty} (\mu_1 \alpha_{T,j} \xi_{T,j} + \mu_2 \lambda_{T,j} (\xi_{T,j}^2 - 1))\) equals
\[
\psi_T(t) = e^{-\frac{i}{2} \mu_2^2 \sigma_{\varphi}^2 t^2} \prod_{j=1}^{\infty} \left( 1 - 2i\mu_2 \lambda_{T,j} t \right)^{-1/2} \exp \left( -i\mu_2 \lambda_{T,j} t - \frac{(\mu_1 \alpha_{T,j})^2 t^2}{2(1 - 2i\mu_2 \lambda_{T,j} t)} \right).
\]
Taking logarithms, using Taylor expansion and the convergences stated in (3.32) and (3.34) provide

\[
\log \psi_T(t) = -\frac{1}{2} \mu_1^2 \sigma_{\varphi}^2 \tau^2 - \sum_{j=1}^{\infty} \left(\frac{1}{2} \log(1 - 2i\mu_2 \lambda_{T,j} \tau t) + i\mu_2 \lambda_{T,j} \tau t + \frac{\mu_1^2 \alpha_{T,j}^2 \tau^2}{2(1 - 2i\mu_2 \lambda_{T,j} \tau t)}\right)
\]

\[
= -\frac{1}{2} \mu_1^2 \sigma_{\varphi}^2 \tau^2 - \sum_{j=1}^{\infty} (\mu_2^2 \lambda_{T,j}^2 + \frac{1}{2} \mu_1^2 \alpha_{T,j}^2) \tau^2 + o(1)
\]

\[
= -\frac{1}{2} \left(\mu_1^2 \sigma_{\varphi}^2 + \sum_{j=1}^{\infty} \mu_1^2 \alpha_{T,j}^2 + 2 \sum_{j=1}^{\infty} \mu_2^2 \lambda_{T,j}^2\right) \tau^2 + o(1)
\]

\[
\rightarrow -\frac{1}{2} \left(\mu_1^2 \sigma_{\varphi}^2 + \mu_2^2 \alpha_{\varphi}\right) \tau^2.
\]

Note that the right hand side is the logarithm of the characteristic function of a normal distribution with mean 0 and variance \(\mu_1^2 \sigma_{\varphi}^2 + \mu_2^2 \alpha_{\varphi}\). Since

\[
\frac{1}{\sqrt{T}} \int_0^T \int_0^{\tau} e^{\alpha(s-t)} dB_s dB_t = \frac{1}{2} \sum_{j=1}^{\infty} \lambda_{T,j}(\xi_{T,j}^2 - 1),
\]

cf. equations (3.30) and (3.31), the asymptotic variance \(\frac{1}{2\pi}\) stated in the assertion is immediate.

\[\square\]

**Lemma 3.16.** Let \(f : [0,1]^2 \to \mathbb{R}\) be a symmetric, continuous and positive semidefinite kernel and let \((\lambda_i)_{i \geq 1}\) and \((e_i(t))_{i \geq 1}\) denote the set of eigenvalues and corresponding eigenfunctions of the integral operator \(G_f : L^2[0,1] \to L^2[0,1]\) with kernel \(f\), i.e. \(G_f g(x) = \int_0^1 g(y) f(x,y) dy\). Then

\[
\int_0^1 \int_0^1 f(s,t) dW_s dW_t = \sum_{i=1}^{\infty} \lambda_i(\xi_i^2 - 1),
\]

where

\[
\xi_i = \int_0^1 e_i(t) dW_t.
\]

The random variables \((\xi_i)_{i \in \mathbb{N}}\) are independent and standard normally distributed random variables.

**Proof.** Since the kernel \(f\) is continuous and symmetric the operator \(G_f\) is self-adjoint and compact. By Mercer’s Theorem it holds that the kernel can be represented as

\[
f(s,t) = \sum_{i=1}^{\infty} \lambda_i e_i(s) e_i(t)
\]

where \(\lambda_i\) and \(e_i, i \in \mathbb{N}\), are the eigenvalues and eigenfunctions of the integral operator \(G_T\), i.e.

\[
\int_0^1 f(s,t) e_i(s) ds = \lambda_i e_i(t), \quad i \in \mathbb{N}.
\]

Moreover, it holds that the functions \(e_i, i \in \mathbb{N}\), form an orthonormal basis of \(L^2[0,1]\). Define the random variables

\[
\xi_i := \int_0^1 e_i(t) dW_t, \quad i \in \mathbb{N},
\]
and note that $(\xi_i)_{i \geq 1}$ is an i.i.d. sequence of standard normally distributed random variables. It follows by (3.35) that

$$\int_0^1 \int_0^1 f(s, t) dW_s dW_t = \sum_{i=1}^{\infty} \lambda_i \int_0^1 \int_0^1 e_i(s) e_i(t) dW_s dW_t = \sum_{i=1}^{\infty} \lambda_i (\xi_i^2 - 1).$$

The last equality follows by Itô's Theorem which states that

$$\int_0^1 \int_0^1 e_i(s) e_i(t) dW_s dW_t = H_2 \left( \int_0^1 e_i(t) dW_t \right),$$

where $H_2$ is the second Hermite polynomial, i.e. $H_2(x) = x^2 - 1$.

Consider now the kernel $f_T : [0, 1]^2 \to \mathbb{R}$ defined by

$$f_T(s, t) = \sqrt{T} e^{-\alpha T |s-t|}, \quad s, t \in [0, 1]. \tag{3.36}$$

**Remark 3.17.** Note that the kernel defined in (3.36) fulfills the requirements of Lemma 3.16: It is obviously symmetric and continuous. Further, the function $t \mapsto e^{-\alpha T |t|}$ is the characteristic function of the Cauchy distribution with location parameter 0 and scale $\alpha T$. Since a characteristic function is in general positive semidefinite it is immediate that the kernel $f_T$ is positive semidefinite, i.e. it holds for arbitrary real numbers $t_1, \ldots, t_n$ and complex numbers $z_1, \ldots, z_n$, $n \in \mathbb{N}$, that

$$\sum_{k=1}^n \sum_{l=1}^n e^{-\alpha T |t_k-t_l|} z_l \bar{z}_k \geq 0.$$ 

**Lemma 3.18.** Let $(\lambda_{T,i})_{i \geq 1}$ denote the set of eigenvalues of the integral operator with kernel (3.36). Then

$$\lim_{T \to \infty} \sum_{i=1}^{\infty} \lambda_{T,i}^2 = \frac{1}{\alpha}$$

and

$$\lim_{T \to \infty} \max_{i \geq 1} |\lambda_{T,i}| = 0.$$

**Proof.** The operator $G_{f_T}$ is self-adjoint and bounded so that its eigenvalues are real-valued, and

$$\max_{i \geq 1} \lambda_{T,i}^2 = \sup_{g \in L^2([0,1]):\|g\|=1} \|G_{f_T} g\|^2$$

where $\| \cdot \|$ denotes the standard $L^2$-norm on $[0, 1]$. By an equality in Lax [32] (Theorem 2, p. 176) it is

$$\sup_{g \in L^2([0,1]):\|g\|=1} \|G_{f_T} g\| \leq \sup_{t \in [0,1]} \int_0^1 |f_T(s, t)| ds.$$

Simple integration yields

$$\int_0^1 |f_T(s, t)| ds = \int_0^1 \sqrt{T} e^{-\alpha T |s-t|} ds = \frac{1}{\alpha \sqrt{T}} (2 - e^{-\alpha T} - e^{-\alpha T} (1-t)),$$

and thus it follows that

$$\max_{i \geq 1} \lambda_{T,i} \leq \frac{2}{\alpha \sqrt{T}} \to 0 \quad \text{as} \quad T \to \infty.$$
The assertion of Mercer’s Theorem given in (3.35) and the orthonormality of the eigenvalues provide the identity

\[
\sum_{i=1}^{\infty} \lambda_{T,i}^2 = \int_0^1 \int_0^1 f_T(s, t)^2 \, ds \, dt = \frac{1}{2\alpha} \left( 2 + \frac{1}{T} \left( e^{-2\alpha T} - 1 \right) \right)
\]

where the last equality is obtained by simple integration.
Chapter 4

A Change Point Problem

The problem of detecting a change in the drift parameters of a generalized time-inhomogeneous Ornstein-Uhlenbeck is approached in the following. The process of interest is defined by the stochastic differential

\[ dX_t = (L(t) - \alpha X_t)dt + \sigma dB_t, \]

see also Section 3.1, and is assumed to be observed in continuous time. The likelihood ratio test is applied for which an explicit representation of the test statistic can be derived if \( L(t) \) is modeled as a finite sum of known basis functions which are linear in the parameter, i.e.

\[ L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t). \]

The main result is that in the case of a periodic mean reversion mechanism, see Section 3.3, the test statistic converges in distribution, under the null hypothesis of no change in \( \theta = (\mu_1, \ldots, \mu_p, \alpha)^t \), to the squared Euclidean length of a multi-dimensional Brownian bridge.

4.1 Change in the Drift

The problem of testing for an abrupt variation of the parameters in a stochastic model has been an important issue in statistical inference for a long time. Initially investigated in the classical i.i.d. case, change point analysis has been extended to the mean and covariance pattern of times series since structural changes often occur in economic models due to changes in policy or social events. Diffusion processes, originally serving as models for the physical process of diffusion, are the most popular continuous time stochastic processes being applied in fields like finance, physics and engineering. Although statistical inference for diffusion processes that are observed in continuous time has drawn many researchers’ attention change point testing for those processes is not a well-established research area yet. For a general review of change point analysis, see Csörgő and Horváth [16].

In this chapter the primary interest in practice is investigated, namely to test whether or not a change has occurred in a given, time-continuous sample. If a change has been detected, the likelihood ratio test statistic can be used to estimate the location of the most likely change. Another important problem is the sequential setting where the objective is to find the change point as soon as possible in order to correct the calibration of the parameters. This interesting issue is beyond the scope of this work and might be a task of future study.
Consider the mean-reverting Ornstein-Uhlenbeck process introduced in (3.1), that is the diffusion process satisfying the stochastic differential
\[ dX_t = (L(t) - \alpha X_t)dt + \sigma dB_t, \quad t \geq 0, \quad X_0 = \zeta, \] (4.1)
where the constants \( \alpha \) and \( \sigma \) are positive and \( \zeta \) with \( \mathbb{E}(\zeta^2) < \infty \) is a real-valued random variable which is independent of the standard Brownian motion \((B_t)_{t \geq 0}\). The mean reversion function \( L(t) \) is non-random. Obviously, the model (4.1) is more general than the ordinary Ornstein-Uhlenbeck process, see Ornstein and Uhlenbeck [36], since it allows for time-dependent dynamics of the mean level.

Moreover, the parametric model for the mean reversion function \( L(t) \) as proposed in (3.2) of the form
\[ L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t) \] (4.2)
with known functions \( \varphi_1(t), \ldots, \varphi_p(t) \) is regarded. The unknown parameter vector is
\[ \theta = (\mu_1, \ldots, \mu_p, \alpha)^t \in \{(x_1, \ldots, x_{p+1})^t \in \mathbb{R}^{p+1} | x_{p+1} > 0\}. \]

As already explained in Section 3.1, the diffusion parameter \( \sigma \) may be supposed to be known since the statistical inference considered here is based on a time-continuous sample.

Note that it is implicitly assumed in the sequel that there exists a unique solution of equation (4.1). Sufficient conditions on the drift term, here
\[ S(t, X_t, \theta) = \sum_{i=1}^{p} \mu_i \varphi_i(t) - \alpha X_t, \]
can be found in e.g. Kuo [29] (Theorem 10.3.5, p. 192), compare also Section 1.2 and 3.1.

The interesting issue is testing whether or not there is a change in the values of the parameter \( \theta \). The process is observed continuously in the time interval \([0, T]\). In the first step of the following investigation, the test statistic will be formulated for a fixed, suspect change point \( \tau := sT, s \in (0, 1) \). Note that in this setting the ratio of the pre- and post-change point sample sizes does not change when \( T \) increases. The aim is to avoid the case where the smaller of those sample sizes is too small compared to the total sample size so that the change point cannot be detected by the test. Then, in the second step, the test statistic for all possible change points \( \{\tau : s \in (0, 1)\} \) will be established.

For the present fix a known, suspect change point \( \tau \) and rewrite the parametric mean reversion model to
\[ dX_t = \left( S(t, X_t, \theta) 1_{\{t \leq \tau\}} + S(t, X_t, \theta') 1_{\{t > \tau\}} \right) dt + \sigma dB_t, \quad 0 \leq t \leq T. \] (4.3)
Here, \( 1_A \) denotes the indicator function of the set \( A \). The test problem of interest can be formulated as
\[ H_0 : \theta = \theta' \text{ (no change)} \quad \text{vs.} \quad H_A : \theta \neq \theta' \text{ (change at time point } \tau \). \] (4.4)
In what follows, the widely used generalized likelihood ratio test will be studied.

Note that the likelihood ratio test statistic to the change point problem formulated in (4.3) and (4.4) depends on the suspect change point \( \tau = sT, s \in (0, 1) \). Since in practice the crucial
4.2. GENERALIZED LIKELIHOOD RATIO TEST

The question is whether there exists some point \( \tau \) at which a change occurs in the data set, the supremum of the test statistic over all possible change points \( \tau \) will be taken as the actual test statistic. The asymptotic behavior of that supremum under the null hypothesis will be investigated.

**Remark 4.1.** Denote by \( \Lambda_T \) the test statistic to the test problem described above. The asymptotic results on \( \Lambda_T \) under the null hypothesis of no change can be used to construct, by means of a Monte Carlo simulation, a test decision function \( \phi_T = 1_{\{\Lambda_T \geq c_\varepsilon\}} \) where \( c_\varepsilon = c_\varepsilon(T) \) is determined by the condition

\[
\lim_{T \to \infty} E_0(\phi_T) = \lim_{T \to \infty} P_0(\Lambda_T \geq c_\varepsilon) = \varepsilon.
\]

The subscript 0 indicates that the expectation and the probability are computed under the null hypothesis.

### 4.2 Generalized Likelihood Ratio Test

Denote by \( P_X \) the measure induced by the observable realizations \( X^T = \{X_t, 0 \leq t \leq T\} \) on the measurable space \( (C[0, T], B[0, T]) \), \( C[0, T] \) being the space of continuous, real-valued functions on \([0, T]\) and \( B[0, T] \) the associated Borel \( \sigma \)-field. Moreover, let \( P_B \) be the measure generated by the Brownian motion on \((C[0, T], B[0, T])\). Then the likelihood function of observations \( X^T \) of the process with stochastic differential (4.3) is defined as the Radon-Nikodym derivative

\[
L(\theta, \theta', X^T) = \frac{dP_X}{dP_B}(X^T).
\]

The likelihood ratio \( \mathcal{R}(X^T) \) in the setting presented here takes the form

\[
\mathcal{R}(X^T) = \sup_{\theta=\theta'} \frac{L(\theta, \theta', X^T)}{L(\theta, \theta', X^T)}.
\]

As usual in many situations of application of the likelihood ratio the transformed ratio \( \Lambda_T \) defined by

\[
\Lambda_T(s) := -2 \ln(\mathcal{R}(X^T))
\]

is regarded as the essential test statistic since it will turn out that its representation is more convenient than the one of the original likelihood ratio. Note that the notation in (4.6) highlights the dependence of \( \Lambda_T \) upon \( s \) determining the unknown change point \( \tau = sT \).

