Statistical estimation for Lévy driven OU-processes and Brownian semimartingales

Meulen, F.H. van der , 1977 -
Statistical estimation for Lévy driven OU-processes and Brownian semimartingales

Thomas Stieltjes Institute
for Mathematics

IBIS UvA BV

INSTITUUT VOOR BEDRIJFS- EN INDUSTRIËLE STATISTIEK

Dit proefschrift is mede mogelijk gemaakt door een financiële bijdrage van het Instituut voor Bedrijfs- en Industriële Statistiek van de Universiteit van Amsterdam (IBIS UvA).

VRIJE UNIVERSITEIT

Statistical estimation for Lévy driven OU-processes and Brownian semimartingales

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op woensdag 30 november 2005 om 15.45 uur
in het auditorium van de universiteit,
De Boelelaan 1105

door

Frank Henri van der Meulen

geboren te Papendrecht

promotor: prof.dr. A.W. van der Vaart

copromotor: dr.ir. G. Jongbloed

Contents

Introduction **1**
Overview of chapters 1–3 2
Overview of chapter 4 4
Publication details 5

1 Lévy processes and induced OU-processes **7**
1.1 Infinite divisibility 7
1.1.1 Positive self-decomposable distributions 10
1.2 Lévy processes 12
1.2.1 Increasing Lévy processes 16
1.3 Simulation of subordinators 16
1.4 Ornstein-Uhlenbeck processes driven by a subordinator 19
1.5 A condition for the OU-process to be β-mixing 23
1.6 Proof of φ-irreducibility of an AR(1)-process 26
1.7 Some additional proofs and results 29

2 Nonparametric estimation for OU-processes driven by a subordinator **33**
2.1 Definition of a cumulant M-estimator 33
2.2 Consistency 39
2.3 Computing the cumulant M-estimator 43
2.3.1 The support reduction algorithm 43
2.3.2 Implementation details 46
2.3.3 Local optimization of the support 48
2.3.4 Testing the support reduction algorithm 49
2.4 Extension to convex decreasing and completely monotone canonical functions 49
2.5 Applications and examples 51
2.5.1 The empirical characteristic function as a preliminary estimator . 51
2.5.2 Data-adaptive choice of the weight-function 53
2.5.3 A simple example 54
2.5.4 Example: data from the OU-process. Observation scheme *(i)* . . . 54
2.5.5 Example: estimating a positive self-decomposable Lévy density . . 56
2.6 Some additional remarks 57
2.7 Computing the density from the canonical function 58
2.8 Estimation of the intensity parameter λ 60
2.9 An alternative estimation method 61
2.10 Some technical results 63

2.11 Bibliographic notes 65

3 Parametric estimation for subordinators and induced OU-processes 67
3.1 Definition of the estimator 67
3.2 Consistency 69
3.3 Asymptotic behavior of the estimator 71
3.4 Example: discrete observations from a Gamma-process 77
3.5 Parametric estimation for OU-processes driven by a subordinator 80
3.5.1 Adaptations to the proof of asymptotic normality of the CME . . 80
3.5.2 Example: discrete observations from an Inverse Gaussian-OU-process 82
3.6 Appendix 84

4 Convergence rates of posterior distributions for Brownian semimartingale models 87
4.1 Introduction 87
4.2 Main result 89
4.3 Special cases 93
4.3.1 Signal in white noise 93
4.3.2 Perturbed dynamical system 93
4.3.3 Ergodic diffusion 98
4.4 Proofs 103
4.4.1 Proof of Theorem 4.2 103
4.4.2 Proof of Lemma 4.8 105
4.4.3 Proof of Lemma 4.9 107
4.4.4 Proof of Lemma 4.3 108
4.4.5 A technical result 109
4.4.6 Likelihoods 109

References 111

Samenvatting (Summary in Dutch) 115

Dankwoord 117

Curriculum Vitae 119

Introduction

In this thesis we study statistical estimation for stochastic processes which are modelled by a *stochastic differential equation* (SDE). In its most general form such an equation is given by

$$dX(t) = \beta(t, X)dt + \sigma(t, X)dZ(t), \quad t \geq 0, \tag{\dagger}$$

in which Z is a semimartingale and typically either β or σ is an unknown function, which we wish to estimate. Depending on the precise form of the driving semimartingale Z and the coefficients β and σ, a variety of models is covered by (†). Examples of such models are continuous time diffusions, jump-diffusions and Lévy processes. For each of these models one can also consider different observation schemes: we can either assume continuous-time or discrete-time observations. In practice, observations are always recorded at discrete points in time only. Therefore, the data constitute a time-series and one may well wonder why one would prefer a more complicated continuous-time model to a time-series model. The main reason for this hinges on the fact that a continuous-time model remains the same under any observation scheme, whereas time-series models have to be adjusted for e.g. irregularly spaced data. Hence, continuous-time models allow a unified study of the processes at different time scales.

During the past decade, many of the models covered by (†) have been utilized within the field of mathematical finance. Since the classical Black-Scholes model for asset returns has been shown to be inconsistent with observed data, many authors have proposed other models, fitting these data better. Eberlein and Keller (1995) propose hyperbolic Lévy processes for return processes, whereas Barndorff-Nielsen and Shephard (2001) stick to the original Black-Scholes framework, though replacing the constant volatility by a latent Lévy-driven Ornstein-Uhlenbeck process. Carr et al.(2002) use a four-parameter pure-jump Lévy process, which they call the CGMY-process. This model includes the variance-gamma model (cf. Madan and Seneta (1990)) and the tempered stable Lévy process of Koponen (1995).

To get an idea of the flexibility of the type of models covered by (†) we now give two examples:

(A) The *Cox-Ingersoll-Ross (CIR) process* (also known as *Feller's square root process*) is defined as the stationary solution to the SDE

$$dX(t) = -\lambda(X(t) - \xi)dt + \omega\sqrt{X(t)}dB(\lambda t), \quad \xi, \lambda, \omega > 0,$$

where B is a standard Brownian Motion. We suppose $\xi \geq \omega^2/2$ so that the process is reflecting at zero.

(B) A *Lévy driven Ornstein-Uhlenbeck-process* (OU-process) is defined as the stationary solution to the SDE

$$dY(t) = -\lambda Y(t)dt + dZ(\lambda t), \quad \lambda > 0,$$

where Z is an increasing pure-jump Lévy process.

Both models have the same covariance structure and, as the stationary distribution of the CIR-process is a Gamma distribution, we can, by an appropriate choice of Z in model *(B)*, match the marginal distributions of the processes. However, the sample paths differ in that the CIR-process has continuous sample paths whereas the process Y moves upward by jumps.

The introduction of increasingly more difficult models has resulted in new statistical estimation problems. In this thesis we deal with two particular cases:

- Z is an increasing Lévy process, also known as a *subordinator*, $\beta(t, X) = -\lambda X$ for a positive constant λ and $\sigma(t, X) = 1$. The resulting process is known as a *Lévy driven Ornstein-Uhlenbeck process* (OU-process). The observations are discrete in time, though we will allow for the sampling interval to tend to zero, as we obtain more observations. We study both parametric and nonparametric estimation for the Lévy density of the driving subordinator.
- Z is a Brownian semimartingale, β and σ are of a general form. The resulting model is known as the *Brownian semimartingale model.* We study the Bayesian rate of convergence of the posterior distribution in the idealized situation that we can monitor the process continuously in time (whence, providing an upper-bound on convergence rates in case of discrete-time observations).

Our main focus is on nonparametric estimation problems, i.e. problems in which the unknown parameter is an element of an infinite-dimensional set.

The structure of this thesis is as follows. In the first three chapters we treat both parametric and nonparametric estimation for subordinators and their induced Ornstein-Uhlenbeck processes. The fourth chapter is on Bayesian convergence rates in Brownian semi-martingale models. In the following two sections we give an overview of both parts.

Overview of chapters 1–3

The first estimation problem we deal with is motivated by the stochastic volatility model as proposed by Barndorff-Nielsen and Shephard (2001). In its simplest form, this model postulates that the return process of stock price $S = (S(t), t \geq 0)$ evolves by the SDE

$$dS(t) = \sqrt{X(t)}dB(t), \qquad (*)$$

in which B is a standard Brownian motion and X is a latent process (the *volatility process*) that is defined as the stationary solution to the SDE

$$dX(t) = -\lambda X(t)dt + dZ(\lambda t). \qquad (**)$$

Here $\lambda > 0$ is an intensity parameter and Z is a *subordinator*, an increasing pure-jump Lévy process, that is independent of B. The given model is an example of a stochastic volatility model, in which only one component of two coupled processes is observed.

In Chapter 1 we first discuss some general results on Lévy processes and infinitely divisible laws, such as the Lévy-Khintchine theorem and the Lévy-Itô-decomposition. The main point of these fundamental theorems is the representation of a general Lévy process in terms of its drift, Brownian motion part and jump part. The last of these three is characterized by the *Lévy measure* of the process, a measure that determines the frequency and magnitude of the jumps made by the process. We then move to some examples and discuss a few techniques to simulate the paths of a subordinator. This immediately enables us to simulate the corresponding OU-process, since the solution to the SDE in ($**$) is given by

$$X(t) = e^{-\lambda t}X(0) + \int_0^t e^{-\lambda(t-s)}dZ(\lambda s).$$

Figure 1.2 in Chapter 1 shows the trajectory of a simulated OU-process. Theorem 1.16 is an important result, it states that if the Lévy measure of Z satisfies a weak integrability condition, then a stationary OU-process exists. Moreover, the invariant distribution π of X is *self-decomposable*, with Lévy density $x \mapsto x^{-1}\rho(x,\infty)$, independent of λ. Here ρ denotes the Lévy measure of Z. If we put $k(x) = \rho(x,\infty)$, then k is of course decreasing and

$$\psi_k(t) := \int e^{itx}d\pi(x) = \exp\left(\int_0^\infty (e^{itx}-1)\frac{k(x)}{x}dx\right), \quad t \in \mathbb{R}.$$

In this way, we have a full characterization of the invariant distribution of X in the Fourier domain, which serves as a basis for our estimation procedure. Chapter 1 closes with a proof that X is β-mixing, a result that is being used in Chapters 2 and 3.

Although in Barndorff-Nielsen and Shephard (2001) the data are discrete observations from the process S in ($*$), we consider the statistically less complicated situation where the volatility process X itself is observed at discrete time instants $t_i = i\Delta$ ($i = 0,\dots,n-1$, $\Delta > 0$). The present work may be extended to handle the more complicated model by the addition of a deconvolution step, and hence may provide a first step towards estimating these models nonparametrically.

In Chapter 2 we deal with nonparametric estimation of k based on discrete time observations from X. It is important to realize that traditional estimation techniques such as maximum-likelihood and Bayesian estimation are cumbersome. This is caused by the intractability of the likelihood. By the Lévy-Khintchine theorem, a Lévy process has its natural parametrization in the Fourier domain and, in general, there exists no closed form for the marginal distribution of a Lévy process. For discrete time observations from a stationary OU-process, the estimation is further complicated by the fact that the observations are dependent, so that the transition density of the process is needed. The latter is even harder to get hold of than the marginal density.

For this reason an alternative estimation method is proposed, which we will now explain briefly. As a starting point, we take a sequence of preliminary estimators $(\tilde\psi_n, n \geq 1)$ such that for each t, $\tilde\psi_n(t) \to \psi_k(t)$, either almost surely or in probability. The canonical example of such an estimator is the *empirical characteristic function*, though other estimators are possible. To prove that the empirical characteristic function is an appropriate preliminary estimator, we can use the β-mixing result of Chapter 1. Given the sequence $(\tilde\psi_n)_n$, we define an estimator $\hat k_n$ for k as the minimizer of the random criterion function

$$k \mapsto \int (\log\tilde\psi_n(t) - \log\psi_k(t))^2 w(t)dt$$

over all decreasing functions k. Here w is a compactly supported weight-function and $\log\psi$ refers to the distinguished logarithm of a characteristic function ψ, which probabilists often call the cumulant function. For this reason, we call this estimator a *cumulant M-estimator* (CME). We give precise conditions for the existence of such an estimator. Moreover, we prove that it is consistent both for the case that Δ is fixed and $n \to \infty$, and for the case that $\Delta_n \downarrow 0$ such that $n\Delta_n \to \infty$. Along with the consistency proof we give some results on uniform convergence of random characteristic functions, which may be of independent interest. The estimator can be approximated numerically by a support-reduction algorithm. We then give some numerical examples, and extend our results to the case where the canonical function k is assumed to be convex decreasing or completely monotone.

In Chapter 3 we study the CME for the case that the canonical function is parametrized by a finite-dimensional parameter θ. Since the shape-restriction on the Lévy density of the invariant distribution π plays no role whatsoever in this case, we formulate the statistical problem as follows: estimate θ in case we have discrete-time observations from a stationary process with invariant law π satisfying

$$\psi_\theta(t) = \int e^{itx} d\pi(x) = \exp\left(\int_0^\infty (e^{itx} - 1) a_\theta(x) dx\right), \quad t \in \mathbb{R}.$$

Here a_θ is the density of a Lévy measure. We give precise condtions under which the CME exists and show its consistency under weak conditions. Moreover, in case we assume that the preliminary estimator is the empirical characteristic function we show that the estimator is asymptotically normal with normalizing rate $\sqrt{n}$. The chapter is complemented with some examples for which the actual computations are worked out.

Overview of chapter 4

In Chapter 4 we suppose that we observe the real-valued stochastic process $X^n = (X_t^n, 0 \leq t \leq T_n)$ defined through the SDE

$$dX_t^n = \beta^{\theta,n}(t, X^n)\, dt + \sigma^n(t, X^n)\, dB_t^n, \quad t \in [0, T_n], \quad X_0^n = X_0,$$

where B^n is a standard Brownian motion. The parameter θ is assumed to belong to a possibly infinite-dimensional set Θ^n. Based on a realization of X^n we wish to make inference on the parameter θ that determines the shape of the "drift coefficient" $\beta^{\theta,n}$. The "diffusion coefficient" σ^n is considered to be known, as it can be determined without error from continuous observation of the process. The natural number $n \in \mathbb{N}$ serves as an indexing parameter for this asymptotic setup, in which n tends to infinity. This *Brownian semimartingale model* contains the *diffusion model*, *the Gaussian white noise model* and *the perturbed dynamical system* as special cases.

The practical problem of computing estimators for θ based on discrete-time observations can be solved by Markov-Chain Monte Carlo (MCMC) techniques. See for instance Eraker (2001) and Olerian et al. (2001). The power of these MCMC-methods motivates us to investigate the properties of Bayesian estimators. The Bayesian paradigm postulates that all inference should be based on the posterior distribution $\Pi^n(\cdot|X^n)$, which is the conditional distribution of the parameter θ given the observation X^n. By minimizing the posterior loss for an appropriate loss function, we can then obtain derived Bayesian point estimators. At a first level we would like our estimation procedure to be consistent:

as we obtain more and more observations, the posterior will concentrate arbitrarily close to the true distribution of the parameter. For this property to hold true, we need to make sure that the support of the prior is sufficiently large. Hence, not every prior distribution will give satisfactory results. At a second level we can judge the quality of the estimation procedure by the rate of convergence at which the posterior concentrates on balls around the true parameter value. Different priors may yield different posteriors which in turn may have different performance in terms of rate of convergence.

Recently, Ghosal and Van der Vaart (2004) have given a general result on the rate of convergence of posterior distributions, in the case of non-i.i.d. observations. However, their result is not applicable to the current situation of interest and hence needs to be modified. In Chapter 4 we present such a modification. Our main result bounds the posterior rate of convergence in terms of the complexity of the model and the amount of prior mass given to balls centered around the true parameter. Deviations from the true parameter θ_0 are measured by the Hellinger semimetric, which is the random semimetric given by

$$h_n(\theta, \theta_0) = \left(\int_0^{T_n} \left(\frac{\beta^{\theta,n} - \beta^{\theta_0,n}}{\sigma^n} \right)^2 (t, X^n) dt \right)^{1/2}, \quad \theta \in \Theta^n.$$

For the special models mentioned above we work out our general result to a more tractable form. We then give some examples for specific priors, including priors based on wavelets and Dirichlet processes.

Publication details

Chapters 1–3 have resulted in the following articles:

> Jongbloed, G. Meulen, F.H. van der, Vaart, A.W. van der (2004) *Nonparametric inference for Lévy driven Ornstein-Uhlenbeck processes.* To appear in Bernoulli.

> Jongbloed, G. and Van der Meulen, F.H. (2005) *Parametric estimation for subordinators and induced OU-processes*, Tentatively accepted for publication in the Scandinavian Journal of Statistics.

Chapter 4 contains the contents of the following article:

> Van der Meulen, F.H., Van der Vaart, A.W. and Van Zanten, J.H. (2005) *Convergence rates of posterior distributions for Brownian semimartingale models*, Submitted to Bernoulli.

Chapter 1

Lévy processes and induced OU-processes

This chapter serves as a preparation to chapters 2 and 3. We review well-known results but also present some new results which may be of independent interest. In the first two sections we discuss infinite divisibility in connection with Lévy processes. In particular, we state the Lévy-Khintchine and the Lévy-Itô decomposition theorem. Special attention is given to positive self-decomposable distributions and subordinators (increasing Lévy processes). We do not give proofs of results, since during the last years there have appeared a number of textbooks on Lévy processes, including Sato (1999), Bertoin (1996) and Applebaum (2004), and for infinitely divisible distributions on the real line, Van Harn and Steutel (2004). In Section 1.3 we discuss two techniques for simulating Lévy processes: the inverse Lévy measure method and the rejection method. Section 1.4 is devoted to stationary OU-processes, induced by a subordinator. We state an existence result for these processes and show that their transition functions are of Feller-type. The latter implies the strong Markov property for these type of processes. Exploiting this feature, we prove in Section 1.5 that a stationary OU-process is β-mixing. In the proof we use that a first-order autoregressive process is φ-irreducible under weak conditions, a fact proved in Section 1.6. The last section of this chapter contains some additional proofs.

Our main motivation for studying these processes and some of their specific properties becomes clear in chapters 2 and 3. These chapters deal with related statistical estimation problems, whereas the present chapter is on probabilistic aspects.

1.1 Infinite divisibility

We start with the definition of an infinitely divisible random variable.

Definition 1.1 A random variable X is called infinitely divisible (ID) if for each $n \in \mathbb{N}$ there exists an i.i.d. sequence of random variables $Y_1^{(n)}, \ldots, Y_n^{(n)}$ such that

$$X \stackrel{d}{=} Y_1^{(n)} + \ldots + Y_n^{(n)}.$$

This definition can easily be put in terms of probability distributions: let μ denote the distribution of X, then X is ID if for each n there exists a probability distribution μ_n

such that $\mu = (\mu_n)^{*n}$. Here $*n$ denotes the n-th convolution power. Since the concept of infinite divisibility only involves the distribution of a random variable, we define a probability measure, or a characteristic function (ch.f.) to be ID if the corresponding random variable is ID. The distribution of an ID random variable can be characterized by its *generating triplet* through the celebrated *Lévy-Khintchine theorem.* Call a σ-finite Borel measure ν on $\mathbb{R}_0 = \mathbb{R} \setminus \{0\}$ a *Lévy measure* if

$$\int_{\mathbb{R}_0} (x^2 \wedge 1)\nu(dx) < \infty. \tag{1.1}$$

Equivalently, we can define a Lévy measure as a σ-finite measure on $\mathbb{R}$ satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}}(x^2 \wedge 1)\nu(dx) < \infty$. Denote by ψ_a the characteristic function (ch.f.) of a random variable or distribution a.

Theorem 1.2 (Lévy-Khintchine) *If μ is an infinitely divisible distribution on $\mathbb{R}$, then there exists a unique triplet (A, ν, γ) (referred to as the generating triplet), consisting of a number $\gamma \in \mathbb{R}$, a Lévy measure ν and a nonnegative number A, such that*

$$\psi_\mu(t) = \exp\Big[-\tfrac{1}{2}At^2 + i\gamma t + \int_{\mathbb{R}_0} \big(e^{itx} - 1 - itx\mathbf{1}_{[-1,1]}(x)\big)\nu(dx)\Big], \quad t \in \mathbb{R}. \tag{1.2}$$

The reverse statement also holds: given a generating triplet (A, ν, γ), there exists an infinitely divisible distribution μ with characteristic function given by (1.2).

Example 1.3

(i) If $\nu \equiv 0$, then μ is Gaussian with mean γ and variance A.

(ii) Let $Z_1, Z_2, \ldots$ be an i.i.d. sequence of random variables with distribution function F satisfying $F(\{0\}) = 0$. Let N be a random variable that is independent of the sequence $\{Z_i\}_{i\geq 1}$ and has the Poisson(λ) distribution. If we define the random variable X by $X := Z_1 + \ldots + Z_N$, then

$$\psi_X(t) = \exp\Big[\int_{\mathbb{R}} \big(e^{itx} - 1\big)\lambda F(dx)\Big], \quad t \in \mathbb{R},$$

from which we infer that the generating triplet of X is given by

$$(0, \lambda F, \lambda \int_{[-1,1]} xF(dx)).$$

The r.v. X is called compound Poisson.

(iii) Combining the preceding two items, we see that a sum of independent Gaussian and compound Poisson random variables almost yields (1.2). The delicate point is that ν can be a Lévy measure with $\nu(\mathbb{R}_0) = \infty$. In the examples for Lévy processes in the next section we also give examples for this case.

A subclass of the ID distributions is the class of *self-decomposable* (SD) distributions.

Definition 1.4 A random variable X is self-decomposable if for every $c \in (0, 1)$ there exists a random variable X_c, independent of X, such that $X \stackrel{d}{=} cX + X_c$.

In particular, all degenerate random variables are self-decomposable. In case X is self-decomposable, its Lévy measure ν takes a special form. It has a density with respect to Lebesgue measure (Sato (1999), Corollary 15.11) and

$$\nu(dx) = \frac{k(x)}{|x|} dx,$$

for $k : \mathbb{R} \to \mathbb{R}$ a function that is increasing on $(-\infty, 0)$, decreasing on $(0, \infty)$ and satisfies the integrability condition $\int_0^\infty (1 \wedge x^2) x^{-1} k(x) dx$. This function is known as the *canonical function*.

For completeness, we also give the definition of a *stable* distribution.

Definition 1.5 A random variable X is stable if there exist sequences $(c_n, n \in \mathbb{N})$ and $(d_n, n \in \mathbb{N})$ with each $c_n > 0$ such that

$$c_n X + d_n \stackrel{d}{=} X_1 + \ldots + X_n,$$

where $X_1, \ldots, X_n$ are independent copies of X.

It can be shown that necessarily $c_n = n^{1/\alpha}$, for some $\alpha \in (0, 2]$. The generating triplet of a stable random variable takes the form (Sato (1999), Section 14)

(i) if $\alpha = 2$: $\nu \equiv 0$, whence X is Gaussian with mean γ and variance A.

(ii) if $\alpha \neq 2$: then $A = 0$ and

$$\nu(dx) = \frac{c_1}{x^{1+\alpha}} \mathbf{1}_{(0,\infty)}(x) dx + \frac{c_2}{|x|^{1+\alpha}} \mathbf{1}_{(-\infty,0)}(x) dx,$$

where $c_1 \geq 0$, $c_2 \geq 0$ and $c_1 + c_2 > 0$.

The various characterizations in terms of Lévy measures reveal that the class of SD distributions encompasses the parametric class of *stable* distributions.