In order to derive an explicit representation of \( \Lambda_T(s) \) one needs to maximize the numerator and denominator of (4.5). This is done by the corresponding maximum likelihood estimators. Recall that the likelihood function of a general diffusion process

\[
dX_t = S(t, X_t, \theta)dt + \sigma dB_t, \quad 0 \leq t \leq T,
\]

is given by

\[
L(\theta, X^T) = \frac{dP_X}{dP_B}(X^T) = \exp \left( \frac{1}{\sigma^2} \int_0^T S(t, X_t, \theta) dX_t - \frac{1}{2\sigma^2} \int_0^T S(t, X_t, \theta)^2 dt \right)
\]

(4.7)
if the condition

$$P \left( \int_0^T S(t, X_t, \theta)^2 dt < \infty \right) = 1$$

for all \(0 \leq T < \infty\) and all \(\theta\) is fulfilled, cf. Section 1.3 or Section 3.2. The maximum likelihood estimator \(\hat{\theta}_{ML}\) is defined by the relation

$$\max_{\theta \in \Theta} L(\theta, X^T) = L(\hat{\theta}_{ML}, X^T). \quad (4.8)$$

In the case of the diffusion (4.1) provided with the drift

$$S(t, X_t, \theta) = \sum_{i=1}^p \mu_i \varphi_i(t) - \alpha X_t$$

the maximum likelihood estimator \(\hat{\theta}_{ML}\) can be represented by Proposition 3.1 as

$$\hat{\theta}_{ML} = \frac{Q_T^{-1} P_T}{R(X^T)} \quad (4.12)$$

where

$$Q_T = \left( \begin{array}{cc} G_T & -a_T \\ -a_T^T & b_T \end{array} \right) \in \mathbb{R}^{(p+1) \times (p+1)}$$

and

$$P_T = \left( \begin{array}{c} \int_0^T \varphi_1(t) dX_t, \ldots, \int_0^T \varphi_p(t) dX_t - \int_0^T X_t dX_t \end{array} \right)^T \in \mathbb{R}^{p+1}$$

with the integrals of \(\varphi_i(t)\) and the quadratic variation of \(X_t\) given by

$$G_T = \left( \begin{array}{cc} \int_0^T \varphi_j(t) \varphi_k(t) dt & 1 \leq j, k \leq p \end{array} \right) \in \mathbb{R}^{p \times p}, \quad a_T = (\int_0^T \varphi_1(t) X_t dt, \ldots, \int_0^T \varphi_p(t) X_t dt)^T$$

and

$$b_T = \int_0^T X_t^2 dt.$$

Due to the linearity of the drift term in (4.3) the log-likelihood function of the process satisfying (4.3) factorizes and is thus given, compare (4.7), by

$$\ln(\mathcal{L}(\theta, \theta^\prime, X^T)) = \frac{1}{\sigma^2} \left( \int_0^T S(t, X_t, \theta) dX_t + \int_\tau^T S(t, X_t, \theta^\prime) dX_t \right)$$

$$- \frac{1}{2\sigma^2} \left( \int_0^T S(t, X_t, \theta)^2 dt + \int_\tau^T S(t, X_t, \theta^\prime)^2 dt \right).$$

Hence, the likelihood ratio (4.5) can be written by denoting \(X^{\tau,T} = \{X_t, \tau \leq t \leq T\}\) as

$$R(X^T) = \sup_{\theta^\prime} \frac{\mathcal{L}(\theta, X^T)}{\mathcal{L}(\hat{\theta}_{ML}, X^T) \mathcal{L}(\hat{\theta}_{ML}, X^{\tau,T})} \quad (4.13)$$

where \(\mathcal{L}(\theta, X^T)\) is given in (4.7) with drift function specified in (4.9). The expressions \(\mathcal{L}(\theta^*, X^{\tau})\) and \(\mathcal{L}(\theta^*, X^{\tau}, T)\) contain analog integrals with integration regions \([0, \tau]\) and \([\tau, T]\), respectively. It follows from (4.8) that

$$R(X^T) = \frac{\mathcal{L}(\hat{\theta}_{ML}^{\tau}, X^T)}{\mathcal{L}(\hat{\theta}_{ML}^{\tau}, X^{\tau,T})}$$

where the maximum likelihood estimates \(\hat{\theta}_{ML}^{\tau}\) and \(\hat{\theta}_{ML}^{\tau,T}\) are computed from the total, the pre- and post-change sample, respectively. This representation of the likelihood ratio is used to prove the following proposition. The detailed proof is established in Section 4.4.
Proposition 4.2. The transformed likelihood ratio test statistic $\Lambda_T(s) = -2\ln(R(X^T))$ where $R(X^T)$ is given in (4.5) can be represented under the null hypothesis as

$$\Lambda_T(s) = -R_T^{t}Q_T^{-1}R_T + R_T^{t}Q_T^{-1}R_s + (R_T - R_s)^t(Q_T - Q_s)^{-1}(R_T - R_s)$$

where $Q_T$ is specified in (4.10)

$$R_T = \left(\int_0^T \varphi_1(t)dB_t, \ldots, \int_0^T \varphi_p(t)dB_t, -\int_0^T X_tdB_t\right)^t.$$

4.3 Asymptotic Behavior of the Test Statistic

For the rest of this investigation the periodic Ornstein-Uhlenbeck model proposed in Section 3.3 which is characterized by periodicity of the functions

$$\varphi_j(t + \nu) = \varphi_j(t)$$

(4.14)
is considered. Thereby, $\nu$ denotes the known period. Under the null hypothesis of no change the mean reversion function $L(t) = \sum_{i=1}^p \mu_i\varphi_i(t)$ is periodic, that is $L(t + \nu) = L(t)$.

Again, it is assumed that the process is observed over some multiple of periods, i.e.

$$T = n\nu, \quad n \in \mathbb{N},$$

and Gram-Schmidt orthogonalization justifies the assumption that $\varphi_1(t), \ldots, \varphi_p(t)$ satisfy

$$\int_0^\nu \varphi_j(t)\varphi_k(t)dt = \begin{cases} \nu, & j = k \\ 0, & j \neq k. \end{cases}$$

(4.15)

Under these assumptions, the matrix $Q_T$ appearing in the test statistic $\Lambda_T(s)$, see Proposition 4.2, takes the form

$$Q_T = \left(\begin{array}{c|c} TI_{p \times p} & a_T \\ \hline a_T^t & b_T \end{array}\right).$$

In practice, the change point $\tau = sT$ is typically unknown such that the practitioner needs to find out whether or not a change has occurred at some time in the data. It is natural to consider the maximal, observable likelihood ratio. This leads to the test statistic

$$\sup_s \Lambda_T(s).$$

One possibility to study the asymptotic behavior of this supremum is to prove weak convergence of the process $(\Lambda_T(\cdot))_T$. Weak convergence of the supremum then follows by the continuous mapping theorem. The proof of the following result is relegated to the last section.

Theorem 4.1. Let $X^T = \{X_t, 0 \leq t \leq T\}$ be observations of the mean reversion process (4.3) satisfying (4.14) and (4.15). Denote by $\Lambda_T(s) = -2\ln(R(X^T))$ the transformed likelihood ratio test statistic to the test problem (4.4). Then it holds under the null hypothesis that

$$\sup_{s \in [s_1, s_2]} \Lambda_T(s) \Rightarrow \sup_{s \in [s_1, s_2]} \left\|W(s) - sW(1)\right\|^2_s \frac{1}{(1 - s)}$$
as $T \to \infty$, $0 < s_1 < s_2 < 1$. Thereby, $\| \cdot \|$ denotes the Euclidean norm and $W$ is a $(p + 1)$-dimensional standard Brownian motion.
Remark 4.3. The results of this chapter comply with the properties of the likelihood ratio test that detects a change point in the parameters of an autoregressive process which is a discrete time version of the Ornstein-Uhlenbeck process. The test statistic given in Proposition 4.2 is the continuous-time analog of the likelihood ratio test statistic of the corresponding test for an autoregressive process of order 1, see Davis et al. [17]. This compliance is not trivial even though the Euler discretization of the mean reversion process (4.1) has an autoregressive structure, namely

\[ X_{t_{i+1}} = L(t_i) \Delta t + (1 - \alpha \Delta t) X_{t_i} + \varepsilon_{t_{i+1}} \]  

(4.16)

where \( \Delta t = t_{i+1} - t_i \). But the time-continuous mean reversion process can be approximated by its discrete-time version (4.16) only if the time steps \( \Delta t \) are small. However, the autoregressive coefficient \((1 - \alpha \Delta t)\) is then close to 1 which may lead to a violation of the causality condition which assures stationarity of an AR(1) process and which is a necessary condition in Davis et al. [17]. Moreover, the limit in Theorem 4.1 is also the same as in the autoregressive case studied in Davis et al. [17]. However, this equality is not very surprising since in many situations in extreme value theory Brownian bridge is obtained as the limit of some extreme value statistic.

\section*{4.4 Proof}

This section is mainly devoted to the proof of Theorem 4.1 which requires some auxiliary results. First of all, the proof of Proposition 4.2 is completed.

\textit{Proof of Proposition 4.2.} The aim is to compute an explicit expression of the ratio given in (4.13). Making use of the particular form of the drift function

\[ S(t, X_t, \theta) = \sum_{i=1}^{p} \mu_i \varphi_i(t) - \alpha X_t \]

yields

\[ \int_0^T S(t, X_t, \theta) dX_t = \theta^T P_T \]

where \( P_T = \begin{pmatrix} \int_0^T \varphi_1(t) dX_t, \ldots, \int_0^T \varphi_p(t) dX_t, -\int_0^T X_t dX_t \end{pmatrix} \) and

\[ \int_0^T S(t, X_t, \theta)^2 dt = \theta^T Q_T \theta \]

where

\[ Q_T = \begin{pmatrix} G_T & -a_T \\ -a_T^T & b_T \end{pmatrix}, \]

\[ G_T = (\int_0^T \varphi_j(t) \varphi_k(t) dt)_{1 \leq j, k \leq p}, \quad a_T = (\int_0^T \varphi_1(t) X_t dt, \ldots, \int_0^T \varphi_p(t) X_t dt)^T \]

and \( b_T = \int_0^T X_t^2 dt \). Hence the likelihood function in the numerator of (4.12)

\[ \mathcal{L}(\theta, X^T) = \exp \left( \frac{1}{\sigma^2} \int_0^T S(t, X_t, \theta) dX_t - \frac{1}{2\sigma^2} \int_0^T S(t, X_t, \theta)^2 dt \right) \]

can written as

\[ \mathcal{L}(\theta, X^T) = \exp \left( \frac{1}{\sigma^2} \theta^T P_T - \frac{1}{2\sigma^2} \theta^T Q_T \theta \right). \]
Recall that $\hat{\theta}^T_{ML} = Q_T^{-1}P_T$ such that
\[ L(\hat{\theta}^T_{ML}, X^T) = \exp \left( \frac{1}{2\sigma^2} P_T^T Q_T^{-1} P_T - \frac{1}{2\sigma^2} P_T^T Q_T^{-1} P_T \right) \]
\[ = \exp \left( \frac{1}{2\sigma^2} P_T^T Q_T^{-1} P_T \right). \tag{4.17} \]

Denote by $\theta$ the true value of the parameter under the null hypothesis, i.e. $\theta = \theta'$. Plugging in equation (3.21) which states that $P_T = Q_T \theta + \sigma R_T$ into representation (4.17) leads to
\[ L(\hat{\theta}^T_{ML}, X^T) = \exp \left( \frac{1}{2\sigma^2} (Q_T \theta + \sigma R_T)^T Q_T^{-1} (Q_T \theta + \sigma R_T) \right) \]
\[ = \exp \left( \frac{1}{2\sigma^2} (\theta' Q_T \theta + \sigma \theta' R_T + \sigma R_T^T \theta + \sigma^2 R_T^T Q_T^{-1} R_T) \right) \]
\[ = \exp \left( \frac{1}{2\sigma^2} \theta'^T Q_T \theta + \frac{1}{\sigma} R_T^T \theta + \frac{1}{2} R_T^T Q_T^{-1} R_T \right). \]

The same procedure yields an analog expression for $L(\hat{\theta}^*_{ML}, X^\tau)$ whereby every index $T$ is replaced by $\tau$. Further, the additivity of the integrals provides
\[ \hat{\theta}^{r,T}_{ML} = (Q_T - Q_\tau)^{-1}(P_T - P_\tau) \]
and
\[ P_T - P_\tau = (Q_T - Q_\tau)\theta + \sigma(R_T - R_\tau) \]
such that
\[ L(\hat{\theta}^{r,T}_{ML}, X^{\tau,T}) = \exp \left( \frac{1}{2\sigma^2} \theta'^T (Q_T - Q_\tau) \theta + \frac{1}{\sigma} (R_T - R_\tau)^T \theta \right. \]
\[ \left. + \frac{1}{2} (R_T - R_\tau)^T (Q_T - Q_\tau)^{-1} (R_T - R_\tau) \right). \]

It follows from all the preceding representations that under the null hypothesis one has
\[ R(X^T) = \frac{L(\hat{\theta}^T_{ML}, X^T)}{L(\hat{\theta}^T_{ML}, X^\tau)} \frac{L(\hat{\theta}^{r,T}_{ML}, X^{\tau,T})}{L(\hat{\theta}^{r,T}_{ML}, X^{\tau,T})} \]
\[ = \exp \left( \frac{1}{2\sigma^2} \theta'^T Q_T \theta + \frac{1}{\sigma} R_T^T \theta + \frac{1}{2} R_T^T Q_T^{-1} R_T - \frac{1}{2\sigma^2} \theta'^T Q_\tau \theta - \frac{1}{\sigma} R_{\tau}^T \theta - \frac{1}{2} R_{\tau}^T Q_\tau^{-1} R_{\tau} \right) \]
\[ - \frac{1}{2\sigma^2} \theta'^T (Q_T - Q_\tau) \theta - \frac{1}{\sigma} (R_T - R_\tau)^T \theta - \frac{1}{2} (R_T - R_\tau)^T (Q_T - Q_\tau)^{-1} (R_T - R_\tau) \]
\[ = \exp \left( \frac{1}{2} (R_T^T Q_T^{-1} R_T - R_\tau^T Q_\tau^{-1} R_\tau - (R_T - R_\tau)^T (Q_T - Q_\tau)^{-1} (R_T - R_\tau) \right) \]
whereby the last equality is obtained by canceling several terms. Taking logarithm and multiplying by -2 provides the assertion.

Now the asymptotic behavior of $\Lambda_T(s)$ which is represented in Proposition 4.2 as
\[ \Lambda_T(s) = -R_T^T Q_T^{-1} R_T + R_{st}^T Q_s^{-1} R_{st} + (R_T - R_{st})^T (Q_T - Q_{st})^{-1} (R_T - R_{st}) \]
is studied in the case of a periodic mean reversion function, see (4.14) and (4.15). The first term
\[ R_T^t Q_T^{-1} R_T = \frac{1}{\sqrt{T}} R_T^t \left( \frac{1}{T} Q_T \right)^{-1} \frac{1}{\sqrt{T}} R_T \]
is covered by the following corollary which is just a summarization of Proposition 3.11, 3.14 and 3.15.

**Corollary 4.4.** It holds that
\[ \frac{1}{\sqrt{T}} R_T \xrightarrow{D} N(0, \Sigma) \]
and
\[ \frac{1}{T} Q_T \to \Sigma, \text{ almost surely,} \]
as \( T \to \infty \). The matrix \( \Sigma \) is given by
\[ \Sigma = \left( \begin{array}{cc} \nu I_{p \times p} & -\Lambda \\ -\Lambda^t & \omega \end{array} \right) \] (4.18)
where \( \Lambda_i = \int_0^\nu \varphi_i(t) \tilde{h}(t) dt, \ i = 1, \ldots, p, \ \omega = \int_0^\nu (\tilde{h}(t))^2 dt + \frac{\nu \sigma^2}{2 \alpha} \) and where \( \tilde{h} : [0, \infty) \to \mathbb{R} \) is defined by
\[ \tilde{h}(t) = e^{-\alpha t} \sum_{j=1}^p \mu_j \int_{-\infty}^t e^{\alpha s} \varphi_j(s) ds. \] (4.19)

As usual \( N(0, A) \) denotes a normally distributed random vector with zero-mean and covariance matrix \( A \).