The three classes of probability distributions initially appeared in conjunction with the problem of determining which probability laws can appear as a weak limit of rescaled partial sums of random variables. The following result is taken from Van Harn and Steutel (2004), Chapter 1, Section 5. We use the symbol "$\rightsquigarrow$" to denote weak convergence.

Theorem 1.6

(i) A random variable X is ID if and only if it can be obtained as

$$X_{n,1} + \cdots + X_{n,k_n} \rightsquigarrow X \quad (n \to \infty),$$

where $k_n \uparrow \infty$, $X_{n,1}, \ldots, X_{n,k_n}$ are independent for every $n \in \mathbb{N}$, and for every $\varepsilon > 0$

$$\lim_{n\to\infty} \max_{1 \leq j \leq k_n} P(|X_{n,j}| \geq \varepsilon) = 0$$

(ii) A random variable X is SD if and only if it can be obtained as

$$\frac{1}{b_n}(Y_1 + \cdots + Y_n - a_n) \rightsquigarrow X \quad (n \to \infty),$$

with $Y_1, Y_2, \ldots$ independent, $a_n \in \mathbb{R}$, and $b_n > 0$ satisfying

$$\lim_{n\to\infty} b_n = \infty, \quad \lim_{n\to\infty} \frac{b_{n+1}}{b_n} = 1.$$

(iii) A random variable X is stable if and only if it can be obtained as

$$\frac{1}{b_n}(Y_1 + \cdots + Y_n - a_n) \rightsquigarrow X \quad (n \to \infty),$$

with $Y_1, Y_2, \ldots$ independent, identically distributed, $a_n \in \mathbb{R}$, and $b_n > 0$.

In chapters 2 and 3 on estimation we need the logarithm of a characteristic function. This logarithm cannot be defined as a fixed branch of an ordinary complex logarithm, since in general such a branch is discontinuous, while a characteristic function is continuous. Nevertheless, we have the following result from complex analysis (a proof can be found for example in Chung (2001), Section 7.6).

Theorem 1.7 *Suppose $\varphi : \mathbb{R} \to \mathbb{C}$ is continuous, $\varphi(0) = 1$ and $\varphi(x) \neq 0$ for all $x \in [-T, T]$. Then there exists a unique continuous function $g : [-T, T] \to \mathbb{C}$ such that $g(0) = 0$ and $\exp(g(x)) = \varphi(x)$. The corresponding statement when $[-T, T]$ is replaced by $(-\infty, \infty)$ is also true.*

The function g is referred to as the *distinguished logarithm*. In case φ is a ch.f., g is called a *cumulant function*. Since an infinitely divisible distribution φ has no zeros (see e.g. Sato (1999), Lemma 7.5), we conclude that we can attach to ψ a unique continuous function g such that $e^{g(t)} = \psi(t)$ and $g(0) = 0$.

1.1.1 Positive self-decomposable distributions

In view of the estimation problem we consider in chapters 2 and 3, we give special attention to the class of positive SD-distributions. A random variable X is self-decomposable on $\mathbb{R}_+$ if and only if its characteristic function takes the form

$$\psi_X(t) = \exp\left[i\gamma_0 t + \int_0^\infty \left(e^{itx} - 1\right)\frac{k(x)}{x}dx\right], \quad t \in \mathbb{R},$$

where $\gamma_0 \geq 0$ and k is a decreasing function on $(0, \infty)$ satisfying $\int_0^1 k(x)dx + \int_1^\infty x^{-1}k(x)dx < \infty$. We choose k to be right-continuous.

For later purposes we remark that *(i)* the infimum of the support of X is given by γ_0, *(ii)* the class of positive SD-distributions is closed under weak convergence (Proposition V.2.3. in Van Harn and Steutel (2004)), *(iii)* the distribution of X is either absolutely continuous with respect to Lebesgue measure or degenerate (Theorem 27.13 in Sato (1999)).

A few examples of positive SD distributions are given below.

Example 1.8

(i) Let X be Gamma(c, α) distributed with density f given by

$$f(x) = \frac{\alpha^c}{\Gamma(c)} x^{c-1} e^{-\alpha x} \mathbf{1}_{\{x>0\}}, \quad c,\, \alpha > 0.$$

The ch.f. and canonical function are given by $\psi(t) = (1 - \alpha^{-1}it)^{-c}$ and $k(x) = ce^{-\alpha x}$ respectively.

(ii) Call a random variable X *tempered stable* (see e.g. Barndorff-Nielsen and Shephard (2002)) if it is infinitely divisible with Lévy measure

$$\rho(dx) = Ax^{-\kappa-1}e^{-Bx}dx, \quad x > 0,$$

for constants $A > 0$, $B \geq 0$ and $\kappa \in (0,1)$. We write $X \sim TS(\kappa, A, B)$. The form of the Lévy measure implies that X is positive SD with canonical function $k(x) = Ax^{-\kappa}e^{-Bx}\mathbf{1}_{\{x>0\}}$, which is decreasing. In general, no closed form expression for the density function of X is known in terms of elementary functions (however, there do exists series expansions, see Chapter XVII.6 in Feller (1971)). The Laplace transform of X is easily computed: for $t > -B$,

$$\begin{aligned}
\log \varphi(t) &:= \log Ee^{-tX} = \int_0^\infty (e^{-tx} - 1)\rho(dx) \\
&= A(t+B)^\kappa \int_0^\infty e^{-u}u^{-\kappa-1}du - AB^\kappa \int_0^\infty e^{-u}u^{-\kappa-1}du \\
&= A\Gamma(-\kappa)((t+B)^\kappa - B^\kappa) = A\tfrac{\Gamma(1-\kappa)}{\kappa}(B^\kappa - (t+B)^\kappa).
\end{aligned}$$

From this expression we compute the ch.f. (as in Example 2.13 in Sato (1999)). Let $D := \{z \in \mathbb{C} : \Re z > -B\}$. As in the proof of Lemma 2.26 in Chapter 2, we can extend the definition of φ to D so that φ is analytic on D. Define the function $\Phi : D \to \mathbb{C}$ by

$$\Phi(z) = \exp\left[A\tfrac{\Gamma(1-\kappa)}{\kappa}(B^\kappa - (z+B)^\kappa)\right].$$

For all $z \in D$, the function $z \mapsto (z+B)^\kappa$ is defined by $(z+B)^\kappa = e^{\kappa \log(z+B)}$, where we take the branch of the logarithm that corresponds to the principal value of the logarithm. That is, if $\log(z+B) = \rho e^{i\theta}$, then we take $\theta \in [-\pi, \pi)$ and $\log|z+B| = \log|\rho| + i\theta$. Then Φ is analytic on D and agrees with φ on $(-B, \infty)$. Therefore, $\varphi = \Phi$ on D. In particular, for $t \in \mathbb{R}$

$$\begin{aligned}
\varphi(-it) &= Ee^{itX} = \Phi(-it) = \exp\left[A\tfrac{\Gamma(1-\kappa)}{\kappa}(B^\kappa - (B-it)^\kappa)\right] \\
&= \exp\left[A\tfrac{\Gamma(1-\kappa)}{\kappa}\left(B^\kappa - |B-it|^\kappa e^{-i\kappa \mathrm{Sgn}(t)\arctan(t/B)}\right)\right].
\end{aligned}$$

If $B = 0$ this remains true under the convention that $\arctan(t/B) = \pi/2$ (its limit value). In case $B > 0$ all positive moments exists, and are given by $EX = -\frac{d}{dt}Ee^{-tX}|_{t=0} = A\Gamma(1-\kappa)B^{\kappa-1}$.

(iii) *Positive α-Stable* distributions arise as a subclass of the TS-distributions. If we take $B = 0$, $\alpha := \kappa$ and

$$A = A_\alpha = \frac{\alpha}{\Gamma(1-\alpha)\cos(\pi\alpha/2)},$$

then X is called positive stable with index α. Its ch.f. equals

$$\psi(t) = \exp\left(-|t|^\alpha \left[1 - i\tan\left(\frac{\pi\alpha}{2}\right)\mathrm{Sgn}(t)\right]\right).$$

Its corresponding canonical function is given by $k(x) = A_\alpha x^{-\alpha}$. Note that $A_{1/2} = 1/\sqrt{2\pi}$. Only if $\alpha = 1/2$, a closed form expression in terms of elementary functions, for the density function of X is known. In this case $f(x) = \frac{1}{\sqrt{2\pi}} x^{-3/2} e^{-1/(2x)} \mathbf{1}_{\{x>0\}}$. The probability distribution with this density is called the Lévy distribution. If Z has a standard normal distribution, then W defined by $W = 1/Z^2$ if $Z \neq 0$ and $W = 0$ otherwise, has a Lévy distribution.

(iv) *Inverse Gaussian distributions* also arise as a subclass of the TS-distributions: if we take $B = \gamma^2/2$, $\kappa = 1/2$ and $A = \delta/\sqrt{2\pi}$, then X is called Inverse Gaussian. The parameters are $\delta > 0$ and $\gamma \geq 0$ and we write $X \sim IG(\delta, \gamma)$. The probability density function of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} \delta e^{\delta\gamma} x^{-3/2} \exp(-(\delta^2 x^{-1} + \gamma^2 x)/2) \mathbf{1}_{\{x>0\}}.$$

See for example Barndorff-Nielsen and Shephard (2001a).

1.2 Lévy processes

Let $Z \equiv (Z(t), t \geq 0)$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$.

Definition 1.9 We say that Z is a Lévy process if

(i) Z(0)=0 a.s.

(ii) Z has independent increments. That is, for any choice of $n \geq 1$ and $0 \leq t_0 < t_1 < \cdots < t_n$, $(Z(t_j) - Z(t_{j-1}), 1 \leq j \leq n)$ are independent.

(iii) Z has stationary increments. That is, for all $0 \leq s < t < \infty$, $Z(t) - Z(s) \stackrel{d}{=} Z(t-s)$.

(iv) Z is stochastically continuous. That is,

$$\forall\, \varepsilon > 0 \quad \forall\, s \geq 0 \quad \lim_{t \to s} P(|Z(t) - Z(s)| > \varepsilon) = 0.$$

From this definition it easily follows that the distribution of $Z(t)$ is ID for each $t \geq 0$. Therefore, the distribution of the process Z is determined by the generating triplet of $Z(1)$. The converse also holds: for each ID distribution μ there exists a Lévy process Z such that the law of $Z(1)$ equals μ (Theorem 7.10 in Sato (1999)). By Theorem 30 in Chapter I of Protter (2004) we can safely assume that the sample paths of Z are right-continuous with existing left-hand limits.

By a careful analysis of the jumps of Z, the Lévy-Itô decomposition theorem can be deduced. To state this result, we first need to introduce some notation. Let $\Delta Z(s) = Z(s) - Z(s-)$ and define a random measure N on $([0, \infty) \times \mathbb{R}_0, \mathcal{B}([0, \infty)) \otimes \mathcal{B}(\mathbb{R}_0))$ by

$$N((s,t], A) = \#\{s < u \leq t \;:\; \Delta Z(u) \in A\} = \sum_{s<u\leq t} \mathbf{1}_A(\Delta Z(s)),$$

for $0 \leq s \leq t < \infty$ and $A \in \mathcal{B}(\mathbb{R}_0)$. Thus, $N((s,t], A)$ counts the number of jumps of Z in the time interval $(s, t]$ with size in A. We write $N(t, A)$ for $N([0,t], A)$. The random measure N gives rise to a σ-finite measure on $\mathcal{B}(\mathbb{R}_0)$ defined by $\nu(\cdot) = EN(1, \cdot)$. This measure is called the *Lévy measure* of Z (and can be shown to satisfy (1.1)). Also, N can be shown to be a *Poisson random measure* with mean measure $Leb \times \nu$ (Leb denotes the Lebesgue measure). This implies

- For each $t > 0$, $\omega \in \Omega$, $N(t, \cdot)(\omega)$ is a counting measure on $\mathcal{B}(\mathbb{R}_0)$.
- for each $A \in \mathcal{B}(\mathbb{R}_0)$, $(N(t, A), t \geq 0)$ is a Poisson process with intensity $\nu(A)$.