Next, the second term of \( \Lambda_T(s) \) is investigated, it can be rewritten to
\[ R_{sT}^t Q_{sT}^{-1} R_{sT} = \frac{1}{\sqrt{T}} R_{sT}^t \left( \frac{1}{T} Q_{sT} \right)^{-1} \frac{1}{\sqrt{T}} R_{sT}. \]
In detail, it will be shown that the process \( \left( \frac{1}{\sqrt{T}} R_{sT} \right)_{s \in [s_1, s_2]} \) converges in distribution to a Gaussian process on \( [s_1, s_2] \) as \( T \to \infty \) and that \( \frac{1}{T} Q_{sT} \) converges in probability uniformly on \( [s_1, s_2] \).

In order to prove weak convergence of a sequence \( (Z_n(\cdot))_{n \geq 1} \) of processes on \( C[0,1] \) the following lemma which is a corollary to Prokhorov’s theorem will be applied. The proof of the lemma can be found in Karatzas and Shreve [27] (Theorem II.4.15, p. 65).

**Lemma 4.5.** Let \( Z_n, \ n \geq 1, \ Z \) be \( C[0,1] \)-valued random variables. If
\[ (Z_n(t_1), \ldots, Z_n(t_m)) \xrightarrow{D} (Z(t_1), \ldots, Z(t_m)) \]
for all \( (t_1, \ldots, t_m) \in [0,1]^m, \ m \in \mathbb{N}, \) and if, in addition, \( (Z_n)_{n \geq 1} \) is tight then
\[ Z_n \xrightarrow{D} Z. \]

The next lemma is a helpful tool for the verification of tightness. Note that a more general assertion with slightly different conditions is proved in Billingsley [7] (Theorem 12.3, p. 95).
4.4. PROOF

**Lemma 4.6.** Let \((Z_n)_{n \geq 1}\) be a sequence of \(C[0,1]\)-valued random variables. If

\[
\sup_{n \geq 1} E|Z_n(0)|^\gamma < \infty \tag{4.20}
\]

and

\[
\sup_{n \geq 1} E|Z_n(t) - Z_n(s)|^\alpha \leq C|t - s|^{1+\beta}, \quad 0 \leq s, t \leq 1, \tag{4.21}
\]

for some positive constants \(\alpha, \beta, \gamma\) and \(C\), then \((Z_n)_{n \geq 1}\) is tight.

**Proof.** Necessary and sufficient conditions for tightness, see Billingsley [7] (Theorem 8.2, p. 55) or Karatzas and Shreve [27] (Theorem II.4.10, p. 63), are the following: It needs to hold that

\[
\lim_{\lambda \to 0} \sup_{n \geq 1} P \left( |Z_n(0)| > \lambda \right) = 0 \tag{4.22}
\]

and that for each \(\eta > 0\)

\[
\lim_{\gamma \to 0} \sup_{n \geq 1} P \left( w(Z_n, \gamma) \geq \eta \right) = 0 \tag{4.23}
\]

where

\[
w(x, \gamma) = \sup_{|s-t| \leq \gamma} |x(s) - x(t)|, \quad 0 < \gamma \leq 1,
\]

is the modulus of continuity of \(x \in C[0,1]\). Note that condition (4.20) implicates (4.22). In order to verify (4.23) define the sequence \((D_m)_{m \geq 1}\) of dyadic grids by

\[
D_m := \left\{ \frac{k}{2^m} : k = 0, 1, \ldots, 2^m \right\}
\]

and the sets

\[
A_m^\gamma := \left\{ \max_{s,t \in D_m, |t-s| < \gamma} |x(s) - x(t)| \geq \gamma \right\}
\]

and

\[
A^\gamma := \left\{ \sup_{s,t \in [0,1], |t-s| < \gamma} |x(s) - x(t)| \geq \gamma \right\}.
\]

It holds that \(A_m^\gamma \not\supset A^\gamma\) as \(m \to \infty\) so that it is sufficient for (4.23) to show that

\[
\lim_{m \to \infty} \sup_{n \geq 1} P \left( w(Z_n, \gamma_m) \geq \eta \right) = 0
\]

for \(\gamma_m = \frac{1}{2^m}\). Making use of Chebychev’s inequality and the condition of the lemma give

\[
P \left( |Z_n(t) - Z_n(s)| \geq \eta \right) \leq \eta^{-\alpha} E|Z_n(t) - Z_n(s)|^\alpha \leq C \eta^{-\alpha} |t - s|^{1+\beta}.
\]

Thus, for \(t = \frac{k}{2^m}, s = \frac{k-1}{2^m}, \eta = 2^{-\delta m}\), where \(0 < \delta < \beta/\alpha\), it is

\[
P \left( \left| Z_n \left( \frac{k}{2^m} \right) - Z_n \left( \frac{k-1}{2^m} \right) \right| \geq 2^{-\delta m} \right) \leq C 2^{\delta m} a 2^{-m(1+\beta)} = C 2^{-m(1+\beta-\delta \alpha)}
\]

and

\[
P \left( \max_{1 \leq k \leq 2^m} \left| Z_n \left( \frac{k}{2^m} \right) - Z_n \left( \frac{k-1}{2^m} \right) \right| \geq 2^{-\delta m} \right) \leq P \left( \bigcup_{k=1}^{2^m} \left\{ \left| Z_n \left( \frac{k}{2^m} \right) - Z_n \left( \frac{k-1}{2^m} \right) \right| \geq 2^{-\delta m} \right\} \right)
\]

\[
\leq \sum_{k=1}^{2^m} P \left( \left| Z_n \left( \frac{k}{2^m} \right) - Z_n \left( \frac{k-1}{2^m} \right) \right| \geq 2^{-\delta m} \right) = C 2^{-m(\beta-\delta \alpha)},
\]
where \((\beta - \delta \alpha) > 1\).

Consider the first \(p\) entries of \(\frac{1}{\sqrt{T}}R_{sT}\) which are of the form

\[
Y_n(s) := \frac{1}{\sqrt{n}} \int_0^{n\nu s} \varphi(u)dB_u, \quad s_1 \leq s \leq s_2,
\]

keeping in mind that \(0 < s_1 \leq s \leq s_2 < 1\), \(T = n\nu\) and that \(\varphi\) denotes one of the periodic functions \(\varphi_1, \ldots, \varphi_p\).

**Proposition 4.7.** It holds for the process defined in (4.24) that

\[
Y_n(\cdot) \overset{D}{\rightarrow} Y(\cdot)
\]

as \(n \to \infty\) where \(Y\) is Brownian motion with covariance function \(\text{Cov}(Y(s), Y(t)) = (s \wedge t)\nu\), where \(s \wedge t = \min(s, t)\).

**Proof.** Since \(\varphi\) is a deterministic function the Itô integral \(Y_n(s)\) is a zero-mean Gaussian process for each \(n\) and

\[
\text{Cov}(Y_n(s), Y_n(t)) = \frac{1}{n} \int_0^{n\nu(s \wedge t)} \varphi^2(u)du
\]

\[
= \frac{1}{n} \left( \sum_{i=1}^M \int_{(i-1)\nu}^{i\nu} \varphi^2(u)du + \int_{M\nu}^{n\nu(s \wedge t)} \varphi^2(u)du \right)
\]

where \(M = \lfloor n(s \wedge t) \rfloor\). Here \(\lfloor x \rfloor\) denotes the integer part of \(x\). It holds for the second summand that

\[
\left| \frac{1}{n} \int_{M\nu}^{n\nu(s \wedge t)} \varphi^2(u)du \right| \leq \frac{1}{n} \int_{M\nu}^{\nu(M+1)} \varphi^2(u)du = \frac{n}{M} \nu \sup_{0 \leq u \leq \nu} \varphi^2(u) \to 0
\]

and for the first one that

\[
\frac{1}{n} \sum_{i=1}^M \int_{(i-1)\nu}^{i\nu} \varphi^2(u)du = \frac{M}{n} \nu \to \nu(s \wedge t)
\]

as \(n \to \infty\). So it has been proved that the expectation and covariance function of the sequence of Gaussian processes \((Y_n)_{n \geq 1}\) converge implicating that the finite-dimensional distributions of the process converge. Moreover, for \(s < t\) the increment

\[
Y_n(t) - Y_n(s) = \frac{1}{\sqrt{n}} \int_{n\nu s}^{n\nu t} \varphi(u)dB_u
\]

is normal with zero-mean and variance equal to

\[
\frac{1}{n} \int_{n\nu s}^{n\nu t} \varphi^2(u)du \leq \sup_{0 \leq u \leq \nu} \varphi^2(u)\nu|t - s|
\]

since \(\varphi\) is periodic and bounded. Thus it holds that

\[
\sup_{n \geq 1} \text{E}[Y_n(t) - Y_n(s)]^{2k} \leq C|t - s|^k, \quad 0 \leq s, t \leq 1,
\]

for \(k \in \mathbb{N}\) and positive constant \(C\). Consequently, the condition for tightness given in Lemma 4.6 is satisfied and the convergence of the process follows by Lemma 4.5. Since the limit process is continuous with covariance function of the form \(\nu(s \wedge t)\), it is a Brownian motion.
4.4. PROOF

Turn to the last entry of $\frac{1}{\sqrt{T}}R_{sT}$ and consider therefor

$$V_n(s) := \frac{1}{\sqrt{n}} \int_0^{n\nu s} X_0 dB_u, \quad s_1 \leq s \leq s_2.$$  

(4.25)

Before studying the finite-dimensional distributions of $V_n(\cdot)$, note that by Lemma 3.8 the solution of the stochastic differential equation (4.1) is given by

$$X_t = e^{-at}X_0 + h(t) + Z_t$$  

(4.26)

where

$$h(t) = e^{-at} \int_0^t e^{as} L(s) ds = e^{-at} \sum_{i=1}^p \mu_i \int_0^t e^{as} \varphi_i(s) ds$$

and

$$Z_t = \sigma e^{-at} \int_0^t e^{as} dB_s.$$  

Proposition 4.8. The finite-dimensional distributions of $V_n(\cdot)$ defined in (4.25) converge. In detail, it is

$$(V_n(t_1), \ldots, V_n(t_k)) \xrightarrow{D} (V(t_1), \ldots, V(t_k))$$

for all $(t_1, \ldots, t_k) \in [0,1]^k$, $k \in \mathbb{N}$. Thereby, $V(\cdot)$ is a zero-mean Gaussian process with covariance function

$$\text{Cov}(V(s), V(t)) = (s \wedge t) \left( \int_0^\nu \tilde{h}(u)^2 du + \frac{\nu g^2}{2\alpha} \right), \quad 0 \leq s, t \leq 1,$$

where the function $\tilde{h}$ is given in (4.19).

Proof. Let $s \in [0,1]$ be fixed. By (4.26) it is

$$V_n(s) = \frac{1}{\sqrt{n}} \int_0^{n\nu s} e^{-au} X_0 dB_u + \frac{1}{\sqrt{n}} \int_0^{n\nu s} h(u)dB_u + \frac{1}{\sqrt{n}} \int_0^{n\nu s} \sigma e^{-au} \int_0^u e^{\alpha r} dB_r dB_u. \quad (4.27)$$

The first summand on the right-hand side of the previous equation converges in probability toward zero since

$$\text{E} \left( \frac{1}{\sqrt{n}} \int_0^{n\nu s} e^{-au} X_0 dB_u \right)^2 = \frac{\text{E}(X_0^2)}{n} \int_0^{n\nu s} e^{-2\alpha u} du \to 0$$

as $n \to \infty$. The second term on the r.h.s. of (4.27) is normal with zero-mean and

$$\text{Cov} \left( \frac{1}{\sqrt{n}} \int_0^{n\nu s} h(u)dB_u, \int_0^{n\nu t} g(u)dB_u \right) = \frac{1}{n} \int_0^{n\nu s \wedge t} \tilde{h}(u)^2 du \to (s \wedge t) \int_0^\nu \tilde{h}(u)^2 du$$

for some fixed $s, t \in [0,1]$. Moreover, the third piece of (4.27) is asymptotically normal, more precisely

$$\frac{1}{\sqrt{n}} \left( \int_0^{n\nu s} \int_0^u e^{\alpha(r-u)} dB_r dB_u, \int_0^{n\nu s} g(u)dB_u \right) \xrightarrow{D} N(0, \begin{pmatrix} \frac{\nu g}{\sigma^2} & 0 \\ 0 & \sigma^2 \end{pmatrix}),$$  

(4.28)
for any $L^2$-function $g$ such that $\sigma_g = \lim_{n \to \infty} \frac{1}{n} \int_0^2 (g(u))^2 du < \infty$, see Proposition 3.15. So it remains to show that the finite-dimensional distributions of the process

\[
\left( \frac{1}{\sqrt{n}} \int_0^{\nu(s)} \int_0^u e^{\alpha(r-u)} dB_r dB_u \right)_{n \geq 1}
\]

converge as $n \to \infty$. In doing so, the investigation is restricted to the two-dimensional case, i.e. it will be proved that

\[
\frac{1}{\sqrt{n}} \left( \int_0^{\nu(s)} \int_0^u e^{\alpha(r-u)} dB_r dB_u, \int_0^{\nu(t)} \int_0^u e^{\alpha(r-u)} dB_r dB_u \right)
\]

covers in distribution for all $s,t \in (0,1)$. The demonstration of the convergence of multidimensional distributions can be treated in an analogous way. The same calculations as in the proof of Proposition 3.15 are carried out: By the time change formula for stochastic integrals, see Øksendal [35] (Theorem 8.5.7, p. 156), applied for the function $g(\tau) := n\nu \tau$, $g'(\tau) = n\nu$, it is

\[
\frac{1}{\sqrt{n}} \int_0^{\nu(s)} \int_0^u e^{\alpha(r-u)} dB_r dB_u = \sqrt{n\nu} \int_0^s \int_0^u e^{\alpha(r-u)} dB_r dB_u
\]

and

\[
\frac{1}{\sqrt{n}} \int_0^{\nu(t)} \int_0^u e^{\alpha(r-u)} dB_r dB_u = \sqrt{n\nu} \int_0^t \int_0^u e^{\alpha(r-u)} dB_r dB_u = \sqrt{n\nu} \int_0^t \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u
\]

where $B_t^{(n)} = \frac{1}{\sqrt{n\nu}} B_{n\nu t}$. Denoting by $(W_t)_{t \geq 0}$ a Brownian motion which is distributed as $(B_t^{(n)})_{t \geq 0}$ and applying Kuo’s [29] symmetrization theorem for double Wiener integrals (Theorem 9.2.8, p. 154) result in

\[
\sqrt{n\nu} \int_0^s \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u dW_r dW_u \overset{D}{\to} N(0, \frac{\nu}{2\alpha} (\lambda_1^2 s + \lambda_2^2 t + 2\lambda_1 \lambda_2 (s \wedge t))).
\]

Assuming without loss of generality that $s < t$, the linear combination (4.30) can be alternatively written as

\[
(\lambda_1 + \lambda_2) \frac{\sqrt{n\nu}}{2} \int_0^s \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u + \lambda_2 \frac{\sqrt{n\nu}}{2} \int_0^t \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u + \lambda_2 A_n
\]

where

\[
A_n = \sqrt{n\nu} \int_0^s \int_0^t e^{-\alpha n \nu(r-u)} dB_r dB_u \overset{P}{\to} 0
\]

since

\[
E(A_n^2) = \frac{n\nu}{4} \int_0^s \int_0^t e^{-2\alpha n \nu(r-u)} du
\]

converge as $n \to \infty$. In doing so, the investigation is restricted to the two-dimensional case, i.e. it will be proved that

\[
\frac{1}{\sqrt{n}} \left( \int_0^{\nu(s)} \int_0^u e^{\alpha(r-u)} dB_r dB_u, \int_0^{\nu(t)} \int_0^u e^{\alpha(r-u)} dB_r dB_u \right)
\]

converges in distribution for all $s,t \in (0,1)$. The demonstration of the convergence of multidimensional distributions can be treated in an analogous way. The same calculations as in the proof of Proposition 3.15 are carried out: By the time change formula for stochastic integrals, see Øksendal [35] (Theorem 8.5.7, p. 156), applied for the function $g(\tau) := n\nu \tau$, $g'(\tau) = n\nu$, it is

\[
\frac{1}{\sqrt{n}} \int_0^{\nu(s)} \int_0^u e^{\alpha(r-u)} dB_r dB_u = \sqrt{n\nu} \int_0^s \int_0^u e^{\alpha(r-u)} dB_r dB_u = \sqrt{n\nu} \int_0^s \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u
\]

where $B_t^{(n)} = \frac{1}{\sqrt{n\nu}} B_{n\nu t}$. Denoting by $(W_t)_{t \geq 0}$ a Brownian motion which is distributed as $(B_t^{(n)})_{t \geq 0}$ and applying Kuo’s [29] symmetrization theorem for double Wiener integrals (Theorem 9.2.8, p. 154) result in

\[
\sqrt{n\nu} \int_0^s \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u dW_r dW_u \overset{D}{\to} N(0, \frac{\nu}{2\alpha} (\lambda_1^2 s + \lambda_2^2 t + 2\lambda_1 \lambda_2 (s \wedge t))).
\]

Assuming without loss of generality that $s < t$, the linear combination (4.30) can be alternatively written as

\[
(\lambda_1 + \lambda_2) \frac{\sqrt{n\nu}}{2} \int_0^s \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u + \lambda_2 \frac{\sqrt{n\nu}}{2} \int_0^t \int_0^u e^{-\alpha n \nu(r-u)} dB_r dB_u + \lambda_2 A_n
\]

where

\[
A_n = \sqrt{n\nu} \int_0^s \int_0^t e^{-\alpha n \nu(r-u)} dB_r dB_u \overset{P}{\to} 0
\]

since

\[
E(A_n^2) = \frac{n\nu}{4} \int_0^s \int_0^t e^{-2\alpha n \nu(r-u)} du
\]
as $n \to \infty$. The two remaining summands in (4.32) are independent due to the disjoint integration regions $[0,s]^2$ and $[s,t]^2$. It is shown in the proof of Proposition 3.15 that each of these summands is asymptotically normal with zero-mean and variance computed according to

$$2(\lambda_1 + \lambda_2)^2 \frac{n\nu^2}{4} \int_0^s \int_0^s e^{-2\alpha \nu|u-v|} dvdu = (\lambda_1 + \lambda_2)^2 \frac{\nu}{2\alpha} \left( s + \frac{1}{2\alpha \nu}(e^{-2\alpha \nu s} - 1) \right)$$

and

$$2\lambda_2^2 \frac{n\nu^2}{4} \int_s^t \int_s^t e^{-2\alpha \nu|u-v|} dvdu = \lambda_2^2 \frac{\nu}{2\alpha} \left( t - s + \frac{1}{2\alpha \nu}(e^{-2\alpha \nu(t-s)} - 1) \right)$$

respectively. Hence the variance of the limit distribution of the linear combination (4.30) can be computed to be

$$(\lambda_1 + \lambda_2)^2 s \frac{\nu}{2\alpha} + \lambda_2^2 (t-s) \frac{\nu}{2\alpha} = \frac{\nu}{2\alpha} (\lambda_1^2 s + 2\lambda_1 \lambda_2 s + \lambda_2^2 t)$$

which is equal to the variance stated in (4.31) in the case $s < t$ considered here.

\[ \square \]

The proof of tightness of $(V_n(\cdot))_{n \geq 1}$, which is a necessary condition for weak convergence in Lemma 4.5, is still an open issue.

Application of Lemma 4.6 yields the following corollary.

**Proposition 4.9.** The sequence $(V_n(\cdot))_{n \geq 1}$ defined in (4.25) is tight.

**Proof.** The conditions of Lemma 4.6 by regarding representation (4.27) have to be verified. Obviously, condition (4.20) is satisfied. The first piece on the r.h.s. of (4.27) is normal with zero-mean and

$$E \left| \frac{1}{\sqrt{n}} \int_{ns}^{nvt} e^{-a\nu X_0 dB_u} \right|^2 \leq \nu E|X_0|^2 e^{-2\alpha \nu s}|t-s| \leq \nu E|X_0|^2 |t-s| < \infty$$

so that

$$\sup_{n \geq 1} E \left| \frac{1}{\sqrt{n}} \int_{ns}^{nvt} e^{-a\nu X_0 dB_u} \right|^{2k} \leq C|t-s|^k, \quad k \in \mathbb{N}.$$ 

That means that this part is tight by Lemma 4.6. The second summand of (4.27) is normal as well with zero-mean and

$$E \left| \frac{1}{\sqrt{n}} \int_{ns}^{nvt} h(u)dB_u \right|^2 = \frac{1}{n} \int_{ns}^{nvt} h(u)^2 du \leq \sup_{u} h(u)^2 \nu |t-s| < \infty$$

since the function $h$ is bounded. The tightness of this term follows by Lemma 4.6 again. Finally, the 4th moment of the increment of the last term in (4.27) is computed by means of Lemma 4.10 which gives for $m = 2$ the relation

$$E \left| \frac{\sigma}{\sqrt{n}} \int_{ns}^{nvt} e^{-a\nu X_0 dB_u} \right|^{4} \leq \frac{36\sigma^4}{n} \nu |t-s| \int_{ns}^{nvt} E \left| e^{-a\nu X_0} \right|^4 du$$

$$= \frac{36\sigma^4}{n} \nu |t-s| \int_{ns}^{nvt} \frac{3}{4\alpha^2} (1 - e^{-2\alpha \nu})^2 du$$

$$= O(|t-s|^2)$$
as the random variable in the expectation is normal, in detail
\[
e^{-\alpha u} \int_0^u e^{\alpha r} dB_r \sim N \left( 0, \frac{1}{2\alpha} (1 - e^{-2\alpha u}) \right).
\]
Hence, the last summand of (4.27) also satisfies condition (4.21) of Lemma 4.6 for \(\alpha = 4\) and \(\beta = 1\) so that it is tight. Since all summands in (4.27) are tight, \((V_n(\cdot))_{n \geq 1}\) is a tight sequence.

The subsequent lemma is stated and proved in Lipster and Shiryayev in [33] (Lemma 4.12, p. 125).

**Lemma 4.10.** Let \(f(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}\) be a non-anticipating function satisfying
\[
\int_0^T f(t, \omega)^2 dt < \infty \text{ a.s. and } E \left( \int_0^T |f(t, \omega)|^{2m} dt \right) < \infty, m \in \mathbb{N}.
\]
Then it holds that
\[
E \left| \int_0^T f(t, \omega) dB_t \right|^{2m} \leq (m(2m - 1))^m T^{m-1} \int_0^T E|f(t, \omega)|^{2m} dt.
\]

Finally the preceding results of this section are collected.

**Corollary 4.11.** It holds that
\[
\frac{1}{\sqrt{T}} R(\cdot) \overset{D}{\rightarrow} R(\cdot)
\]
on \([s_1, s_2]\), as \(T \to \infty\), where \(R\) is a \((p + 1)\)-dimensional Gaussian process with \(R \sim N(0, s\Sigma)\), whereby \(\Sigma\) is specified in (4.18). The covariance function of \(R\) is of the form \(\text{Cov}(R_i(s), R_j(t)) = (s \land t)\Sigma_{ij}, i, j = 1, \ldots, p + 1\).

**Proof.** By Proposition 4.7, 4.8 and 4.9 in combination with Lemma 4.5 each entry of \(\frac{1}{\sqrt{T}} R_s T\) converges on \([s_1, s_2]\) to a Gaussian process. Further, the convergence stated in (4.28) can be extended to
\[
\frac{1}{\sqrt{n}} \left( \int_0^{n s} \int_0^u e^{\alpha(r-u)} dB_r dB_u, \int_0^{n t} \varphi(u) dB_u \right) \overset{D}{\rightarrow} N(0, \begin{pmatrix} \nu(s \land t)/\alpha & 0 \\ 0 & \nu(s \land t) \end{pmatrix})
\]
for all \(s, t \in [s_1, s_2]\), since \(\lim_{n \to \infty} \frac{1}{n} \int_0^{n t} (\varphi(u))^2 du = \nu t\), by making use of the same line of arguments as used in Proposition 3.15. Note that
\[
\text{Cov}(Y_n(s), V_n(t)) = \frac{1}{n} \text{Cov} \left( \int_0^{n s} \varphi(u) dB_u, \int_0^{n t} X_0 dB_u \right) \to (s \land t) \int_0^\nu \varphi(u) \tilde{h}(u) du.
\]

In order to prove convergence of the matrix \(\frac{1}{\sqrt{T}} Q_T\) Lemma 3.9 is applied. It states that the stochastic process \((W_k)_{k \in \mathbb{N}}\) defined by
\[
W_k(s) = \tilde{X}_{(k-1)\nu + s}, \quad 0 \leq s \leq \nu, k \in \mathbb{N},
\]
where
\[
\tilde{X}_t = \tilde{h}(t) + \tilde{Z}_t,
\]
is stationary and ergodic. Hereby \( \tilde{h}(t) \) is given in (4.19),

\[
\tilde{Z}_t = \sigma e^{-at} \int_{-\infty}^{t} e^{as} d\tilde{B}_s
\]

and \((\tilde{B}_s)_{s \in \mathbb{R}}\) denotes a two-sided Brownian motion. Further, by Lemma 3.10

\[
|\tilde{X}_t - X_t| \to 0, \tag{4.33}
\]

almost surely, as \( t \to \infty \).

**Proposition 4.12.** It holds for the matrix \( Q_{sT} \) defined in (4.4) that

\[
\frac{1}{T} Q_{sT} \xrightarrow{P} s\Sigma
\]

as \( T \to \infty \), uniformly on \([s_0, 1]\), where \( \Sigma \) is given in (4.18).

**Proof.** Consider the entries of \( \frac{1}{T} Q_{sT} \). It holds

\[
\frac{1}{n \nu s} \int_0^{n \nu s} \varphi_i(t) \varphi_j(t) dt = \frac{1}{n \nu s} \sum_{i=1}^{M} \int_{(i-1)\nu}^{i \nu} \varphi_i(t) \varphi_j(t) dt + \frac{1}{n \nu s} \int_{M \nu}^{n \nu s} \varphi_i(t) \varphi_j(t) dt
\]

where \( M = \lfloor ns \rfloor \). Further, it is

\[
\sup_{s \in [s_0, 1]} \left| \frac{1}{n \nu s} \sum_{i=1}^{M} \int_{(i-1)\nu}^{i \nu} \varphi_i(t) \varphi_j(t) dt - 1 \right| = \sup_{s \in [s_0, 1]} \left| \frac{ns}{n} - 1 \right| \to 0
\]

and

\[
\sup_{s \in [s_0, 1]} \left| \frac{1}{n \nu s} \int_{M \nu}^{n \nu s} \varphi_i(t) \varphi_j(t) dt \right| \le \sup_{s \in [s_0, 1]} \frac{1}{n \nu s} \int_{M \nu}^{(M+1)\nu} |\varphi_i(t) \varphi_j(t)| dt
\]

\[
= \frac{1}{n \nu s_0} \int_{0}^{\nu} |\varphi_i(t) \varphi_j(t)| dt \to 0.
\]

Now, regard the entry \( \frac{1}{n} \int_0^{n \nu s} \varphi(t) X_t dt \), \( 0 \le s \le 1 \) and define \( U_T(s) := \frac{1}{T} \int_0^{T s} \varphi(t) \tilde{X}_t dt \). It can be concluded from (4.33) that

\[ U_T(s) - \frac{1}{T} \int_0^{T s} \varphi(t) X_t dt \to 0 \quad \text{a.s.} \]

as \( T \to \infty \) for each \( 0 \le s \le 1 \). Hence it suffices to analyze \( U_T \) which can be written as

\[
\sup_{0 \le s \le 1} U_T(s) = \sup_{s_0 \le s \le 1} \frac{1}{T} \left( \sum_{i=1}^{M} \int_{(i-1)\nu}^{i \nu} \varphi(t) \tilde{X}_t dt + \int_{M \nu}^{T s} \varphi(t) \tilde{X}_t dt \right)
\]

where \( M = \lfloor Ts/\nu \rfloor \). Due to the stationarity of \((\tilde{X}_{k\nu+t})_{k \geq 1}\), see Lemma 3.9, and the periodicity of \( \varphi \) it holds that

\[
\frac{1}{T} E \left| \int_{M \nu}^{T s} \varphi(t) \tilde{X}_t dt \right| \le \frac{1}{T} E \left( \int_{M \nu}^{(M+1)\nu} |\varphi(t) \tilde{X}_t| dt \right)
\]

\[
= \frac{1}{T} E \left( \int_{0}^{\nu} |\varphi(t) \tilde{X}_t| dt \right) \to 0
\]
as $T \to \infty$. Hence
\[
\sup_{s_0 \leq s \leq 1} \frac{1}{T} \int_{M \nu}^{Ts} \varphi(t) \tilde{X}_t \, dt \overset{p}{\to} 0.
\]
Further, one has
\[
\frac{1}{M} \sum_{i=1}^{M} \int_{(i-1)\nu}^{i\nu} \varphi(t) \tilde{X}_t \, dt \to E \left( \int_{0}^{\nu} \varphi(t) \tilde{X}_t \, dt \right)
\]
by the ergodic theorem. This convergence is uniform in $s$ because for all $\varepsilon > 0$ there exists a $M_0 = M_0(\varepsilon, \omega)$ so that it holds for all $M \geq M_0$ that
\[
\left| \frac{1}{M} \sum_{i=1}^{M} \int_{(i-1)\nu}^{i\nu} \varphi(t) \tilde{X}_t \, dt - E \left( \int_{0}^{\nu} \varphi(t) \tilde{X}_t \, dt \right) \right| \leq \varepsilon
\]
since $\inf_{s_0 \leq s \leq 1} [Ts/\nu] = [Ts_0/\nu] \to \infty$. Thus it can be summarized that
\[
\sup_{s_0 \leq s \leq 1} \left| \frac{1}{M} \sum_{i=1}^{M} \int_{(i-1)\nu}^{i\nu} \varphi(t) \tilde{X}_t \, dt - sE \left( \int_{0}^{\nu} \varphi(t) \tilde{X}_t \, dt \right) \right| \to 0
\]
as $T \to \infty$. Finally, look at the term $\frac{1}{T} \int_{0}^{Ts} X_t^2 \, dt$. Due to (3.22) it holds that
\[
\lim_{T \to \infty} \frac{1}{T} \int_{0}^{Ts} X_t^2 \, dt = \limsup_{T \to \infty} \frac{1}{T} \int_{0}^{Ts} e^{-\alpha t} X_0 + h(t) + Z_t \, dt < \infty, \tag{4.34}
\]
almost surely, since $h(t)$ is bounded and $X_0$ and $Z_t$ are bounded almost surely. It follows from (4.34) and (4.33) that
\[
\frac{1}{T} \int_{0}^{Ts} \tilde{X}_t^2 \, dt - \frac{1}{T} \int_{0}^{Ts} X_t^2 \, dt = \frac{1}{T} \int_{0}^{Ts} (\tilde{X}_t + X_t)(\tilde{X}_t - X_t) \, dt \to 0.
\]
Consequently, again by the ergodic theorem and the same method as above, it is
\[
\frac{1}{T} \int_{0}^{Ts} \tilde{X}_t^2 \, dt \to sE \left( \int_{0}^{\nu} \tilde{X}_t^2 \, dt \right)
= \begin{cases} 
& sE \left( \int_{0}^{\nu} (\tilde{h}(t) + \tilde{Z}_t)^2 \, dt \right) \\
& = s \left( \int_{0}^{\nu} \tilde{h}(t) \, dt + \frac{\nu \sigma^2}{2\alpha} \right).
\end{cases}
\]
\[
\square
\]

The proof of Theorem 4.1 can be finally provided.