We define the compensated Poisson random measure $\tilde{N}$ by

$$\tilde{N}(t, A) := N(t, A) - t\nu(A), \quad t \geq 0.$$

Theorem 1.10 *[Lévy-Itô decomposition] Let Z be a Lévy process. Then Z has a decomposition*

$$Z(t) = \gamma t + \sqrt{A}B(t) + \int_{\{|x|\leq 1\}} x\tilde{N}(t, dx) + \int_{\{|x|>1\}} xN(t, dx) \tag{1.3}$$

where B is a standard Brownian Motion, independent of N, $\gamma \in \mathbb{R}$ and $A \geq 0$.

Main steps of the proof: Since $t \mapsto \int_{\{|x|>1\}} xN(t, dx)$ is a Lévy process, the process Y, obtained from Z by removing all jumps that exceed 1,

$$Y(t) := Z(t) - \int_{\{|x|>1\}} xN(t, dx)$$

is also a Lévy process. As Y is a Lévy process with bounded jumps, all its moments $E|Y(t)|^m$, $m = 1, 2, \ldots$ exist (Protter (2004), Theorem 34 in Chapter 1). Let $\gamma = EY(1)$ and define the centered process $\tilde{Y}$ by

$$\tilde{Y}(t) := Y(t) - EY(t) = Z(t) - \int_{\{|x|>1\}} xN(t, dx) - \gamma t,$$

This process is a mean zero Lévy process (with jumps bounded by 1) and hence a martingale. For each $n \geq 1$, let $A_n := \{x \in \mathbb{R} : \frac{1}{n+1} \leq |x| \leq \frac{1}{n}\}$. It can be shown that the $L^2 - \lim_{n\to\infty} \int_{A_n} x\tilde{N}(t, dx)$ exists and equals $\tilde{Y}_d(t) := \int_{\{|x|\leq 1\}} x\tilde{N}(t, dx)$. [Here L^2 refers to the space of equivalence classes of L^2-martingales, endowed with the norms $(\|\cdot\|_t, t \geq 0)$ where $\|X\|_t = \left(E \sup_{0\leq s\leq t} |X(s)|^2\right)^{1/2}$. This space is complete. Now $L^2 - \lim_{n\to\infty} X_n = X$ if and only if $\|X_n - X\|_t \to 0$ as $n \to \infty$ for all $t \geq 0$.] Next, we split up

$$\tilde{Y}(t) = \tilde{Y}_c(t) + \tilde{Y}_d(t),$$

where the remainder part $\tilde{Y}_c$ is defined by

$$\tilde{Y}_c(t) := L^2 - \lim_{n\to\infty} (Y(t) - \tilde{N}(t, A_n)).$$

The proof is concluded by showing that the martingale $\tilde{Y}_c$ has continuous sample paths and that $Ee^{iu\tilde{Y}_c(t)} = e^{-tAu^2/2}$, which implies that $\tilde{Y}_c$ is a Brownian Motion with variance A. □

In view of (1.3), we interpret A as the variance of the Gaussian part and ν as the parameter for the jump-part of the Lévy process Z. In Example 1.11*(iii)* we explain the need for compensating the small jumps of N.

According to Definition 11.9 in Sato (1999) a basic classification of Lévy processes can be made in the following way: if Z is a Lévy process generated by (A, ν, γ), then it is said to be of *(i)* type A, if $A = 0$ and $\nu(\mathbb{R}) < \infty$.

(ii) type B, if $A = 0$, $\nu(\mathbb{R}) = \infty$ and $\int_{|x|\le 1} |x|\nu(dx) < \infty$.

(iii) type C, if $A \neq 0$ or $\int_{|x|\le 1} |x|\nu(dx) = \infty$.

In case $\nu(\mathbb{R}) = \infty$, the process makes infinitely many jumps in each finite interval: the jump times are countable and dense in $[0, \infty)$ (Theorem 21.3 in Sato (1999)).

If Z is of type A or B (C), then almost surely, the sample paths of Z are of finite (infinite) variation on bounded time intervals (Theorem 21.9 in Sato (1999)). We now give a number of examples of Lévy processes, which should give the reader a flavor of the type of behavior Lévy processes can have.

Example 1.11

(i) Let $(N(t), t \ge 0)$ be a Poisson process with parameter λ and let $(\xi_j, j \ge 1)$ be a sequence of i.i.d. random variables, independent of N. Let ν be a finite measure on $\mathbb{R}$ such that $\nu(\{0\}) = 0$. Suppose $P(\xi_1 \in \cdot) = \nu(\cdot)/\nu(\mathbb{R})$. Define $X(t) = \sum_{j=1}^{N(t)} \xi_j$, then X is a compound Poisson process. By calculating the characteristic function of $X(1)$, it is easily seen that X is a Lévy process with generating triplet $(0, \nu, 0)$.

(ii) A standard Brownian motion with drift γ is a Lévy process with generating triplet $(1, 0, \gamma)$.

(iii) We now give an example of a pure-jump Lévy process with infinite Lévy measure. We follow Shiryaev (1999), p. 204/205. Let $(\lambda_k, k \ge 1)$ be a sequence of positive numbers and let $(\beta_k, k \ge 1)$ be a sequence of non-zero real numbers such that

$$\sum_{k=1}^{\infty} \lambda_k \beta_k^2 < \infty. \tag{1.4}$$

Let $(N^{(k)}, k \ge 1)$ be a sequence of independent Poisson process with parameters $(\lambda_k, k \ge 1)$. Set for each $n \in \mathbb{N}$

$$Z^{(n)}(t) := \sum_{k=1}^{n} \beta_k (N^{(k)}(t) - \lambda_k t),$$

then

$$Ee^{iuZ^{(n)}(t)} = \exp\Big(t\Big[-iu\sum_{k=1}^{n} \beta_k\lambda_k + \int (e^{iux} - 1)\nu^{(n)}(dx)\Big]\Big), \quad u \in \mathbb{R}. \tag{1.5}$$

Hence, the generating triplet of $Z(n)$ is given by $(0, \nu^{(n)}, \gamma_n)$ for

$$\nu^{(n)}(dx) = \sum_{k=1}^{n} \lambda_k \delta_{\{\beta_k\}}(dx)$$

and $\gamma_n = -\sum_{k=1}^{n} \beta_k \lambda_k \mathbf{1}_{\{|\beta_k|>1\}}$. Each $Z^{(n)}$ is a square integrable martingale, and for $n \ge m \ge 1$

$$\begin{aligned} E\Big(\sup_{0\le s\le t} \big|Z^{(n)}(s) - Z^{(m)}(s)\big|\Big)^2 &\le 4E\big|Z^{(n)}(t) - Z^{(m)}(t)\big|^2 \\ &\le \sum_{k=m+1}^{n} \beta_k^2 E(N^{(k)}(t) - \lambda_k t)^2 = t\sum_{k=m+1}^{n} \lambda_k \beta_k^2. \end{aligned}$$

where we applied Doob's inequality (Karatzas and Shreve (1991), p. 14) at the first inequality. Therefore, using (1.4), $(Z^{(n)})_n$ is a Cauchy sequence in the complete space L^2 (as defined in the proof of Theorem 1.10), and hence has a limit Z, which is also a Lévy process. From Theorem 8.7 in Sato (1999) we infer that the generating triplet of Z is given by $(0, \nu, \gamma)$ with $\nu(dx) = \sum_{k=1}^{\infty} \lambda_k \delta_{\{\beta_k\}}(dx)$ and $\gamma = -\sum_{k=1}^{\infty} \beta_k \lambda_k \mathbf{1}_{\{|\beta_k|>1\}}$. Note that ν is indeed a Lévy measure: $\nu(\{0\}) = 0$ and, by (1.4), $\int (|x|^2 \wedge 1)\nu(dx) < \infty$. The process Z can be represented as $\sum_{k=1}^{\infty} \beta_k (N^{(k)}(t) - \lambda_k t)$.

As $\nu(\mathbb{R}) = \sum_{k=1}^{\infty} \lambda_k$, any sequence (λ_k) which sums to infinity and satisfies (1.4) gives rise to a Lévy process with infinitely many jumps on bounded time intervals.

From (1.5) we see that we cannot take the limit $n \to \infty$ in a straightforward manner if

$$\sum_{k=1}^{\infty} \beta_k \lambda_k \mathbf{1}_{\{|\beta_k| \leq 1\}} = \int_{\{|x| \leq 1\}} |x| \nu(dx) = \infty.$$

In this particular case we really need to compensate in the term $\int (e^{iux} - 1)\nu^{(n)}(dx)$ to obtain integrability. This is also the reason why we initially considered a sum of *compensated* Poisson processes.

In the following, assume $\int_{\{|x| \leq 1\}} |x| \nu(dx) = \infty$ (so Z is of type C) and also that all $|\beta_k| \leq 1$ (take e.g. $\lambda_k = \beta_k = 1/\sqrt{k}$). Then the Lévy process Z can be represented as $Z(t) = \int_{\{|x| \leq 1\}} x(N(t, dx) - t\nu(dx))$ where N is the Poisson random measure with mean measure $Leb \times \nu$. The process is an L^2-bounded martingale, which is neither continuous nor of finite variation. To see the last property, for $z > 0$

$$E \exp\Big(-z \sum_{0<s\leq 1} |\Delta Z(s)|\Big) = \exp\Big(-\int_{\mathbb{R}_0} (1 - e^{-z|x|})\nu(dx)\Big)$$

(Applebaum (2004), Theorem 2.3.8). As $\int_{\{|x| \leq 1\}} |x| \nu(dx) = \infty$,

$$\sum_{0<s\leq 1} |\Delta Z(s)| = \infty, \quad \text{with probability 1.}$$

If all β_k are positive, Z has positive jumps only and is fluctuating, not increasing. "An explanation is that such a process can exist only with infinitely strong drift in the negative direction, which cancels the divergence of the sum of jumps: but it causes a random continuous motion in the negative direction" (Sato (1999), p. 138). "These processes are the most mysterious by all" (Rogers and Williams (2000), p. 78). "The Lévy measure has become so fine that it is no longer capable of distinguishing small jumps from drift" (Applebaum (2004), p. xvii).

(iv) A *Gamma Lévy process* is a well-known example of a Lévy process with $\nu(\mathbb{R}) = \infty$. It is defined as the Lévy process for which $P(Z(t) \leq x)$ $(t > 0)$ has Gamma distribution with parameters ct and α (see Example 1.8*(i)*). As seen from this example, the Lévy measure of Z is given by

$$\nu(dx) = \mathbf{1}_{(0,\infty)}(x)\, c\, t \frac{e^{-\alpha x}}{x} dx,$$

which clearly satisfies $\nu(0,\infty) = \infty$. The *variance Gamma process* is the Lévy process given by

$$X(t) = \theta Z(t) + \sigma B(Z(t)), \quad t \geq 0,$$

where $\theta \in \mathbb{R}$, $\sigma > 0$ and B is a standard Brownian motion, independent of Z. This process was put forward by Madan and Seneta (1990) in the field of financial mathematics.

1.2.1 Increasing Lévy processes

A *subordinator* is (by definition) an increasing Lévy process. The next lemma (Sato (1999), Theorem 21.5) implies that a subordinator is always of type A or B.