**Proof of Theorem 4.1.** It follows by Slutzky’s theorem and from Corollary 4.4, 4.11 and Proposition 4.12 that
\[
\Lambda_T(s) = -R_T Q_T^{-1} R_T + R_T Q_T^{-1} R_s T + (R_T - R_s T)^{\alpha} (Q_T - Q_s T)^{-1} (R_T - R_s T)
\]
\[
\overset{p}{\to} -\frac{\|W(1)\|^2}{s} + \frac{\|W(s)\|^2}{1-s} + \frac{\|W(1) - W(s)\|^2}{1-s}
\]
\[
= \frac{\|W(s) - sW(1)\|^2}{s(1-s)}
\]
in $C[s_1, s_2]$. The assertion about the supremum of $\Lambda_T(s)$ is justified by the continuous mapping theorem.
\[
\square
\]
Chapter 5

How to Deal with Jumps?

The present chapter deals with parameter estimation for the drift of jump diffusion processes which are driven by a Lévy process and which are provided with a drift term that is linear in the parameter. That is processes solving the stochastic differential equation

\[ dX_t = f(t, X_t)\theta dt + dL_t, \quad 0 \leq t \leq T, \]

where \((L_t)_{t \geq 0}\) is a homogeneous Lévy process and \(\theta\) the unknown parameter. In the framework of time-continuous data a time-continuous least squares estimator is proposed. This estimator does not coincide with the trajectory fitting estimator, see Section 2.2.3 in Kutoyants [31] for example, which is sometimes referred to as time-continuous least squares estimator in literature. In contrast to the commonly used maximum likelihood estimator the estimator presented in that chapter has the practical advantage that its calculation does not require the evaluation of the continuous part of the sample path which is challenging in many situations.

In the important case of a periodic Ornstein-Uhlenbeck-type jump diffusion which is a discontinuous version of the model presented in Section 3.3 strong consistency is proved.

5.1 Maximum Likelihood vs. Time-Continuous Least Squares

In drift estimation for time-continuously observed diffusion processes maximum likelihood estimation is, as well as in many other fields of statistical inference, the most commonly used estimation method. For a continuous diffusion process with stochastic differential

\[ dX_t = f(t, X_t, \theta)dt + dB_t, \quad 0 \leq t \leq T, \]  

(5.1)

where \((B_t)_{t \geq 0}\) denotes Brownian motion and \(\theta\) the unknown parameter, maximum likelihood estimation is based on Girsanov’s theorem which provides an expression of the likelihood function, see Section 1.3. The resulting maximum likelihood estimator requires the computation of integrals of the form

\[ \int_0^T f(t, X_t, \theta)dX_t, \]  

(5.2)

compare Section 1.3 and Section 3.2. Asymptotic properties of the maximum likelihood estimates from time-continuous realizations of the process in (5.1) can be found e.g. in Bishwal [8], Kutoyants [30] and in Section 3.3 for a concrete diffusion process. Given time-continuous observations \(\{X_t, 0 \leq t \leq T\}\) the integral in (5.2) is approximated by an Itô sum using time-discrete increments of the sample path.
Especially in mathematical finance, an important generalization of model (5.1) is the jump diffusion process allowing for the possibility of discontinuities and a wide variety of marginal distributions, see Barndorff-Nielsen and Shephard [4] and Cont and Tankov [15] for some applications. The jump diffusion process is defined as the solution to

\[ dX_t = f(t, X_t, \theta) \, dt + dL_t, \quad 0 \leq t \leq T, \]

(5.3)

where \((L_t)_{t \geq 0}\) is a homogeneous Lévy process, and the application of the maximum likelihood approach to the Girsanov density yields estimators that are based on

\[ \int_0^T f(t, X_t, \theta) \, dX_t^c \]

where \(X_t^c\) is the continuous part of the process. This integral cannot be computed without further ado since the continuous part is not observed separately in practice. Large sample results on the maximum likelihood estimator for jump diffusion processes are derived in Sørensen [38] and [39].

In this treatment an alternative continuous-time estimator for the drift parameter \(\theta\) of the jump diffusion process given in (5.3) will be presented whereby the drift term is linear in the parameter, that means that the process solves

\[ dX_t = f(t, X_t) \theta \, dt + dL_t. \]

(5.4)

This estimator will be derived by making use of the least squares method. In detail, the discretized version of the stochastic differential equation (5.4) will be regarded first and ordinary least squares estimation will be applied. Thereby, a time-discrete estimator will be obtained and the limit as the discretization step \(\Delta t\) goes toward zero will be taken thereafter resulting in a continuous time estimator of the drift parameter.

The crucial point of this work is the fact that, unlike the maximum likelihood estimator, the proposed estimator requires the computation of integrals of the form (5.2) which can be calculated from the given data and which do not require further investigation detaching the continuous part from the sample path.

In Sections 5.5 and 5.6 strong consistency of the time-continuous least squares estimator for the time-inhomogeneous, mean-reverting Ornstein-Uhlenbeck process of the form (5.4) provided with a periodic drift, see Section 3.3 for its continuous analog, will be proved. Thereby, the similar ideas as used in the proof of Theorem 3.1 in Subsection 3.4 are adjusted and applied to this discontinuous model. Note that mean reversion, periodicity and the occurrence of jumps are meaningful properties of, for example, energy commodity and particularly electricity data, cf. Geman [22].

In the case of the ordinary Ornstein-Uhlenbeck process with jumps whose continuous version is presented in Section 2.1, Hu and Long [23] study the large sample behavior of the time-continuous trajectory fitting estimator, various time-discrete estimation techniques for that model are investigated in Brockwell et al. [12], Diop and Yode [19], Hu and Long [24] and Spiliopoulos [40].

### 5.2 Jump Diffusion Process

Suppose a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)\) is given where \(\mathcal{F}_0\) contains all \(P\)-null sets of \(\mathcal{F}\) and where \(\mathcal{F}_t\) is right continuous, i.e. \(\mathcal{F}_t = \cap_{u>t} \mathcal{F}_u\). Let \((L_t)_{t \geq 0}\) be a Lévy process,
that is an $\mathcal{F}_t$-adapted process which is continuous in probability and which has independent and stationary increments. Throughout this work its unique càdlàg (right-continuous with left limits) modification is considered. The famous Lévy-Itô decomposition (see Applebaum [2], Theorem 2.4.16 on p. 108 or Subsection 1.1) gives the path-wise representation

$$L_t = bt + \sigma B_t + \int_0^t \int_{|x|<1} x\tilde{q}_L(dt, dx) + \int_0^t \int_{|x|\geq 1} xq_L(dt, dx)$$  \hspace{1cm}(5.5)$$

where $(B_t)_{t\geq 0}$ is a (standard) Brownian motion, $b \in \mathbb{R}$, $\sigma > 0$ and where $q_L$ denotes the Poisson random measure associated with $(L_t)_{t \geq 0}$ while $\tilde{q}_L$ is the corresponding compensated measure. In detail, $q_L$ is a random measure on $\mathbb{R}_+ \times (\mathbb{R}\setminus\{0\})$ defined by

$$q_L(t, A) = \#\{0 \leq s \leq t : \Delta L_s \in A\} = \sum_{0 \leq s \leq t} 1_A(\Delta L_s)$$

for all Borel sets $A \in \mathbb{R}\setminus\{0\}$. Thereby, the notation $\Delta L_s = L_s - L_{s-}$, $L_{s-} = \lim_{\varepsilon \to 0} L_{s-\varepsilon}$ and $1_A$ for the indicator function of the set $A$ is used. Further, the compensated Poisson random measure is given by $\tilde{q}_L(dt, dx) = q_L(dt, dx) - dt\mu(dx)$ where $\mu$ is the Lévy measure associated with $(L_t)_{t \geq 0}$ satisfying $\int_{\mathbb{R}\setminus\{0\}}(x^2 + 1)\mu(dx) < \infty$. It holds that $b = E(L_1) - \int_{|x|\geq 1} x\mu(dx)$.

The model of interest is the jump diffusion process $(X_t)_{t \geq 0}$ solving the stochastic differential equation

$$dX_t = f(t, X_t)\theta dt + dL_t, \quad X_0 = \xi,$$  \hspace{1cm}(5.6)$$

where

$$f(t, x) = (f_1(t, x), \ldots, f_p(t, x)), \quad p \in \mathbb{N},$$

and where each $f_i(t, x)$ is a known, real-valued function on $[0, \infty) \times \mathbb{R}$. Further, let the random variable $\xi$ be independent of the Lévy process and $E(\xi^2) < \infty$. The drift parameter $\theta \in \mathbb{R}^p$ is unknown and has to be estimated. The distribution of $\xi$ is required to not depend on $\theta$ otherwise the Radon-Nikodym derivative given in Proposition 5.5 would contain an additional factor, see Kutoyants [31] (p. 37) for details.

Note that equation (5.6) is a short form of the integral equation

$$X_t - \xi = \int_0^t f(s, X_s)\theta ds + L_t.$$  

As in the previous models considered in this thesis, it is implicitly assumed that the well-known Lipschitz and linear growth conditions on the drift function $f$ are satisfied (see Jacod and Shiryaev [25], Theorem III.2.32 on p. 145) such that a unique càdlàg solution to (5.6) exists.

### 5.3 Time-Continuous Least Squares Estimation

This section introduces a least squares estimator for the drift parameter $\theta$. The derivation is based on a discretization of the stochastic differential equation (5.6) to which the ordinary least squares approach is applied. Taking the limit as the refinement improves yields a time-continuous estimator.

The stochastic differential equation (5.6) can be discretized on a time interval $[0, T]$ to the difference equation

$$X_{(i+1)\Delta t} - X_{i\Delta t} = f(i\Delta t, X_{i\Delta t})\theta \Delta t + (L_{(i+1)\Delta t} - L_{i\Delta t}), \quad i = 0, 1, \ldots, N,$$  \hspace{1cm}(5.7)$$
where $N = \lfloor T/\Delta t \rfloor - 1$ and where $\Delta t > 0$ denotes the constant time increment. Here $[x]$ denotes the integer part of $x$. The structure of (5.7) is similar to that of the classical linear model. Even though the conditions of the Gauss-Markov Theorem for linear models are not fulfilled we want to apply least squares estimation which is based on the minimization of the functional

$$g : \theta \mapsto \sum_{i=0}^{N} (X_{(i+1)\Delta t} - X_{i\Delta t} - f(i\Delta t, X_{i\Delta t})\theta \Delta t)^2.$$ 

**Lemma 5.1.** The solution vector $\tilde{\theta}_{T,\Delta t}$ to the minimization problem $g(\theta) \to \min$ is given by

$$\tilde{\theta}_{T,\Delta t} = Q_{T,\Delta t}^{-1}R_{T,\Delta t}$$

where $Q_{T,\Delta t} = \left( \sum_{i=0}^{N} f_j(i\Delta t, X_{i\Delta t})f_k(i\Delta t, X_{i\Delta t})\Delta t \right)_{1 \leq j, k \leq p} \in \mathbb{R}^{p \times p}$ and

$$R_{T,\Delta t} = \left( \sum_{i=0}^{N} f_1(i\Delta t, X_{i\Delta t})(X_{(i+1)\Delta t} - X_{i\Delta t}), \ldots, \sum_{i=0}^{N} f_p(i\Delta t, X_{i\Delta t})(X_{(i+1)\Delta t} - X_{i\Delta t}) \right)^t \in \mathbb{R}^p.$$ 

**Proof.** By general theory of least squares estimation in linear models, the solution to the minimization problem $g(\theta) \to \min$ is given by

$$\theta_{\min} = (A^tA)^{-1}A^tD$$

where

$$A = \Delta t \begin{pmatrix} f_1(0, X_0) & \cdots & f_p(0, X_0) \\ f_1(\Delta t, X_{\Delta t}) & \cdots & f_p(\Delta t, X_{\Delta t}) \\ \vdots & \ddots & \vdots \\ f_1(N\Delta t, X_{N\Delta t}) & \cdots & f_p(N\Delta t, X_{N\Delta t}) \end{pmatrix}, \quad D = \begin{pmatrix} X_{\Delta t} - X_0 \\ X_{2\Delta t} - X_{\Delta t} \\ \vdots \\ X_{(N+1)\Delta t} - X_{N\Delta t} \end{pmatrix}.$$ 

Hence, the products in equation (5.8) can be calculated to be

$$A^tD = \Delta tR_{T,\Delta t}$$

and

$$A^tA = \Delta tQ_{T,\Delta t}.$$ 

Thus $\tilde{\theta}_{T,\Delta t} = (A^tA)^{-1}A^tD = Q_{T,\Delta t}^{-1}R_{T,\Delta t}$. 

Now a continuous-time estimator can be derived from the least squares estimator by considering $\Delta t \to 0$. Note that any càdlàg function can be uniformly approximated on finite intervals by a sequence of step functions since it has countably many discontinuities on finite intervals. Hence, it is Riemann-integrable. This justifies the following convergence of the entries of $Q_{T,\Delta t}$ (as $\Delta t \to 0$):

$$\sum_{i=0}^{N} f_i(i\Delta t, X_{i\Delta t})f_m(i\Delta t, X_{i\Delta t})\Delta t \to \int_0^T f_1(t, X_t)f_m(t, X_t)dt$$

(5.9)
since \( f_j(t, X_t) \) has càdlàg paths (because \( X_t \) has càdlàg paths) and the left-hand side of (5.9) is a Riemann sum. Regarding the entries of \( R_{T, \Delta t} \) it holds that

\[
\sum_{i=0}^{N} f_j(i \Delta t, X_{i \Delta t}) \cdot (X_{(i+1) \Delta t} - X_{i \Delta t}) \to \int_0^T f_j(t, X_t-)dX_t
\]

uniformly on compacts in probability since \( X_t \) is a semi-martingale with càdlàg paths, see Protter [37] (Theorem II.21 on p. 64).

The following proposition has been thus proved.

**Proposition 5.2.** As \( \Delta t \to 0 \), the least squares estimator \( \hat{\theta}_{T, \Delta t} \) converges in probability to \( \hat{\theta}_T = Q_T^{-1} R_T \) where \( Q_T = \left( \int_0^T f_j(t, X_t)f_k(t, X_t)dt \right)_{1 \leq j, k \leq p} \in \mathbb{R}^{p \times p} \) and

\[
R_T = \left( \int_0^T f_1(t, X_t-)dX_t, \ldots, \int_0^T f_p(t, X_t-)dX_t \right)^t \in \mathbb{R}^p.
\]

The estimator \( \hat{\theta}_T \) will be called henceforth continuous time least squares estimator.

**Remark 5.3.** It was implicitly assumed that \( Q_T \) is invertible. This condition holds for many reasonable models, like for jump diffusions of Ornstein-Uhlenbeck form. However, in the case of a singular matrix \( Q_T \), one has to find solutions \( \gamma \in \mathbb{R}^p \) to the normal equations

\[
Q_T \gamma = R_T
\]

and make further constraints in order to determine a proper estimator. Note that the same ambiguity of the solution vector is possible in the case of maximum likelihood estimation which is presented in the next section.

### 5.4 Maximum Likelihood Estimation

The practical advantage of the continuous time least squares estimator introduced in the previous section over the maximum likelihood estimator is demonstrated now. In order to do so, the Lévy process which drives the jump diffusion given in (5.6) is required to fulfill the quite strong conditions (5.10) and (5.11), see below. These constraints are needed for technical reasons and the same shortcoming of the maximum likelihood estimator holds if (5.10) and (5.11) are not satisfied, see Remark 5.7.

Let \( D[0, T] \) denote the space of càdlàg functions from \( [0, T] \) to \( \mathbb{R} \). Denote by \( P_X \) and \( P_L \) the measures induced by the process \( (X_t)_{0 \leq t \leq T} \) solving equation (5.6) and by the Lévy process \( (L_t)_{0 \leq t \leq T} \), respectively. That is

\[
P_X(B) = P(\omega : X^T(\omega) \in B), \quad P_L(B) = P(\omega : L^T(\omega) \in B)
\]

for all Borel sets \( B \in D[0, T], \ X^T(\omega) = \{X_t(\omega), 0 \leq t \leq T\} \).

The following lemma is a direct application of Theorem 2.1 in Sørensen [39] (p. 74).

**Lemma 5.4.** If \( \int_0^T (f(t, X_t)\theta)^2 dt < \infty \) \( P_X \)- and \( P_L \)-almost surely for all \( \theta \), then \( P_X \) and \( P_L \) are equivalent.
Let the Lévy process satisfy
\[ E(\exp(pL_1)) < \infty \] (5.10)
for all \( p \in \mathbb{R} \). This condition guarantees the existence of moments of all orders of the Lévy process and the finiteness of \( \int_{|x| \geq 1} x \mu(dx) \) allowing for the representation
\[ L_t = \sigma B_t + \int_0^t \int_{\mathbb{R}} x \tilde{q}_L(dt, dx) + tE(L_1). \]
Assume further that
\[ \int_{|x| < 1} |x| \mu(dx) < \infty \] (5.11)
which implies finiteness of the total variation of almost all sample paths \( t \mapsto (L_t - \sigma B_t) \) such that the sum of absolute ‘small’ jumps \( \sum_{s \leq t} |\Delta L_s| \mathbf{1}_{|\Delta L_s| < 1}(\Delta L_s) \) is convergent for almost every path.

Referring to Theorem 3.2 in Chan [13] the following statement is valid:

**Proposition 5.5.** Let \( P_X \) be absolutely continuous with respect to \( P_L \) and let
\[ E_L \left( \exp \left( \int_0^T f(t, \xi_t -) \theta \, d\xi_t^c - \frac{1}{2} \int_0^T (f(t, \xi_t)\theta)^2 \, dt \right) \right) < \infty \]
where \( E_L \) denotes expectation with respect to \( P_L \) and where \((\xi_t)_{0 \leq t \leq T}\) is an element in the sample space generated by the process solving (5.6) with a Lévy process satisfying (5.10) and (5.11). Further, the function \( \xi_t^c = \xi_t - \sum_{s \leq t} \Delta \xi_s \) is the continuous part of the sample path \( \xi_t \). Then the Radon-Nikodym derivative \( \frac{dP_X}{dP_L}(\xi) \) is given by
\[ \frac{dP_X}{dP_L}(\xi) = \exp \left( \int_0^T f(t, \xi_t -) \theta \, d\xi_t^c - \frac{1}{2} \int_0^T (f(t, \xi_t)\theta)^2 \, dt \right). \]

**Remark 5.6.** In the model specified in (5.6), the Radon-Nikodym derivative has a simpler form than the one in Chan [13] because \( X^T(\omega) \) and \( L^T(\omega) \) exhibit the same jumps since
\[ \Delta X_t = \lim_{\varepsilon \to 0} (X_t - X_{t-\varepsilon}) = \lim_{\varepsilon \to 0} \int_{t-\varepsilon}^t dX_s = \lim_{\varepsilon \to 0} \int_{t-\varepsilon}^t f(t, X_s) \theta \, dt + \lim_{\varepsilon \to 0} (L_t - L_{t-\varepsilon}) = \Delta L_t. \]
Hence, it holds for the point process \( q_X \) associated with \((X_t)_{t \geq 0}\) that
\[ q_X(t, A) = q_L(t, A) \]
for all Borel sets \( A \in \mathbb{R} \setminus \{0\} \) and all \( t \). Consequently, the change of measure from \( P_L \) to \( P_X \) does not change the ‘weights’ of the discontinuities. So, in our framework, the density of \( P_X \) with respect to \( P_L \) does not include any term that accounts for the jumps.

**Remark 5.7.** The assumption (5.11) simplifies the representation of the continuous part of the sample path and is not required in Chan [13]. Note that a more general statement without the need of (5.10) and (5.11) can be found in Sørensen [39] (Theorem 2.1) whereby the corresponding integrator is again the continuous part of the sample path.
5.4. MAXIMUM LIKELIHOOD ESTIMATION

Suppose the sample path \( X^T = \{ X_t, 0 \leq t \leq T \} \) of the process with stochastic differential given in (5.6) is observed. The maximum likelihood estimator \( \hat{\theta}_T \) is defined by

\[
\hat{\theta}_T := \text{arg max}_\theta \frac{dP_X}{dP_L}(X^T)
\]

where the Radon-Nikodym density \( dP_X/dP_L \) obviously depends on the parameter \( \theta \), see the previous proposition. The partial derivatives of the logarithm of this functional are of the form

\[
\frac{\partial}{\partial \theta_i} \ln \left( \frac{dP_X}{dP_L}(X^T) \right) = \int_0^T \frac{\partial}{\partial \theta_i} f(t, X_t) \theta dX_t^i - \int_0^T f(t, X_t) \theta \frac{\partial}{\partial \theta_i} f(t, X_t) \theta dt,
\]

(5.12)

\[^i = 1, \ldots, p, \]

and the single derivatives of the linear drift function are given by \( \frac{\partial}{\partial \theta_i} f(t, X_t) \theta = f_i(t, X_t) \). Setting the derivatives in (5.12) equal zero results in a system of equations

\[
\int_0^T f_i(t, X_t) \theta dX_t^i - \int_0^T f(t, X_t) \hat{\theta}_T f_i(t, X_t) dt = 0, \quad i = 1, \ldots, p,
\]

which can be written as

\[
\tilde{R}_T - Q_T \hat{\theta}_T = 0
\]

where \( Q_T = \left( \int_0^T f_j(t, X_t) f_k(t, X_t) dt \right)_{1 \leq j, k \leq p} \in \mathbb{R}^{p \times p} \) and

\[
\tilde{R}_T = \left( \int_0^T f_1(t, X_t) dX_t^1, \ldots, \int_0^T f_p(t, X_t) dX_t^p \right)^t \in \mathbb{R}^p.
\]

Proposition 5.8. The maximum likelihood estimator \( \hat{\theta}_T \) is given by

\[
\hat{\theta}_T = Q_T^{-1} \tilde{R}_T
\]

where \( Q_T \) and \( \tilde{R}_T \) are defined above.

Note that the expression for the maximum likelihood estimator \( \hat{\theta}_T \) is similar to the least squares estimator \( \hat{\theta}_T \) given in Proposition 5.2. The discrepancy lies in the vectors \( \tilde{R}_T \) and \( R_T \). The entries of the latter are of the form \( \int_0^T f_i(t, X_t) dX_t^i \) and can be calculated in practice without any difficulty. If time-discrete observations are available these integrals can be approximated by sums. In contrast to that, the integrals \( \int_0^T f_i(t, X_t) dX_t^i \) in \( \tilde{R}_T \) cannot be computed without further investigation due to the integrator which is the continuous part of the sample path. In practice, a discontinuous path is observed such that the continuous part of this path has to be determined by means of further techniques detaching discontinuities. This is a challenging issue unless the paths of the Lévy process have a finite number of jumps along the time interval \([0, T]\). In the case of time-discrete observations the always arising problem is to distinguish the jumps from the continuous points since the entire time-discrete sample looks discontinuous.

Remark 5.9. In the case of an ordinary diffusion process without jumps, that is the process solving

\[
dX_t = f(t, X_t) \theta dt + dB_t, \quad X_0 = \xi,
\]
where \((B_t)_{t \geq 0}\) is a Brownian motion, the Radon-Nikodym derivative in Proposition 5.5 takes the form

\[
\frac{dP_X}{dP_B}(X^T) = \exp \left( \int_0^T f(t, X_t) \theta \, dX_t - \frac{1}{2} \int_0^T (f(t, X_t) \theta)^2 \, dt \right)
\]

for data \(X^T\). This expression is in accordance with the famous Girsanov Theorem, see Lipster and Shiryayev [33] (Theorem 7.6, p. 246), Section 3.2 or Subsection 1.3. The first integral is computed with respect to the entire path since there do not occur any discontinuities in this model. Note that the derivation of the maximum likelihood estimator goes in line with the considerations given above, i.e. differentiating and solving the resulting system of equations, such that \(\hat{\theta}_T = Q_T^{-1} R_T = \hat{\theta}_T\), see Proposition 3.1. That means, in this continuous diffusion model, the maximum likelihood method provides the same estimator as our least squares methodology presented in the previous section.

5.5 Consistency of Least Squares Estimator

In order to substantiate the convenience of the least squares estimator introduced in Section 5.3 (strong) consistency of this estimator for a concrete jump diffusion model is shown. In the quite general setup with regard to the drift function \(f\) in (5.6) consistency requires general conditions on the convergence of the matrix \(Q_T\) in Proposition 5.2 which are not helpful for the application in concrete models.

Consider the time-inhomogeneous, mean-reverting Ornstein-Uhlenbeck process with jumps defined as the solution to

\[
dX_t = \Phi(t, X_t) \theta \, dt + dL_t, \quad X_0 = \xi,
\]

where

\[
\Phi(t, x) = (\varphi_1(t), \ldots, \varphi_p(t), -x), \quad p \in \mathbb{N},
\]

with known, real-valued functions \(\varphi_1, \ldots, \varphi_p\) and \(E(\xi^2) < \infty\), \(\xi\) being independent of the Lévy process. The parameter vector is denoted by \(\theta = (\theta_1, \ldots, \theta_p, \alpha)^t \in \mathbb{R}^p \times \mathbb{R}_+\). As in Section 3.3, consider the case where the drift function \(\Phi\) is periodic in \(t\), i.e.

\[
\Phi(t + \nu, x) = \Phi(t, x) \quad \text{for all } x
\]

where \(\nu\) is a known period. This assumed periodicity leads to the requirement \(\varphi_j(t + \nu) = \varphi_j(t)\). Due to Gram-Schmidt orthogonalization, it can be assumed without loss of generality that \(\varphi_1(t), \ldots, \varphi_p(t)\) form an orthonormal system in \(L^2([0, \nu], \frac{1}{\nu} \, d\lambda)\), that means that

\[
\int_0^\nu \varphi_j(t) \varphi_k(t) \, dt = \begin{cases} 
\nu, & j = k \\
0, & j \neq k.
\end{cases}
\]

Henceforth, it will be assumed that an integral multiple of the period length is observed, i.e. that

\[
T = N \nu \quad \text{for some integer } N.
\]

for notational simplicity, let \(\nu = 1\).

The driving process \((L_t)_{t \geq 0}\) is the right-continuous modification of a Lévy process of the form as described in Section 5.2, that is, as usual, a stochastic process that is continuous in probability with stationary and independent increments. For this model, it is additionally required

\[
E(L_t^2) < \infty
\]
for all $t$ which is equivalent to the requirement $\int_{|x| \geq 1} x^2 \mu(dx) < \infty$ where $\mu$ denotes the Lévy measure. Further, we consider a centered process, that is

$$E(L_t) = 0$$  \hspace{1cm} (5.17)$$

implying that $(L_t)_{t \geq 0}$ is a martingale.

Note that under the assumptions (5.14) and (5.15) the matrix $Q_T$ in Proposition 5.2 simplifies to

$$Q_T = \begin{pmatrix} T I_p & -a_T \\ -a_T^T & b_T \end{pmatrix}$$  \hspace{1cm} (5.18)$$

where $a_T = (\int_0^T \varphi_1(t)X_t \, dt, \ldots, \int_0^T \varphi_p(t)X_t \, dt)^t$, $b_T = \int_0^T X_t^2 \, dt$ and where $I_p$ denotes the $(p \times p)$-identity matrix.

The main result of this section can be formulated now. Its proof is postponed to Section 5.6.

**Theorem 5.1.** Let $\{X_t, 0 \leq t \leq T\}$ be observations of the periodic Ornstein-Uhlenbeck process introduced above satisfying (5.14), (5.15), (5.16) and (5.17). Then the least squares estimator given in Proposition 5.2 is consistent, i.e.

$$\hat{\theta}_T \to \theta, \text{ almost surely,}$$

as $T \to \infty$.

**Remark 5.10.** In the particular case where the driving process is Brownian motion, asymptotic normality of the least squares estimator is given in Theorem 3.2. In that continuous model, least squares and maximum likelihood give the same estimator, see Remark 5.9.

### 5.6 Proof

The following demonstration of the consistency applies the same line of arguments as used in the proof of Theorem 5.1, see Subsection 3.4. These arguments have to be adapted to the discontinuous case that is considered here.

The next Proposition is proved basically in the same manner as Proposition 3.7.

**Proposition 5.11.** The least squares estimator $\hat{\theta}_T$ can be represented as

$$\hat{\theta}_T = \theta + Q_T^{-1} S_T$$  \hspace{1cm} (5.19)$$

where

$$S_T = \begin{pmatrix} \int_0^T \varphi_1(t) \, dL_t \\ \vdots \\ \int_0^T \varphi_p(t) \, dL_t \\ -\int_0^T X_t \, dL_t \end{pmatrix}$$  \hspace{1cm} (5.20)$$

and where $Q_T$ is given in (5.18).

**Proof.** By definition it is $\theta_T = Q_T^{-1} R_T$, where $Q_T$ is defined in (5.18) and

$$R_T = \begin{pmatrix} \int_0^T \varphi_1(t) \, dX_t \\ \vdots \\ \int_0^T \varphi_p(t) \, dX_t \\ -\int_0^T X_t \, dX_t \end{pmatrix}$$
in the model considered here, see Proposition 5.2. Due to the stochastic differential equation (5.13)

\[ dX_t = \left( \sum_{j=1}^{p} \theta_j \varphi_j(t) - \alpha X_t \right) dt + dL_t \] (5.21)

the stochastic integrals in \( R_T \) are

\[ \int_0^{T} \varphi_i(t)dX_t = \sum_{j=1}^{p} \theta_j \int_0^{T} \varphi_i(t) \varphi_j(t) dt - \alpha \int_0^{T} \varphi_i(t) X_t dt + \int_0^{T} \varphi_i(t) dL_t, \quad i = 1, \ldots, p, \]

\[ \int_0^{T} X_{t-}dX_t = \sum_{j=1}^{p} \theta_j \int_0^{T} \varphi_j(t) X_t dt - \alpha \int_0^{T} X_t^2 dt + \int_0^{T} X_{t-} dL_t. \]

Observe that the set \( \{ t \in [0, T] : X_t \neq X_{t-} \} \) is countable and has thus zero-mass with respect to \( dt \). Hence, it follows from these representations combined with (5.14) and (5.15) that

\[ R_T = \begin{pmatrix} T I_p & -aT \\ -d_T & b_T \end{pmatrix} \theta + S_T \]

such that \( Q_T^{-1} R_T = \theta + Q_T^{-1} S_T \).

Due to representation (5.19) the aim in the sequel is to show that

\[ Q_T^{-1} S_T = (T Q_T^{-1}) \left( \frac{1}{T} S_T \right) \]

converges to zero almost surely as \( T \to \infty \). Therefore, it will be proved that \( T Q_T^{-1} \) converges to a finite limit and that \( \frac{1}{T} S_T \) tends to zero, almost surely respectively.

The next lemma differs from Lemma 3.8 only in the fact that the Itô formula for jump diffusions has to be applied now in order to obtain an explicit solution of the process.