Lemma 1.12 *A Lévy process Z with generating triplet (A,ν,γ) is a subordinator if and only if $A = 0$, $\int_{(-\infty,0)} \nu(dx) = 0$, $\int_{(0,1]} x\nu(dx) < \infty$ and $\gamma_0 = \gamma - \int_{(0,1]} x\nu(dx) \geq 0$. As a consequence,*

$$Ee^{itZ(1)} = \exp\left(i\gamma_0 t + \int_0^\infty (e^{itx} - 1)\nu(dx)\right), \quad \forall t \in \mathbb{R},$$

and the Lévy-Ito decomposition for Z takes the following form:

$$Z(t) = \gamma_0 t + \int_0^\infty xN(t,dx) = \gamma_0 t + \sum_{0\leq s\leq t} \Delta Z(s).$$

In this case, γ_0 is called the *drift*-parameter. Since the sample functions of a subordinator are increasing, they are of finite variation. However, if $\nu(\mathbb{R}_+) = \infty$, the number of jumps in bounded time-intervals is infinite, in which case we call the process of infinite activity. In the other case, if $\nu(\mathbb{R}_+) < \infty$, the process is a compound Poisson process and hence of finite activity. Example of infinite activity subordinators are the Inverse Gaussian Lévy process and the Gamma process.

1.3 Simulation of subordinators

A subordinator is either a compound Poisson process or an infinite activity process. Simulation of the first type of processes is straightforward, whereas for the latter type there exist series representations. We discuss a part of the article by Rosiński (2001), in which series representations for general Lévy processes are treated.

Let Z be an infinite activity subordinator with Lévy measure Q and suppose we want to simulate its sample paths on an interval $[0,T]$. By the Lévy-Itô Theorem this comes down to simulating the sample paths of $t \mapsto \int xN(t,dx)$, where N is a Poisson random measure (PRM) on $([0,T] \times (0,\infty), \mathcal{B}([0,T]) \otimes \mathcal{B}((0,\infty)))$ with mean measure $Leb_T \times Q$. Here Leb_T denotes Lebesgue measure, restricted to $[0,T]$. We write $N \sim \text{PRM}(Leb_T \times Q)$. Starting from a unit intensity Poisson process, we can construct the PRM N by *transformation* and *marking*. We now give the main results on these concepts. Throughout this section, we denote by δ_x the Dirac measure at x.

Transformation Let E, E' be locally compact spaces with countable bases. Let $\mathcal{E}, \mathcal{E}'$ be the associated σ-fields. If $T : (E, \mathcal{E}) \to (E', \mathcal{E}')$ is measurable and N is a PRM(μ) on E with points $\{X_n\}$, then $N' := N \circ T^{-1}$ is a PRM(μ') on E' with points $\{T(X_n)\}$, where $\mu' := \mu \circ T^{-1}$.

Marking Let N be a PRM(μ) as above. Let $\{J_n\}$ be i.i.d. random elements with values in E' and common distribution F. Suppose the Poisson process N and the sequence $\{J_n\}$ are independent. Then the point process

$$\sum_n \delta_{(X_n, J_n)},$$

on $E \times E'$ is PRM($\mu \times F$).

A proof of these results can be found in Resnick (1987), pp. 134-135.

Let $(\Gamma_i, i \geq 1)$ denote the arrival times of a unit intensity Poisson process on $\mathbb{R}$, then M defined by $M = \sum_{i=1}^{\infty} \delta_{\Gamma_i/T}$ is PRM($T \cdot Leb$), Leb denoting the Lebesgue measure on $\mathbb{R}$. Now let $\varphi : [0, \infty) \to \mathbb{R}$ be a measurable map, taking points Γ_i/T to points $\varphi(\Gamma_i/T)$, then the transformation result shows that

$$M' = \sum_{i=1}^{\infty} \delta_{\varphi(\Gamma_i/T)} \sim PRM((T \cdot Leb) \circ \varphi^{-1}).$$

Let M'' be the process on $[0,1] \times [0, \infty)$ obtained by adding independent marks $U_1, U_2, \ldots$, each uniformly distributed on $[0, T]$, to M'. That is,

$$M'' := \sum_{i=1}^{\infty} \delta_{(U_i, \varphi(\Gamma_i/T))},$$

then the marking result gives that $M'' \sim$PRM$((Leb_T/T) \times (T \cdot Leb) \circ \varphi^{-1})$ =PRM$(Leb_T \times (Leb \circ \varphi^{-1}))$.

By an appropriate choice of the transformation φ, we can obtain a Poisson random measure with mean measure $Leb_T \times Q$. In this section we discuss two such methods: the "inverse Lévy measure method" and the "rejection method". Both yield a series representation for Z. We start with the inverse Lévy measure method.

For the upper-tail mass of a Lévy measure Q, define the left-continuous decreasing function Q_+ by $Q_+(x) := Q([x, \infty))$ $(x > 0)$. Define an inverse function $Q_+^{\leftarrow}$ by

$$Q_+^{\leftarrow}(x) = \inf\{y > 0 : Q_+(y) \geq x\}$$

The set $A(x) = \{s : Q_+(s) \geq x\}$ is closed, which implies that $\inf A(x) \in A(x)$. This fact together with the monotonicity of $x \mapsto Q_+(x)$ implies that

$$Q_+^{\leftarrow}(y) \geq t \quad \text{if and only if} \quad y \leq Q_+(t)$$

Now let $B = [a, b)$ for $0 < a < b < \infty$ and take $\varphi(x) = Q_+^{\leftarrow}(x)$. Then

$$\begin{aligned} (Leb \circ \varphi^{-1})(B) &= Leb\{x : a \leq Q_+^{\leftarrow}(x) < b\} = Leb\{x : Q_+(b) < x \leq Q_+(a)\} \\ &= Q([a, \infty)) - Q([b, \infty)) = Q([a, b)). \end{aligned}$$

Therefore, M'' and N have the same mean measure. Proposition 2.1 in Rosiński (2001) asserts that the processes N and M'' can be defined on the same probability space such that

$$N = \sum_{i=1}^{\infty} \delta_{(U_i, Q_+^{\leftarrow}(\Gamma_i/T))}, \quad a.s.$$

This implies that a sample paths of the subordinator Z with Lévy measure Q can be obtained as

$$Z(t) = \sum_{i=1}^{\infty} Q_+^{\leftarrow}(\Gamma_i/T)\mathbf{1}_{\{U_i \le t\}}, \quad a.s. \quad t \in [0,T].$$

In case $Q_+^{\leftarrow}$ is difficult to compute, we can use the *rejection method.* Suppose Z_0 is another Lévy process with Lévy measure Q_0 such that $\frac{dQ}{dQ_0} \le 1$. Let (U_i) again be an i.i.d. sequence of Uniform$(0,T)$ random variables. Suppose that the PRM of Z_0 is given by N_0 and admits a representation $N_0 \stackrel{d}{=} \sum_{i=1}^{\infty} \delta_{(U_i, J_i^0)}$. Let $(V_i, i \ge 1)$ be an i.i.d. sequence of Uniform$(0,1)$ random variables, independent of N_0. Then $M := \sum_{i=1}^{\infty} \delta_{(U_i, V_i, J_i^0)}$ is PRM$(Leb_T \times Leb_1 \times Q_0)$ on $([0,T] \times [0,1] \times \mathbb{R}_0)$. Define $\varphi : ([0,T] \times [0,1] \times \mathbb{R}_0) \to ([0,T] \times \mathbb{R}_0)$ by

$$\varphi(u,v,j) = \big(u, j\mathbf{1}\{\tfrac{dQ}{dQ_0}(j) \ge v\}\big).$$

Set $M' := \sum_{i=1}^{\infty} \delta_{\varphi(U_i, V_i, J_i^0)}$, then M' is $PRM((Leb_T \times Leb_1 \times Q_0) \circ \varphi^{-1})$. Now, for $A \in \mathcal{B}([0,T])$ and $B \in \mathcal{B}(\mathbb{R}_0)$

$$\begin{aligned} &(Leb_T \times Leb_1 \times Q_0) \circ \varphi^{-1}(A \times B) \\ &\quad = (Leb_T \times Leb_1 \times Q_0) \circ \varphi^{-1}\Big(u \in A \text{ and } (v,j) : j\mathbf{1}\Big\{\frac{dQ}{dQ_0}(j) \ge v\Big\} \in B\Big) \\ &\quad = Leb_T(A) \int_B \int_{0 \le v \le \frac{dQ}{dQ_0}(j)} dLeb_1(v) Q_0(dj) = Leb_T(A)Q(B). \quad (1.6) \end{aligned}$$

Whence, M' has the same mean measure as N, and we can conclude in the same way as for the inverse Lévy measure method that

$$N = \sum_{i=1}^{\infty} \delta_{(U_i, J_i)}, \quad a.s., \text{ where } \quad J_i = J_i^0 \mathbf{1}_{\{\frac{dQ}{dQ_0}(J_i^0) \ge V_i\}},$$

which implies

$$Z(t) = \sum_{i=1}^{\infty} J_i \mathbf{1}_{\{U_i \le t\}}, \quad a.s. \quad t \in [0,T].$$

From this representation it is clear that the jumps of Z are obtained from the ones of Z_0 by a random cut-off.

We illustrate the inverse Lévy measure method for an α-stable process and the rejection method for a tempered stable process. Let $U_1, U_2, \ldots$ be a sequence of independent $Unif(0,T)$ random variables. Let $E_1, E_2, \ldots$ be a sequence of independent $Exp(1)$ random variables. Let $\Gamma_1 < \Gamma_2 < \ldots$ be the arrival times of a unit intensity Poisson process. Suppose all sequences are independent from each other.

Example 1.13

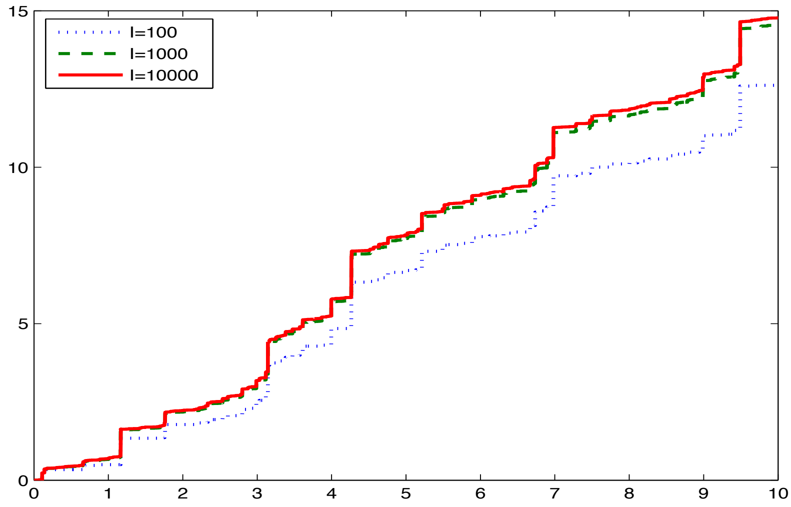

Figure 1.1: Simulation of an Inverse Gaussian Lévy process via the rejection method, using different truncation indices.

(i) Suppose we want to simulate from a $TS(\kappa, A, 0)$ Lévy process. Then $Q_+^{\leftarrow}(x) = (A/(\kappa x))^{1/\kappa}$ and we obtain by the inverse Lévy measure method

$$Z(t) = \sum_{i=1}^{\infty} \left(\frac{AT}{\kappa \Gamma_i}\right)^{1/\kappa} \mathbf{1}_{\{U_i \leq t\}}, \quad a.s. \quad t \in [0, T].$$

(ii) For a $TS(\kappa, A, B)$-Lévy process it is difficult to compute $Q_+^{\leftarrow}$. Therefore we use the rejection method with $Z_0 \sim TS(\kappa, A, 0)$. We have $\frac{\rho(dx)}{\rho_0(dx)} = e^{-Bx} \leq 1$ and we have just seen how to simulate the jumps of Z_0 via the inverse Lévy measure method. Therefore

$$Z(t) = \sum_{i=1}^{\infty} J_i^0 \mathbf{1}_{\{J_i^0 \leq E_i/B\}} \mathbf{1}_{\{U_i \leq t\}}, \quad a.s. \quad t \in [0, T], \tag{1.7}$$

where $J_i^0 = \left(\frac{AT}{\kappa \Gamma_i}\right)^{1/\kappa}$. Figure 1.1 shows a simulated path of an Inverse Gaussian Lévy process, with $\delta = 2$ and $\gamma = 1$. We truncated the infinite series in (1.7) at respectively $I = 100$, $I = 1000$ and $I = 10000$.