**Lemma 5.12.** The solution to the stochastic differential equation (5.13) is given by

\[ X_t = e^{-\alpha t} X_0 + h(t) + Z_t \] (5.22)

where

\[ h(t) = e^{-\alpha t} \sum_{i=1}^{p} \theta_i \int_0^{t} e^{\alpha s} \varphi_i(s) ds \]

and

\[ Z_t = e^{-\alpha t} \int_0^{t} e^{\alpha s} dL_s. \]

**Proof.** The Itô lemma, see Protter [37] (Theorem II.32, p. 71) or Kuo [29] (Theorem 7.6.1 and 7.6.4, p. 111 and 113), states that it holds for a semimartingale \( (Y_t)_{t \geq 0} \) and a \( C^{1,2} \) function \( F(t,x) \) that

\[
F(t,Y_t) = F(0,Y_0) + \int_0^t \frac{\partial F}{\partial t}(s,Y_s) ds + \int_0^t \frac{\partial F}{\partial x}(s,Y_{s-}) dY_s + \frac{1}{2} \int_0^t \frac{\partial^2 F}{\partial x^2}(s,Y_{s-}) d[Y]_s^c \\
+ \sum_{0 < s \leq t} \left( F(s,Y_s) - F(s,Y_{s-}) - \frac{\partial F}{\partial x}(s,Y_{s-}) \Delta Y_s \right)
\]
where \([Y]_s^c\) is the continuous part of the quadratic variation process \([Y]_s\) and \(\Delta Y_s = Y_s - Y_{s-}\).

Now for \(F(t, x) = e^{\alpha t}x\) and the stochastic differential (5.13) which can be written as in (5.21) it follows that

\[
F(t, X_t) = X_0 + \int_0^t \alpha e^{\alpha s}X_s ds + \int_0^t e^{\alpha s}dX_s + \sum_{0 < s \leq t} \left( e^{\alpha s}X_s - e^{\alpha s}X_{s-} - e^{\alpha s}\Delta X_s \right)
\]

\[
= X_0 + \sum_{j=1}^p \theta_j \int_0^t e^{\alpha s}\varphi_j(s) ds + \int_0^t e^{\alpha s}dL_s
\]

by making use of the equality \(X_s - X_{s-} = \Delta X_s\). Finally, multiply by \(e^{-\alpha t}\).

Due to the time-dependence of \(h\) and \(Z\) in (5.22) the process \((X_t)_{t \geq 0}\) is not stationary in the ordinary sense such that the ergodic theorem is not directly applicable. So the idea is again to introduce a solution to the stochastic differential equation (5.13) with time index \(t \in \mathbb{R}\) instead of \(t \geq 0\) which is stationary and ergodic when it is interpreted as a sequence of path-valued random variables.

As in Subsection 3.4, define the process

\[
\tilde{X}_t = \tilde{h}(t) + \tilde{Z}_t
\]

(5.23)

where \(\tilde{h} : [0, \infty) \to \mathbb{R}\) is given by

\[
\tilde{h}(t) = e^{-\alpha t} \sum_{j=1}^p \theta_j \int_{-\infty}^t e^{\alpha s}\varphi_j(s) ds
\]

(5.24)

and

\[
\tilde{Z}_t = e^{-\alpha t} \int_{-\infty}^t e^{\alpha s}d\tilde{L}_s
\]

(5.25)

whereby

\[
\tilde{L}_s := L_s1_{\{s \geq 0\}}(s) + \bar{L}_s1_{\{s < 0\}}(s)
\]

by taking \(\bar{L}_s\), when \(s < 0\), to be an independent copy of \(-L_{-(s-)}\) (see p. 214 in Applebaum [2]). Constructed in this way, the process \((\tilde{L}_t)_{t \in \mathbb{R}}\) is a continuation of \((L_t)_{t \geq 0}\) to \(\mathbb{R}\) such that \((\tilde{L}_t)_{t \in \mathbb{R}}\) is also a Lévy process with càdlàg paths.

Let \(D[0, 1]\) be the space of càdlàg functions on \([0, 1]\).

**Lemma 5.13.** The sequence \((W_k)_{k \in \mathbb{N}}\) of \(D[0, 1]\)-valued random variables defined by

\[
W_k(s) := X_{k-1+s}, \quad 0 \leq s \leq 1,
\]

is stationary and ergodic.

**Proof.** The proof is carried out in exactly the same way as the proof of Proposition 3.9. The crucial property of the driving process is the independence and stationarity of its increments which hold true in the case of the Lévy process. Hence, all calculations may be adopted whereby the extended Brownian motion \((\tilde{B}_t)_{t \in \mathbb{R}}\) has to be replaced by the extended Lévy martingale \((\tilde{L}_t)_{t \in \mathbb{R}}\).
The conditions (5.16) and (5.17) are incorporated into the next lemma. Denote by \((L_t)_{t \geq 0}\) the uniquely determined previsible process with non-decreasing sample paths such that
\[
L_t^2 - \langle L \rangle_t \geq 0
\] (5.26) is a martingale. Usually \((L_t)_{t \geq 0}\) is referred to as compensator or bracket process associated with \((L)_{t \geq 0}\). Its existence is guaranteed by the Doob-Meyer decomposition of \(L_t^2\), see Theorem 1.3.

**Lemma 5.14.** Let \((L_t)_{t \geq 0}\) be a Lévy process satisfying (5.16) and (5.17). Then the bracket process \((\langle L \rangle_t)_{t \geq 0}\) in (5.26) is of the form
\[
\langle L \rangle_t = ct, \quad t \geq 0,
\]
where \(c\) is a finite, non-negative constant.

**Proof.** By the Doob-Meyer decomposition the process \((L_t^2 - \langle L \rangle_t)_{t \geq 0}\) is a martingale. It can be proved that \((L_t^2 - \langle L \rangle_t)_{t \geq 0}\) is also a martingale, where \(\langle |L| \rangle_{t \geq 0}\) denotes the quadratic variation process. Hence, \((L_t^2 - \langle L \rangle_t)_{t \geq 0}\) is a martingale. Further, the process \((\langle L \rangle_t)_{t \geq 0}\) is a Lévy process implying that \((L_t^2 - E(\langle L \rangle_t))_{t \geq 0}\) is a martingale. According to Cont and Tankov [15] (Example III.8.5, p. 266) the quadratic variation process has the form
\[
[L]_t = \sigma^2 t + \int_0^t \int_{\mathbb{R} \setminus \{0\}} x^2 q_L(dt, dx)
\]
where \(q_L\) denotes the jump measure associated with \((L_t)_{t \geq 0}\). Since by (5.16) and the fundamental properties of the Lévy measure, see Section 5.2, it holds that \(\int_{\mathbb{R} \setminus \{0\}} x^2 \mu(dx) < \infty\) and \(E(q_L(dt, dx)) = dt \mu(dx)\), it follows that \(E([L]_t) = ct\).

\(\square\)

**Lemma 5.15.** As \(t \to \infty\), one has
\[
|\tilde{X}_t - X_t| \to 0, \quad \text{almost surely.}
\]

**Proof.** It can be seen that
\[
|\tilde{X}_t - X_t| \leq e^{-\alpha t}|X_0| + |\tilde{h}(t) - h(t)| + |\tilde{Z}_t - Z_t|
\]
\[
\leq e^{-\alpha t}|X_0| + e^{-\alpha t} \sum_{i=1}^p \theta_i \int_{-\infty}^0 e^{\alpha s} |\varphi_i(s)| ds + |e^{-\alpha t} \int_{-\infty}^0 e^{\alpha s} d\tilde{L}_s|.
\]
Obviously, the first two terms on the right-hand side converge toward zero as \(t \to \infty\). Further, since \(E(L_t^2) < \infty\) by assumption, it holds by the Itô isometry for stochastic integrals with martingales as integrators that
\[
E\left((\int_{-\infty}^0 e^{\alpha s} d\tilde{L}_s)^2\right) = E\left(\int_{-\infty}^0 e^{2\alpha s} d\langle \tilde{L} \rangle_s\right)
\]
where \(\tilde{L}_s\) denotes the compensator of the Lévy martingale \(\tilde{L}_s\) obtained from the Doob-Meyer decomposition such that \((L_t^2 - \langle L \rangle_s)_{s \leq t}\) is a martingale. By Lemma 5.14 it is \(\langle \tilde{L} \rangle_s = cs\), where \(c\) is a finite, non-negative constant, so that
\[
E\left(\int_{-\infty}^0 e^{2\alpha s} d\langle \tilde{L} \rangle_s\right) = \int_{-\infty}^0 e^{2\alpha s} c ds = \frac{c}{2\alpha} < \infty.
\]
Hence it has been shown that \( E \left( \int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s \right)^2 < \infty \) which implies that \( |\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s| < \infty \) almost surely. It follows that
\[
e^{-\alpha t} |\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s| \to 0
\]
as \( t \to \infty \). \(\square\)

Now turn to the matrix \( Q_T \). Due to its simplified form in this model, see representation (5.18), its inverse can be explicitly computed.

**Lemma 5.16.** The inverse of the matrix \( Q_T \) given in (5.18) can be computed to be
\[
Q_T^{-1} = \frac{1}{T} \begin{pmatrix} I_p + \gamma T \Lambda T \Lambda^T & -\gamma T \Lambda_T \\ -\gamma T \Lambda^T & \gamma T \end{pmatrix}
\]
where \( \Lambda_T = (\Lambda_{T,1}, \ldots, \Lambda_{T,p})^T = \frac{1}{T} \alpha_T \), see (5.18), and
\[
\gamma_T = \left( \frac{1}{T} \int_0^T X_t^2 dt - \sum_{i=1}^p \Lambda_{T,i}^2 \right)^{-1}.
\]

**Proof.** See Lemma 3.4. \(\square\)

**Proposition 5.17.** It holds that
\[
T Q_T^{-1} \to C,
\]
almost surely, as \( T \to \infty \). The \((p+1) \times (p+1)\) matrix \( C \) is given by
\[
C = \begin{pmatrix} I_p + \gamma \Lambda \Lambda^T & -\gamma \Lambda \\ -\gamma \Lambda^T & \gamma \end{pmatrix}
\]
whereby \( \Lambda = (\Lambda_1, \ldots, \Lambda_p)^T \) and
\[
\Lambda_i = \int_0^1 \varphi_i(t) \tilde{h}(t) dt, \quad i = 1, \ldots, p,
\]
\[
\gamma = \left( \int_0^1 (\tilde{h}(t))^2 dt + E(\tilde{Z}_0^2) - \sum_{i=1}^p \Lambda_i^2 \right)^{-1}.
\]
The function \( \tilde{h} \) and the random variable \( \tilde{Z}_t \) are specified in (5.24) and (5.25).

**Proof.** The assertion can be proved by means of the same ideas and calculations as used in the proof of Proposition 3.11 by making use of the ergodicity and stationarity of the sequence \((\tilde{X}_{k-1+s})_{k \in \mathbb{N}}\) justified by Lemma 5.13, the asymptotic equality stated in Lemma 5.15 and the convergence (5.27) implicating
\[
|\frac{1}{T} \int_0^T (Z_t - \tilde{Z}_t) dt| \leq \frac{1}{T} \int_0^T |Z_t - \tilde{Z}_t| dt = \frac{1}{T} \int_0^T e^{-\alpha t} |\int_{-\infty}^{0} e^{\alpha s} d\tilde{L}_s| dt \to 0
\]
such that the ergodic theorem gives
\[
\frac{1}{T} \int_0^T \tilde{Z}_t dt \to E(\tilde{Z}_0) = 0.
\]
\(\square\)
Lemma 5.18. The term $\frac{1}{\sqrt{T}} S_T$ is bounded in $L^2$.

Proof. Note that $\frac{1}{\sqrt{T}} \int_0^T \varphi_i(t) \, dL_t$ is $L^2$-bounded since

$$E \left( \frac{1}{\sqrt{T}} \int_0^T \varphi_i(t) \, dL_t \right)^2 = E \left( \frac{1}{T} \int_0^T \varphi_i(t)^2 \, d\langle L \rangle_t \right) = \frac{c}{T} \int_0^T \varphi_i(t)^2 \, dt < \infty$$

by making use of the Itô isometry and Lemma 5.14. For the last entry of $\frac{1}{\sqrt{T}} S_T$ one has to prove the boundedness of

$$E \left( \frac{1}{\sqrt{T}} \int_0^T X_t \, dL_t \right)^2 = \frac{1}{T} E \left( \int_0^T X_t^2 \, d\langle L \rangle_t \right) = \frac{1}{T} E \left( \int_0^T \left( 2e^{-\alpha t} X_0 h(t) + 2e^{-\alpha t} X_0 Z_t + e^{-2\alpha t} X_0^2 + 2h(t) Z_t + h(t)^2 + Z_t^2 \right) \, d\langle L \rangle_t \right).$$

Again, $\langle L \rangle_t = ct$. Since $Z_t$ is a zero-mean random variable the expectation of the second and fourth term is zero. Moreover, $E(Z_t^2) = \frac{c^2}{2\alpha}(1 - e^{-2\alpha t}) < \infty$ such that

$$\sup_{T \geq 0} \frac{1}{T} E \left( \int_0^T Z_t^2 \, dt \right) < \infty.$$

Further, the function $h$ is bounded and $E(X_0^2) < \infty$ resulting in

$$\sup_{T \geq 0} \frac{1}{T} E \left( \int_0^T e^{-\alpha t} X_0 h(t) \, dt \right) < \infty$$

and $\sup_{T \geq 0} \frac{1}{T} \int_0^T h(t)^2 \, dt < \infty$. \hfill $\Box$

Proposition 5.19. It holds

$$\lim_{T \to \infty} \frac{1}{T} S_T = 0,$$

almost surely, as $T \to \infty$.

Proof. Observe that $S_T$ is a martingale since the Lévy process is a martingale due to condition (5.17). By Lemma 5.18, $\frac{1}{\sqrt{T}} S_T$ is $L^2$-bounded. Doob’s maximal inequality for time-discontinuous submartingales, see Theorem 2.1.5 in Applebaum [2] (p. 74), provides for any $\epsilon > 0$ that

$$P \left( \sup_{2^k \leq T \leq 2^{k+1}} \frac{1}{T} |S_T| \geq \epsilon \right) \leq P \left( \sup_{2^k \leq T \leq 2^{k+1}} |S_T| \geq \epsilon 2^k \right) \leq \frac{4}{\epsilon^2 2^{2k}} E \left| S_{2^{k+1}} \right|^2 = O(2^{-k}).$$

The Borel-Cantelli theorem yields $\lim \sup_{T \to \infty} \frac{1}{T} |S_T| \leq \epsilon$, almost surely. \hfill $\Box$

Proof of Theorem 5.1. This follows directly from Proposition 5.17 and Proposition 5.19. \hfill $\Box$
Chapter 6

Simulation

This chapter discusses a small simulation study that was conducted to support the theory investigated in the previous chapters by confirming the consistency and asymptotic normality of the time-continuous maximum likelihood estimator, see Section 3.3, or by illustrating the convergence of the likelihood ratio test statistic presented in Section 4.3. In order to do so, a generalized mean reversion process as proposed in Section 3.1 provided with a certain periodic mean reversion function is simulated. Thereby, pseudo-random numbers are generated using the standard random number generator of the numerical computing software ‘Matlab’ (The Mathworks, Inc., see www.mathworks.com) providing simulated time-discrete data. In detail, a countable and finite number of observations using a small discretization $\Delta t$ and a large number $n$ of the period $\nu$ is generated.

Maximum Likelihood Estimator

For some $\Delta t > 0$ sufficiently small, time-discrete observations $X_{\Delta t}, X_{2\Delta t}, \ldots, X_{N\Delta t}$, at equivalently spaced time points of the generalized mean reversion process

$$dX_t = (L(t) - \alpha X_t)dt + \sigma dB_t, \quad t \geq 0,$$

where

$$L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t),$$

are generated. Thereby, it is required that $N \Delta t = n \nu$ where $n \in \mathbb{N}$ and $\nu > 0$ is the period of the periodic function $L(t)$, that means realizations over a multiple of entire periods are available. The time-discrete version of the maximum likelihood estimator introduced in Section 3.3 is applied to the generated data, i.e. the integrals appearing in the continuous time estimator, see Proposition 3.1, are replaced by adequate sums. This yields the estimator $\hat{\theta}_{ML, \Delta t} = (\hat{\mu}_1, \ldots, \hat{\mu}_p, \hat{\alpha})$ given by

$$\hat{\theta}_{ML, \Delta t} = Q_{T, \Delta t}^{-1} P_{T, \Delta t}$$

where $Q_{T, \Delta t} \in \mathbb{R}^{(p+1) \times (p+1)}$ and $P_{T, \Delta t} \in \mathbb{R}^{p+1}$ are defined as

$$Q_{T, \Delta t} = \Delta t \begin{pmatrix} G_{T, \Delta t} & a_{T, \Delta t} \\ a_{T, \Delta t}^T & b_{T, \Delta t} \end{pmatrix}, \quad P_{T, \Delta t} = \begin{pmatrix} \sum_{i=0}^{N} \varphi_1(i \Delta t)(X_{(i+1)\Delta t} - X_{i\Delta t}) \\ \vdots \\ \sum_{i=0}^{N} \varphi_p(i \Delta t)(X_{(i+1)\Delta t} - X_{i\Delta t}) \\ - \sum_{i=0}^{N} X_{i\Delta t}(X_{(i+1)\Delta t} - X_{i\Delta t}) \end{pmatrix}$$
and where \( G_{T,\Delta t} = \left( \sum_{i=0}^{N} \varphi_j(i \Delta t) \varphi_k(i \Delta t) \right)_{1 \leq j, k \leq p} \in \mathbb{R}^{p \times p} \), \( b_{T,\Delta t} = \sum_{i=0}^{N} X_i^2 \Delta t \) and \( a_{T,\Delta t} = -(\sum_{i=0}^{N} \varphi_i(i \Delta t) X_i \Delta t, \ldots, \sum_{i=0}^{N} \varphi_p(i \Delta t) X_i \Delta t) \). This proceeding including data simulation and parameter estimation is repeated independently \( m \) times, \( m \) sufficiently large, so that an empirical distribution of the maximum likelihood estimates \( \hat{\theta}_{ML}^{(k)}, k = 1, \ldots, m \), is obtained. The mean reversion function \( L \) in (6.2) is chosen to be

\[
L(t) = \mu_0 + \mu_1 \sin(2\pi t/\nu) + \mu_2 \cos(2\pi t/\nu).
\]

The parameter setting and the outcome of six simulations denoted by (a),..., (f) are summarized in Table 6.1.

Observe that the estimates in the cases (b) and (d) seem to be quite poor. This is due to the small number of periods in the data set \( (n = 5) \) in these cases. Another reason for this unsatisfactory estimation performance is a phenomenon called multicolinearity which sometimes occurs in multiple regression. Here, it can be explained as follows: The matrix \( Q_{T,\Delta t} \) specified in (6.3) can be represented as

\[
Q_{T,\Delta t} = \Delta t \begin{pmatrix} a_{T,\Delta t} & a_{T,\Delta t} \\ a_{T,\Delta t}^T & b_{T,\Delta t} \end{pmatrix} = \frac{1}{\Delta t} A^t A
\]

where

\[
A = \Delta t \begin{pmatrix} \varphi_1(0) & \ldots & \varphi_p(0) & -X_0 \\ \varphi_1(\Delta t) & \ldots & \varphi_p(\Delta t) & -X_{\Delta t} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(N\Delta t) & \ldots & \varphi_p(N\Delta t) & -X_{N\Delta t} \end{pmatrix}.
\]

Now, if the columns of the matrix \( A \) are perfectly correlated or if there is a perfect linear relationship among the columns, then the rank of \( A \) is less than \( p + 1 \) and \( A^t A \) is not invertible. Usually, such a perfect multicolinearity is unlikely. But in the situation where the columns are highly correlated or satisfy an almost linear relationship, \( A \) is barely invertible and may be ill-conditioned. As a result, the computation of the inverse of \( A^t A \) may be complicated and may lead to numerical inaccuracy affecting the preciseness of the maximum likelihood estimate which requires the calculation of \( (A^t A)^{-1} \).

In the simulation cases (c) and (e) the diffusion coefficient \( \sigma \) is increased such that the simulated realizations of the process exhibit more fluctuations and a strong linear dependence of \( X_{\Delta t}, X_{2\Delta t}, \ldots, X_{N\Delta t} \) on \( \sin(2\pi j \Delta t/\nu) \) and \( \cos(2\pi j \Delta t/\nu) \), \( j = 1, \ldots, N \), is avoided. It can be seen that the increase of the value of \( \sigma \) improves the estimation results which suffer on the other hand a larger variance of the estimators than in the cases (b) and (d). This phenomenon is highlighted in Table 6.2 where the simulations (g), (h) and (i) differ in the the values of \( \sigma \). Table 6.3 provides empirical correlations of the columns of \( A \) in the case of the mean reversion function given in (6.4), i.e. correlations between the generated data set \( X_{\Delta t}, X_{2\Delta t}, \ldots, X_{N\Delta t} \) and both \( \sin(2\pi \Delta t/\nu), \sin(2\pi 2\Delta t/\nu), \ldots, \sin(2\pi N\Delta t/\nu) \) and the set \( \cos(2\pi \Delta t/\nu), \cos(2\pi 2\Delta t/\nu), \ldots, \cos(2\pi N\Delta t/\nu) \). A high correlation between these samples, known as regressors in a multiple regression model, or between linear combinations of those indicates multicolinearity.
Table 6.1: Parameter setting and empirical quantities of the centered estimates \( \hat{\theta} - \theta \) in six simulations termed (a),...,(f), respectively, of time-discrete observations of the process (6.1) with mean reversion function (6.4). It is \( X_0 = 20, \nu = 200 \) and \( N \Delta t = n \nu \).

<table>
<thead>
<tr>
<th>simulation no.</th>
<th>parameter setting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
</tr>
<tr>
<td>( m ) (# iterations)</td>
<td>( 10^3 )</td>
</tr>
<tr>
<td>( n ) (# periods)</td>
<td>50</td>
</tr>
<tr>
<td>( N ) (# observations)</td>
<td>( 10^4 )</td>
</tr>
<tr>
<td>( \Delta t )</td>
<td>1</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>14</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td>7</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>mean</th>
<th>variance</th>
<th>kurtosis</th>
<th>skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>( \hat{\mu}_0 - \mu_0 )</td>
<td>0.725530</td>
<td>0.035152</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_1 - \mu_1 )</td>
<td>0.154469</td>
<td>0.009153</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_2 - \mu_2 )</td>
<td>0.581399</td>
<td>0.008707</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} - \alpha )</td>
<td>0.036357</td>
<td>0.000088</td>
</tr>
<tr>
<td>(b)</td>
<td>( \hat{\mu}_0 - \mu_0 )</td>
<td>-4.593931</td>
<td>0.100380</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_1 - \mu_1 )</td>
<td>-2.535962</td>
<td>0.027052</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_2 - \mu_2 )</td>
<td>-2.050593</td>
<td>0.024218</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} - \alpha )</td>
<td>-0.229388</td>
<td>0.000252</td>
</tr>
<tr>
<td>(c)</td>
<td>( \hat{\mu}_0 - \mu_0 )</td>
<td>0.077560</td>
<td>0.807408</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_1 - \mu_1 )</td>
<td>-0.140978</td>
<td>0.906945</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_2 - \mu_2 )</td>
<td>0.228517</td>
<td>0.878482</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} - \alpha )</td>
<td>0.003693</td>
<td>0.000088</td>
</tr>
<tr>
<td>(d)</td>
<td>( \hat{\mu}_0 - \mu_0 )</td>
<td>1.681520</td>
<td>0.003741</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_1 - \mu_1 )</td>
<td>0.637114</td>
<td>0.000997</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_2 - \mu_2 )</td>
<td>1.054908</td>
<td>0.00916</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} - \alpha )</td>
<td>0.084071</td>
<td>0.000009</td>
</tr>
<tr>
<td>(e)</td>
<td>( \hat{\mu}_0 - \mu_0 )</td>
<td>-0.007553</td>
<td>0.007393</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_1 - \mu_1 )</td>
<td>-0.205005</td>
<td>0.009643</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_2 - \mu_2 )</td>
<td>0.223805</td>
<td>0.01285</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} - \alpha )</td>
<td>-0.000159</td>
<td>0.000008</td>
</tr>
<tr>
<td>(f)</td>
<td>( \hat{\mu}_0 - \mu_0 )</td>
<td>-0.149110</td>
<td>7.262317</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_1 - \mu_1 )</td>
<td>-0.208887</td>
<td>2.765054</td>
</tr>
<tr>
<td></td>
<td>( \hat{\mu}_2 - \mu_2 )</td>
<td>0.014685</td>
<td>2.553611</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} - \alpha )</td>
<td>-0.007309</td>
<td>0.016212</td>
</tr>
</tbody>
</table>
Table 6.2: Parameter setting and empirical quantities of the centered estimates $\hat{\theta} - \theta$ in three simulations as in Table 6.1 whereby the settings of the cases (g), (h) and (i) differ in the choice of the diffusion coefficient $\sigma$. It is $X_0 = 70$, $\nu = 250$ and $N\Delta t = n\nu$.

<table>
<thead>
<tr>
<th>simulation no.</th>
<th>parameter setting</th>
<th>(g)</th>
<th>(h)</th>
<th>(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$ (# iterations)</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>$n$ (# periods)</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>$N$ (# observations)</td>
<td>$12.5 \cdot 10^4$</td>
<td>$12.5 \cdot 10^4$</td>
<td>$12.5 \cdot 10^4$</td>
<td></td>
</tr>
<tr>
<td>$\Delta t$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>2</td>
<td>20</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: Empirical correlations $\rho$ calculated for one of the $m$ iterations of each of the cases (g), (h) and (i).

| $\rho\{|L(j\Delta t), j = 1, \ldots, N\}, \{X_j\Delta t, j = 1, \ldots, N\}$ | (g) | (h) | (i) |
|----------------|-----|-----|-----|
| $\rho\{|\sin(2\pi j\Delta t/\nu), j = 1, \ldots, N\}, \{X_j\Delta t, j = 1, \ldots, N\}$ | $0.9213$ | $0.2508$ | $0.0756$ |
| $\rho\{|\cos(2\pi j\Delta t/\nu), j = 1, \ldots, N\}, \{X_j\Delta t, j = 1, \ldots, N\}$ | $0.5286$ | $0.1424$ | $0.0426$ |
| $\rho\{|L(j\Delta t), j = 1, \ldots, N\}, \{X_j\Delta t, j = 1, \ldots, N\}$ | $0.7657$ | $0.2092$ | $0.0632$ |
Likelihood Ratio

The convergence of the likelihood ratio test statistic derived in Chapter 4 as a tool for detecting a change point in the periodic mean reversion model is considered in the following. By Theorem 4.1 it holds under the null hypothesis of no change point in the data set that

\[
\sup_{s \in [s_1, s_2]} \Lambda_T(s) \overset{D}{\rightarrow} \sup_{s \in [s_1, s_2]} \frac{\| W(s) - sW(1) \|_2^2}{s(1 - s)}.
\]

(6.5)

In the same manner as described in the previous section, time-discrete observations of the mean reversion process (6.1) with mean reversion function (6.4) are generated via computer simulations. For each of the iterations \( k = 1, \ldots, m \) a finite set \( \Lambda_{kN\Delta t}(1/N), \Lambda_{kN\Delta t}(2/N), \ldots, \Lambda_{kN\Delta t}(1) \) is calculated such that the maximum of the constrained set

\[
\Lambda_{kN\Delta t}(\kappa_1/N), \Lambda_{kN\Delta t}((\kappa_1 + 1)/N), \ldots, \Lambda_{kN\Delta t}(\kappa_2/N)
\]

where \( \kappa_1, \kappa_2 \in \mathbb{N} \) with \( 0 < \kappa_1 < \kappa_2 < N \) is obtained. The resulting empirical distribution of the \( m \) maxima is considered afterwards. At the same time the limit in (6.5) is simulated. This is done by generating a discrete sample of the multi-dimensional Brownian motion \( (W_t)_{t \in [0, 1]} \) on the same time grid \( 1/N, 2/N, \ldots, 1 \) as above. Then the maximum of the fraction in (6.5) over the constrained grid \( \kappa_1/N, (\kappa_1 + 1)/N, \ldots, \kappa_2/N \) is determined. Replicating this procedure \( m \) times provides an approximation of the distribution of the limit object in (6.5).

In the simulations denoted by (A),...,(F) and summarized next, the parameter settings are more or less the same as in simulation (a) of Table 6.1. The aim of these simulations is to capture the dependence of the convergence stated in (6.5) upon the choice of \( \kappa_1 \) and \( \kappa_2 \) and the number of periods in the data set.

In Figure 6.1, the quantile-quantile plots compare the distributions of the objects in (6.5) by plotting the empirical quantiles of both simulated data sets against each other. In case (A) both distributions correspond quite well to each other since the small number of periods is compensated by a narrow interval \( [\kappa_1/N, \kappa_2/N] \) whereas in (C) the wide interval \( [\kappa_1/N, \kappa_2/N] \) makes the points lie far away from the line suggesting a big discrepancy between the distributions.

In contrast to that, in the case of a quite large number of periods (\( n = 50 \)) the pattern of the points in the plots (D), (E) and (F) follows the line \( y = x \), except for some outliers, indicating that both distributions are close to each other. It can be concluded from all considered cases that the closer the bounds \( \kappa_1/N \) and \( \kappa_2/N \) to 0 and 1, respectively, the larger the discrepancy between the quantiles of the distributions.

Table 6.4: Parameter setting in the simulations (A),...,(F). In all these simulations it is \( X_0 = 20, \nu = 200, N\Delta t = n\nu, m = 100, \mu_0 = 14, \Delta t = 1, \sigma = 10, \mu_1 = \mu_2 = 7 \), compare (a) in Table 6.1.
Figure 6.1: Quantile-quantile plots of the empirical distribution of the test statistic and its limit object given in (6.5) in the simulations (A)....(F). The dashed line is the function $y = x$ helping evaluate the similarity of the empirical distributions.
Bibliography


Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Herold Dehling for his continuous support of my research, his immense knowledge and for the pleasant atmosphere in his research group. His personal guidance and wide experience have been of great value in my career. I owe a great deal to Dr. Brice Franke who was always interested in discussions on stochastic analysis.

Particular gratitude goes to Prof. Dr. Reg Kulperger for giving me the opportunity to spend four months at the University of Western Ontario in Canada. During that time, I could benefit from Professor Kulperger's broad range of knowledge in statistics. I am also grateful to Prof. Dr. Yuri Kutoyants, Université du Maine in France, who is an expert in the field of statistical inference for time-continuously observed diffusions and who hosted me in Le Mans, France, in order to respond to my questions.

I am indebted to the energy company E.ON Ruhrgas AG in Essen, Germany, and especially to the risk management department under the leadership of Dr. Volker Brand, for a fruitful cooperation which was a major contribution to the motivation of this thesis as it delivered an insight into stochastic applications and problems in the energy industry. For professional support, I am especially grateful to Dr. Holger Haaf, Carsten Teller, Dr. Wolfgang Terbeck and Christoph Ziegler. I appreciate additionally the financial support by the Collaborative Research Center 'Statistical modeling of non-linear dynamic processes' (SFB 823) of the German Research Foundation during the final preparation of this manuscript.

Last but not least, I wish to give my deepest appreciation to my parents Christina and Arkadius and my brother Markus who encouraged me during my entire education and supported me in my decisions. I am also grateful to my girlfriend Anika for her motivation, patience and understanding while I was working on this thesis.