Remark 1.14 Another way to handle the infinite activity case, is to "throw away" all jumps of Z smaller than some prescribed small number and subsequently use a compound Poisson approximation for Z.

1.4 Ornstein-Uhlenbeck processes driven by a subordinator

In this section we define OU-processes, generated by a subordinator. Furthermore, we state the close relationship between self-decomposability and Lévy driven OU-processes. Additional details on this can be found in Section 17 in Sato (1999).

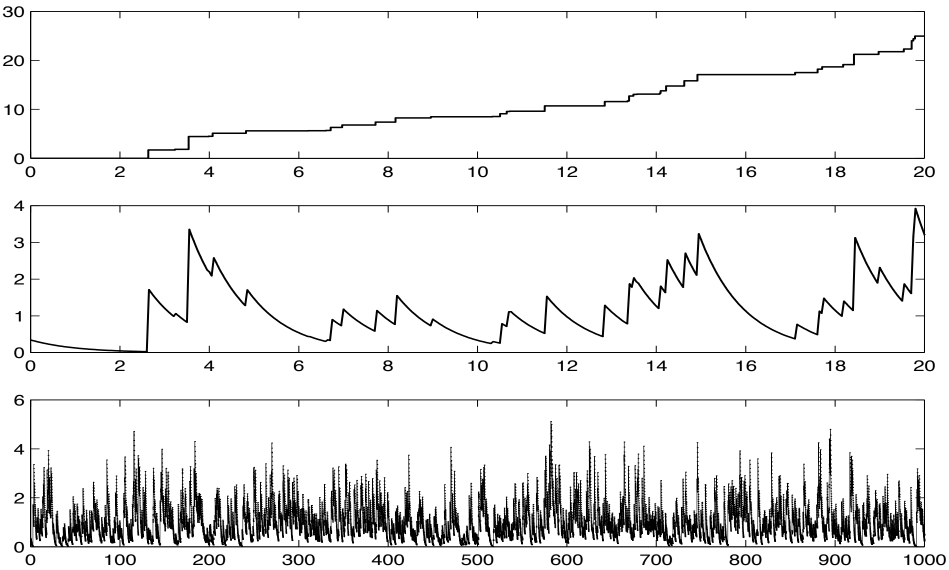

Figure 1.2: Top: simulation of the BDLP (compound Poisson process of intensity 2 with exponential jumps of expectation 1/2). Middle: corresponding OU-process with Gamma$(2, 2)$ marginal distribution. Bottom: OU-process on longer time-horizon.

For a given positive number λ and a given subordinator Z without drift component, consider the stochastic differential equation

$$dX(t) = -\lambda X(t)dt + dZ(\lambda t), \quad t \geq 0\,. \tag{1.8}$$

A solution X to this equation is called a Lévy-driven *Ornstein-Uhlenbeck (OU) process*, and the process Z is referred to as the *background driving Lévy process (BDLP)*. The auto-correlation of X at lag h can be expressed in the "intensity parameter" λ as $e^{-\lambda|h|}$.

It is easily verified that a (strong) solution $X = (X(t),\, t \geq 0)$ to the equation (1.8) is given by

$$X(t) = e^{-\lambda t}X(0) + \int_{(0,t]} e^{-\lambda(t-s)} dZ(\lambda s), \quad t \geq 0\,. \tag{1.9}$$

Up to indistinguishability, this solution is unique (Sato (1999), Section 17). Furthermore, since X is given as a stochastic integral with respect to a càdlàg semi-martingale, the OU-process $(X(t),\, t \geq 0)$ can be assumed càdlàg itself. The stochastic integral in (1.9) can be interpreted as a pathwise Lebesgue-Stieltjes integral, since the paths of Z are almost surely of finite variation on each interval $(0, t]$, $t \in (0, \infty)$. Figure 1.2 shows a simulation of a stationary OU-process with Gamma$(2, 2)$ marginal distribution.

Denote by $(\mathcal{F}_t^0)_{t\geq 0}$ the natural filtration of (X_t). That is, $(\mathcal{F}_t^0) = \sigma(X(u),\, u \in [0, t])$. As noted in Shiga (1990), Section 2, $(X(t), \mathcal{F}_t^0)$ is a temporally homogeneous Markov process. Denote by $(E, \mathcal{E})$ the state space of X, where $\mathcal{E}$ is the Borel σ-field on E. We take $E = [0, \infty)$. The transition kernel of X, denoted by $P_t(x, B)$ $(x \in E,\, B \in \mathcal{E})$, has characteristic function (Sato (1999), Lemma 17.1)

$$\int e^{izy} P_t(x, dy) = \exp\left(ize^{-\lambda t}x + \lambda \int_0^t g(e^{\lambda(u-t)}z)du \right), \quad z \in \mathbb{R}, \tag{1.10}$$

where g is the cumulant of $Z(1)$ (for the definition of a cumulant: see Theorem 1.7). The proof of this roughly runs as follows: first note that for a measurable and suitably integrable function h

$$E\left[\exp\left(i\langle z, \int_s^t h(u)dZ(u)\rangle\right)\right] = \exp\left[\int_s^t g(h(u)z)du\right], \quad s \le t. \tag{1.11}$$

This relation is easily proved for piecewise constant functions, the general form follows from approximation, which enables us to extend the result to (at least) continuous functions h. As a second step, we apply (1.11) with $h(u) = e^{-\lambda(t-u)}$ to obtain (1.10).

Let $b\mathcal{E}$ denote the space of bounded $\mathcal{E}$-measurable functions. The transition kernel induces an operator $P_t : b\mathcal{E} \to b\mathcal{E}$ by

$$P_t f(x) := \int f(y)P_t(x, dy) = \int f(e^{-\lambda t}x + y)P_t(0, dy)\,. \tag{1.12}$$

The second equality follows directly from the explicit solution (1.9). We call P_t the transition operator. Let $C_0(E)$ denote the space of continuous functions on E vanishing at infinity (i.e. $\forall \varepsilon > 0$ there exists a compact subset K of E such that $|f| \le \varepsilon$ on $E \setminus K$).

Proposition 1.15 *The transition operator of the OU-process is of Feller-type. That is,*
(i) $P_t C_0(E) \subseteq C_0(E)$ *for all* $t \ge 0$,
(ii) $\forall f \in C_0(E)$, $\forall x \in E$, $\lim_{t\downarrow 0} P_t f(x) = f(x)$.

For general notions concerning Markov processes of Feller type we refer to Chapter 3 in Revuz and Yor (1999).

Proof Let $f \in C_0(E)$, whence f is bounded. If $x_n \to x$ in E, then $f(e^{-\lambda t}x_n + y) \to f(e^{-\lambda t}x + y)$ in $\mathbb{R}$, by the continuity of f, for any $y \in \mathbb{R}$. By dominated convergence, $P_t f(x_n) \to P_t f(x)$, as $n \to \infty$. Hence, $P_t f$ is continuous. Again by dominated convergence, $P_t f(x) \to 0$, as $x \to \infty$.

For the second part, by dominated convergence $\int_0^t g(e^{\lambda(u-t)}z)du = \int_0^t g(e^{-\lambda u}z)du \to 0$, as $t \downarrow 0$. Here we use the continuity of the cumulant g and $g(0) = 0$. Then it follows from (1.10) that

$$\lim_{t\downarrow 0} \int e^{izy} P_t(x, dy) = e^{izx}\,.$$

Thus $P_t(x, \cdot)$ converges weakly to $\delta_x(\cdot)$ (Dirac measure at x):

$$\lim_{t\downarrow 0} \int f(y)P_t(x, dy) = \int f(y)\delta_x(dy) = f(x)\,, \quad \forall f \in C_b(E)\,.$$

Here $C_b(E)$ denotes the class of bounded, continuous functions on E. The result follows since $C_0(E) \subseteq C_b(E)$. □

The Feller property of X implies X is a *Borel right Markov process* (see the definitions in chapter 9 of Getoor (1975)). We will need this result in Section 1.5.

Since P_t is Feller, X satisfies the strong Markov property (Revuz and Yor (1999), Theorem III.3.1). In order to state a useful form of the latter property, we define a *canonical OU-process* on the space $\Omega = D[0, \infty)$, by setting $X(t, \omega) = \omega(t)$, for $\omega \in \Omega$

(here $D[0,\infty)$ denotes the space of càdlàg functions on $[0,\infty)$, equipped with its σ-algebra generated by the cylinder sets). By the Feller property, this process exists (Revuz and Yor (1999), Theorem III.2.7). Let ν be a probability measure on $(E,\mathcal{E})$ and denote by Q_ν the distribution of the canonical OU-process on $D[0,\infty)$ with initial distribution ν. We write Q_x in case ν is the Dirac measure at x. For $t \in [0,\infty)$, we define the shift maps $\theta_t : \Omega \to \Omega$ by $\theta_t(\omega(\cdot)) = \omega(\cdot + t)$.

Next, we enlarge the filtration by including certain null sets. Denote by $\mathcal{F}^\nu_\infty$ the completion of $\mathcal{F}^0_\infty = \sigma(\mathcal{F}^0_t, t \geq 0)$ with respect to Q_ν. Let $(\mathcal{F}^\nu_t)$ be the filtration obtained by adding to each $\mathcal{F}^0_t$ all the Q_ν-negligible sets of $\mathcal{F}^\nu_\infty$. Finally, set $\mathcal{F}_t = \bigcap_\nu \mathcal{F}^\nu_t$ and $\mathcal{F}_\infty = \bigcap_\nu \mathcal{F}^\nu_\infty$, where the intersection is over all initial probability measures ν on $(E,\mathcal{E})$. In the special case of Feller processes, it can be shown that the filtration $(\mathcal{F}_t)$ obtained in this way is automatically right continuous (thus, it satisfies the "usual hypotheses"). See Proposition III.2.10 in Revuz and Yor (1999). Moreover, X is still Markov with respect to this completed filtration (Revuz and Yor (1999), Proposition III.2.14). The strong Markov property can now be formulated as follows. Let Z be an $\mathcal{F}_\infty$-measurable and positive (or bounded) random variable. Let T be an $\mathcal{F}_t$-stopping time. Then for any initial measure ν,

$$E_\nu(Z \circ \theta_T | \mathcal{F}_T) = E_{X(T)}(Z)\,, \quad P_\nu - \text{a.s. on } \{T < \infty\}\,. \tag{1.13}$$

Here E_x denotes expectation under Q_x and $\mathcal{F}_T = \{A \in \mathcal{F} \,:\, A \bigcap \{T \leq t\} \in \mathcal{F}_t, \,\forall\, t \geq 0\}$. The expectation on the right-hand side is interpreted as $E_x Z$, evaluated at $x = X(T)$.

In Section 1.5 we will apply the strong Markov property to random times such as $\sigma_A := \inf\{t \geq 0 \,:\, X(t) \in A\}$ with $A \in \mathcal{E}$. By Theorem III.2.17 in Revuz and Yor (1999), σ_A is an $(\mathcal{F}_t)$-stopping time.

The following theorem gives a condition in terms of the process Z (called (A)), such that there exists a stationary solution to (1.8). Moreover, it shows that under this condition, the marginal distribution of this stationary solution is *self-decomposable* with canonical function determined by the Lévy measure of the underlying process Z.

Theorem 1.16 *Suppose Z is an increasing Lévy process with Lévy measure ρ (which is by definition the Lévy measure of $Z(1)$). Suppose ρ satisfies the integrability condition*

$$(A) \qquad \int_1^\infty \log x \rho(dx) < \infty\,,$$

then $P_t(x,\cdot)$ converges weakly to a limit distribution π as $t \to \infty$ for each $x \in E$ and each $\lambda > 0$. Moreover, π is self-decomposable with canonical function $k(x) = \rho(x,\infty)\mathbf{1}_{(0,\infty)}(x)$. Furthermore, π is the unique invariant probability distribution of X.

Theorem 24.10*(iii)* in Sato (1999) implies that π has support $[0,\infty)$.

We end this section with two examples of Lévy driven OU-processes. These examples are closely related to the ones given in examples 1.8*(i)* and 1.8*(iv)*.

Example 1.17

(i) Let $(X(t),\, t \geq 0)$ be the OU-process with $\pi = Gamma(c,\alpha)$. From the previous theorem and Example 1.8*(i)* it follows that the BDLP $(Z(t),\, t \geq 0)$ has Lévy measure ρ satisfying $\rho(dx) = c\,\alpha e^{-\alpha x} dx$ (for $x > 0$). Since $\rho(0,\infty) < \infty$, Z is a

compound Poisson process. By examining the characteristic function of $Z(1)$, we see that the process Z can be represented as $Z(t) = \sum_{i=1}^{N(t)} Y_i$, where $(N(t),\, t \geq 0)$ is a Poisson process of intensity c, and $Y_1, Y_2, \ldots$ is a sequence of independent random variables, each having an exponential distribution with parameter α. Figure 1.2 corresponds to the case $c = \alpha = 2$.

(ii) Let $(X(t),\, t \geq 0)$ be the OU-process with $\pi = IG(\delta, \gamma)$. Similarly to *(i)* we obtain for the Lévy measure ρ of the BDLP Z the following expression

$$\rho(dx) = \left(\frac{\delta}{2\sqrt{2\pi}} \frac{1}{x\sqrt{x}} e^{-\gamma^2 x/2} + \frac{\delta\gamma^2}{2\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\gamma^2 x/2} \right) dx \,, \quad x > 0 \,.$$

Write $\rho = \rho^{(1)} + \rho^{(2)}$. Then $(Z(t)\,, t \geq 0)$ can be constructed as the sum of two independent Lévy processes $Z^{(1)}$ and $Z^{(2)}$, where $Z^{(i)}$ has Lévy measure $\rho^{(i)}$ $(i = 1, 2)$. It is easily seen that $Z^{(1)}(1) \sim IG(\delta/2, \gamma)$. Note that

$$\int_0^\infty \rho^{(2)}(dx) = \int_0^\infty \frac{\delta\gamma^2}{2\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\gamma^2 x/2} dx < \infty,$$

so that $Z^{(2)}$ is a compound Poisson process. Some calculations show that we can construct $Z^{(2)}$ as $Z^{(2)}(t) = \frac{1}{\gamma^2} \sum_{i=1}^{N(t)} W_i^2$, where $(N(t),\, t \geq 0)$ is a Poisson process of intensity $\delta\gamma/2$, and $W_1, W_2, \ldots$ is a sequence of independent standard normal random variables. As $\rho(0, \infty) = \infty$, this OU-process is a process of infinite activity: it has infinitely many jumps in bounded time intervals.

1.5 A condition for the OU-process to be β-mixing

Let $X = (X(t), t \geq 0)$ be a stationary Lévy driven OU-process. The following theorem is the main result of this section.

Theorem 1.18 *If condition (A) of Theorem 1.16 holds, then the Ornstein-Uhlenbeck process X is β-mixing.*

This result will be used in Section 2.5 to obtain consistency proofs for some estimators, that will be defined in the next chapter. For the remainder of this section we will assume (A) holds. Theorem 1.16 then implies that there exists a unique invariant probability measure π_0.

A general definition on β-mixing numbers is given in e.g. Rio (2000). By Proposition 1 in Davydov (1973), the *β-mixing coefficients* for a stationary continuous-time Markov process X with invariant distribution π and transition kernel P are given by

$$\beta_X(t) = \int_E \pi(dx) \|P_t(x, \cdot) - \pi(\cdot)\|_{TV} \,, \quad t > 0 \,.$$

Here, $\|\cdot\|_{TV}$ denotes the total variation norm and π the initial distribution. The process is said to be β-mixing if $\beta_X(t) \to 0$, as $t \to \infty$. The analogous definitions for the discrete-time case are obvious. Dominated convergence implies that the following condition is sufficient for X to be β-mixing:

$$\lim_{t \to \infty} \|P_t(x, \cdot) - \pi(\cdot)\|_{TV} = 0 \,, \quad \forall\, x \in E \,. \tag{1.14}$$

That is, it suffices to prove that the transition probabilities converge in total-variation to the invariant distribution for each initial state $x \in E$. The next theorem, taken from Meyn and Tweedie (1993a) (Theorem 6.1), can be used to verify this condition. Below we explain the concepts that are used in this theorem.

Theorem 1.19 *Suppose that X is positive Harris recurrent with invariant probability distribution π. Then (1.14) holds if and only if some skeleton chain is φ-irreducible.*

In this section we first prove that the 1-skeleton-chain, obtained from X, is φ-irreducible (Corollary 1.23). Subsequently, we show that X is positive Harris recurrent (Lemma 1.24). By an application of Theorem 1.19, Theorem 1.18 then follows immediately.

We start with some definitions from the general theory of stability of continuous time Markov processes. These correspond to the ones used in Theorem 1.19. For more details, see Meyn and Tweedie (1993a). Recall from Section 1.4 that Q_x denotes the distribution of the OU-process started at x. For a measurable set A we let

$$\sigma_A = \inf\{t \geq 0 \mid X(t) \in A\}, \qquad \eta_A = \int_0^\infty \mathbf{1}_{\{X(t)\in A\}} dt\,.$$

Thus, σ_A denotes the first *hitting time* of the set A and η_A denotes the *time spent* in A by X. A Markov process is called *φ-irreducible* if for some non-zero σ-finite measure φ,

$$\varphi(A) > 0 \quad \Longrightarrow \quad E_x(\eta_A) > 0, \quad \forall x \in E\,,\ A \in \mathcal{E}\,.$$

The Markov process X is called *Harris recurrent* if for some non-zero σ-finite measure φ,

$$\varphi(A) > 0 \quad \Longrightarrow \quad Q_x(\eta_A = \infty) = 1, \quad \forall x \in E\,,\ A \in \mathcal{E}\,.$$

If X is a Borel right Markov process, then this condition can be shown to be equivalent to the following (Kaspi and Mandelbaum (1994)): for some non-zero σ-finite measure ψ,

$$\psi(A) > 0 \quad \Longrightarrow \quad Q_x(\sigma_A < \infty) = 1, \quad \forall x \in E\,,\ A \in \mathcal{E}\,. \tag{1.15}$$

The latter condition is generally more easily verifiable. The process is called *positive Harris recurrent* if it is Harris recurrent and admits an invariant *probability* measure.

The Δ-*skeleton* is defined as the Markov chain obtained by sampling the original process X at deterministic time points $\Delta, 2\Delta, \ldots$. In a slight abuse of notation, we shall henceforth denote this chain by (X_n) (thus, $X_n \equiv X_{n\Delta}$). The next proposition says that the 1-skeleton obtained from X constitutes a first order auto-regressive time series, with infinitely divisible noise terms.

Proposition 1.20 *If $\Delta = 1$, then the chain (X_n) satisfies the first order auto-regressive relation*

$$X_n = e^{-\lambda} X_{n-1} + W_n(\lambda)\,, \quad n \geq 1\,, \tag{1.16}$$

where $(W_n(\lambda))_n$ is an i.i.d. sequence of random variables distributed as

$$W_\lambda := \int_0^1 e^{\lambda(u-1)} dZ(\lambda u)\,.$$

Moreover, W_λ is infinitely divisible with Lévy measure κ given by

$$\kappa(B) = \int_B w^{-1} \rho(w, e^\lambda w] dw\,, \quad B \in \mathcal{E}\,. \tag{1.17}$$

The proof is given in Section 1.6.

Remark 1.21 Since

$$e^{-\lambda}Z(\lambda) \le \int_0^1 e^{\lambda(u-1)} dZ(\lambda u) \le Z(\lambda)\,,$$

W_λ has the same tail behaviour as $Z(\lambda)$. In particular, if $Z(1)$ has infinite expectation, so does W_λ.

The chain (X_n) is φ-irreducible if there exists a non-zero σ-finite measure φ, such that for all $B \in \mathcal{E}$ with $\varphi(B) > 0$, $\sum_{n=1}^{\infty} P_n(x, B) > 0$, for all $x \in E$. Proposition 6.3.5 in Meyn and Tweedie (1993) asserts that (X_n) as defined in (1.16) is φ-irreducible if the common distribution of the innovation-sequence $(W_n(\lambda))$ has a non-trivial absolutely continuous component with respect to Lebesgue measure. (For completeness, we give a proof of this result for the simple case of a one-dimensional autoregressive process in the next section.)

Lemma 1.22 *Let P^{W_λ} be the distribution of W_λ. Then P^{W_λ} has a non-trivial absolutely continuous component with respect to Lebesgue measure.*

Proof It follows from Proposition 1.20 that P^{W_λ} is infinitely divisible with Lévy measure κ. From (1.17), we see that κ is absolutely continuous with respect to Lebesgue measure.

First consider the case $\kappa[0,\infty) < \infty$. Then P^{W_λ} is compound Poisson, and hence (see equation 27.1 in Sato (1999)),

$$P^{W_\lambda}(\cdot) = e^{-\kappa[0,\infty)} \left(\delta_{\{0\}}(\cdot) + \sum_{k=1}^{\infty} \frac{\kappa^{*k}(\cdot)}{k!} \right)\,, \tag{1.18}$$

where δ_0 denotes the Dirac measure at 0 and * denotes the convolution operator. Since the convolution of two non-zero finite measures σ_1 and σ_2 is absolutely continuous if either of them is absolutely continuous (Sato (1999), Lemma 27.1), it follows from the absolute continuity of κ that the second term on the right-hand side of (1.18) constitutes the absolutely continuous part of P^{W_λ}.

Next consider the case $\kappa[0,\infty) = \infty$. Define for each $n = 1, 2, \ldots$, $\kappa_n(B) := \kappa(B \cap (1/n, \infty)$ for Borel sets B in $(0,\infty)$. Set $c_n = \kappa_n[0,\infty)$. Then $c_n < \infty$ and κ_n is absolutely continuous. Let $P_n^{W_\lambda}$ be the distribution corresponding to κ_n. As in the previous case we have

$$P_n^{W_\lambda}(\cdot) = e^{-c_n} \left(\delta_{\{0\}}(\cdot) + \sum_{k=1}^{\infty} \frac{\kappa_n^{*k}(\cdot)}{k!} \right)\,,$$

and $P_n^{W_\lambda}$ has an absolutely continuous component with respect to Lebesgue measure. Since P^{W_λ} contains $P_n^{W_\lambda}$ as a convolution factor, it follows that P^{W_λ} has a non-trivial absolutely continuous component with respect to Lebesgue measure. $\square$

Corollary 1.23 *The 1-skeleton chain (X_n) is φ-irreducible.*

Lemma 1.24 *Under condition (A), X is positive Harris-recurrent.*

Proof Let $\sigma_a = \inf\{t \geq 0 \,:\, X(t) = a\}$. We will prove $Q_x(\sigma_a < \infty) = 1$, for all $x, a \in E$. Then condition (1.15) is satisfied for any non-zero measure ψ on E.

First, we consider the case $x \geq a$. Since we assume (A), Lemma 1.28 from Section 1.7 applies:

$$\int_0^1 \frac{dz}{z} \exp\left(-\int_z^1 \frac{\lambda_\rho(y)}{y} dy\right) = +\infty\,. \tag{1.19}$$

Here λ_ρ is given as in (1.25). Theorem 3.3 in Shiga (1990) now asserts that $Q_x(\sigma_a < \infty) = 1$. for every $x \geq a > 0$.

Next, suppose $x < a$. As before, let (X_n) denote the skeleton chain obtained from X. Define $\tau_a = \inf\{n \geq 0 \,:\, X_n \geq a\}$, then for each $m \in \mathbb{N}$,

$$\begin{aligned}
Q_x(\tau_a > m) &= Q_x(X_1 < a, \ldots, X_m < a) \\
&= Q_0(X_1 + e^{-\lambda}x < a, \ldots, X_m + e^{-\lambda m}x < a) \\
&\leq Q_0(X_1 < a, \ldots, X_m < a) \\
&= Q_0(W_1 < a, \ldots, e^{-\lambda}X_{m-1} + W_m < a) \\
&\leq P(W_1 < a, \ldots, W_m < a) = [P(W_\lambda < a)]^m \in [0, 1)\,.
\end{aligned}$$

The last assertion holds since the support of any non-degenerate infinitely divisible random variable is unbounded (Sato (1999), Theorem 24.3). From this, it follows that

$$Q_x(\tau_a < \infty) \geq \lim_{m\to\infty} (1 - [P(W_\lambda < a)]^m) = 1\,.$$

It is easy to see that $\{\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty\} \subseteq \{\sigma_a < \infty\}$ (here θ denotes the shift-operator as defined in Section 1.4). Hence,

$$\begin{aligned}
Q_x(\sigma_a < \infty) &\geq Q_x(\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty) = E_x\{E_x(\mathbf{1}_{\{\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty\}} \mid \mathcal{F}_{\tau_a})\} \\
&= E_x\{\mathbf{1}_{\{\tau_a < \infty\}} E_x(\mathbf{1}_{\{\sigma_a \circ \theta_{\tau_a} < \infty\}} \mid \mathcal{F}_{\tau_a})\} = E_x\{E_{X(\tau_a)}\mathbf{1}_{\{\sigma_a < \infty\}}\} = 1\,.
\end{aligned}$$

The second inequality holds since $\{\tau_a + \sigma_a \circ \theta_{\tau_a} < \infty\} = \{\tau_a < \infty\} \cap \{\sigma_a \circ \theta_{\tau_a} < \infty\}$. The third equality follows from the strong Markov property, as formulated in (1.13). The last equality follows from the case $x \geq a$.

Hence, for all $x \in E$, we have proved that $Q_x(\sigma_a < \infty) = 1$. Thus X is Harris-recurrent.

By Theorem 1.16, the invariant measure of a Lévy driven OU-process is a probability measure, which shows that X is *positive* Harris-recurrent. $\square$

Remark 1.25 The β-mixing property of general (multi-dimensional) OU-processes is also treated in Masuda (2004), Section 4. There it is assumed that the OU-process is strictly stationary, and moreover that $\int |x|^\alpha \pi(dx) < \infty$, for some $\alpha > 0$. The latter assumption is stronger than our assumption (A), but also yields the stronger conclusion that $\beta_X(t) = O(e^{-at})$, as $t \to \infty$, for some $a > 0$ (i.e. the process X is *geometrically ergodic*). It seems hard to extend the argument in Masuda (2004) under assumption (A).

1.6 Proof of φ-irreducibility of an AR(1)-process

Let $X = (X_n, n \geq 0)$ be the first order autoregressive process defined by

$$X_n = \alpha X_{n-1} + \eta_n, \quad \alpha \in (0, 1) \tag{1.20}$$

in which we assume that $(\eta_n, n \geq 1)$ is a sequence of independent random variables with common distribution P^η. Denote the state-space of X by $(E, \mathcal{E})$ $(E \subseteq \mathbb{R})$. The chain X is φ-irreducible if there exists a non-zero σ-finite measure φ, such that for all $B \in \mathcal{E}$ with $\varphi(B) > 0$, $\sum_{n=1}^{\infty} P_n(x, B) > 0$, for all $x \in E$. Here P_n denotes the n-step transition kernel of the chain. Let Q_x denote the law of the process X, if started at $x \in E$. The following definitions are taken from Meyn and Tweedie (1993), Chapter 6. A kernel T is a *continuous component* of P_1 if

(i) T possesses the *strong Feller property*, i.e. the map $f \mapsto \int f(y)T(\cdot, dy)$ maps bounded measurable functions to bounded continuous functions.

(ii) For all $x \in E$ and $A \in \mathcal{E}$, the measure $T(x, \cdot)$ satisfies $P_1(x, A) \geq T(x, A)$.

T is called *non-trivial* if $T(x, E) > 0$ for all $x \in E$. We say x^* is a *reachable point* for the chain (X_n) if for any initial state $x_0 \in E$ there exists an $N \geq 1$ such that $P_N(x_0, \mathcal{O}_{x^*}) > 0$. Here, $\mathcal{O}_{x^*}$ is an arbitrary neighborhood of x^*.

Theorem 1.26 *If P^η is non-singular with respect to Lebesgue measure, then the chain (X_n) defined by (1.20) is φ-irreducible.*

Proof The proof is split into three parts:

Step (i): The kernel P_1 has a non-trivial continuous component T. The proof is analogous to the proof of Proposition 6.3.3. in Meyn and Tweedie (1993). Let f be a positive, bounded $\mathcal{E}$-measurable function and $x_0 \in E$, then

$$P_1 f(x_0) = E[f(\alpha x_0 + \eta_1)] = \int_E f(\alpha x_0 + w) P^\eta(dw) \geq \int_E f(\alpha x_0 + w)\gamma(w)dw.$$

Here γ denotes the density of the absolutely continuous component of P^η. Define $Tf(x_0)$ by the right-hand side of the preceding display. We see that T is a non-trivial continuous component of P_1 if and only if for each $f \in b\mathcal{E}$ the function $x \mapsto Tf(x)$ is bounded and continuous on E. The function Tf is bounded since f is bounded and γ is a sub-probability density. Let $M = \sup_{y \in E} |f(y)|$. To prove continuity, we rewrite Tf as

$$Tf(x) = \int_E f(y)\gamma(y - \alpha x)dy\,.$$

Since γ is integrable, the mapping $x \mapsto \gamma(\cdot - \alpha x)$ from $\mathbb{R}$ to $L^1(\mathbb{R})$ is uniformly continuous (Theorem 9.5 in Rudin (1987)). Hence, for a given $\varepsilon > 0$ there exists a $\delta > 0$ such that if $u, v \in E$ with $|u - v| < \delta$, then $\int_E |\gamma(y - \alpha u) - \gamma(y - \alpha v)|dy < \varepsilon/M$. Now

$$|Tf(u) - Tf(v)| \leq \sup_{y \in E} |f(y)| \int_E |\gamma(y - \alpha u) - \gamma(y - \alpha v)|dy < M\varepsilon/M = \varepsilon\,,$$

which shows the asserted continuity.

Step (ii): The chain (X_n) has a reachable point. Let x_0 be the initial state of the chain (X_n). Let $w^* \in \operatorname{supp}(\eta)$ and define $x_n = \alpha^n x_0 + \sum_{j=0}^{n-1} \alpha^j w^*$. Since $\alpha \in (0, 1)$ $x_n \to \sum_{j=0}^{\infty} \alpha^j w^* =: x^*$, as $n \to \infty$. We show x^* is a reachable state for the chain (X_n).

Let $\delta > 0$. Since $x_n \to x^*$, there exists an $N \in \mathbb{N}$ such that $|x_N - x^*| < \delta/2$. Now,

$$p := P_N(x_0, (x^* - \delta, x^* + \delta)) = Q_{x_0}(|X_N - x^*| < \delta) \geq Q_{x_0}(|X_N - x_N| < \delta/2)\,.$$

Choose $\varepsilon > 0$ such that $\varepsilon \sum_{k=0}^{N-1} \alpha^k \leq \delta/2$, then

$$p \geq Q_{x_0}\left(|X_N - x_N| \leq \varepsilon \sum_{k=0}^{N-1} \alpha^k \right) .$$

If $|\eta_{n-k} - w^*| \leq \varepsilon$ for all $k = 0, \ldots, N-1$, then $|X_N - x_N| \leq \varepsilon \sum_{k=0}^{N-1} \alpha^k$ and hence

$$p \geq P(|\eta_{n-k} - w^*| \leq \varepsilon\,,\; k = 0, \ldots, N-1) = [P(|\eta_1 - w^*| \leq \varepsilon]]^N\,,$$

where the last equality holds since the sequence $(\eta_k, k \geq 1)$ is i.i.d. The last probability in the preceding display is at least $[\int_{\{x:\,|x-w^*|\leq\varepsilon\}} \gamma_w(dx)]^N$, which is strictly positive, since $w^* \in \mathrm{supp}(W_\lambda)$.

Step (iii): The chain (X_n) is φ-irreducible. The proof is similar to the proof of Proposition 6.2.1 in Meyn and Tweedie (1993). Define for $x \in E$ and $A \in \mathcal{E}$, $U(x, A) = \sum_{n=1}^{\infty} P_n(x, A)$. It is easily seen that

$$U(x, A) \geq \int U(x, dy) P_1(y, A)\,. \tag{1.21}$$

Now we use the continuous component T and the reachable point x^* to define the measure φ on $(E, \mathcal{E})$ by $\varphi(\cdot) = T(x^*, \cdot)$. Note that φ is non-zero, since T is non-trivial.

Let $A \in \mathcal{E}$ be such that $\varphi(A) = T(x^*, A) > 0$. Since T is a continuous component, the kernel $T\mathbf{1}_A(\cdot)$ is lower semi-continuous for every $A \in \mathcal{E}$ (Meyn and Tweedie (1993), Proposition 6.1.1*(i)*). This implies that there exists an open set O, containing x^*, such that $T(y, A) > 0$, for all $y \in O$. Now, using (1.21), we get for all $x \in E$,

$$U(x, A) \geq \int_O U(x, dy) P_1(y, A) \geq \int_O U(x, dy) T(y, A) > 0,$$

which shows that (X_n) is φ-irreducible. $\square$

Remark 1.27 If the chain is strong Feller, then we can take $T = P_1$ in the proof of the previous theorem, which simplifies the proof (step (i) becomes superfluous). In general, a discretely observed OU-process need not be strong Feller, as the following illustrates.

Let $f(x) = \mathbf{1}_{(0,\infty)}(x)$, then $f \in b\mathcal{E}$ and

$$\begin{aligned} P_t f(x) &= P_t(0, (-e^{-\lambda t}x, \infty)) = P(X^{(0)}(t) \in (-e^{-\lambda t}x, \infty)) \\ &= \begin{cases} 1 & \text{if} \quad x > 0 \\ 1 - P(X^{(0)}(t) = 0) & \text{if} \quad x = 0 \end{cases}, \end{aligned}$$

where $(X^{(0)}(t), t \geq 0)$ denotes the OU-process started at zero. If Z is an increasing compound Poisson process with Lévy measure ρ, then $\rho(0, \infty) < \infty$. Hence,

$$P(X^{(0)}(t) = 0) = P(\text{no jumps of } Z \text{ in } [0, \lambda t]) = e^{-\rho(0,\infty)\lambda t} > 0,$$

which implies that $x \mapsto P_t f(x)$ is not continuous at zero.

1.7 Some additional proofs and results

Proof of Proposition 1.20 The solution of the OU-equation is given in (1.9). If we discretize the expression for this solution we obtain

$$X_n = e^{-\lambda} X_{n-1} + \int_0^1 e^{\lambda(u-1)} dZ(\lambda(u+n-1)) \,, \quad n \geq 1 \,.$$

Since Z has stationary and independent increments we can write

$$X_n = e^{-\lambda} X_{n-1} + W_n(\lambda) \,, \tag{1.22}$$

where $(W_n(\lambda))_n$ is an i.i.d. sequence of random variables, each distributed as W_λ.

Next, we show that the distribution of $(\tilde{X}(t))$, defined by

$$\tilde{X}(t) := \int_0^t e^{-\lambda(t-s)} dZ(\lambda s) \tag{1.23}$$

is infinitely divisible for each $t \geq 0$. Since $W_\lambda \stackrel{d}{=} \tilde{X}(1)$ we then obtain infinite divisibility for the noise variables. Note that $(\tilde{X}_t)$ is simply the OU-process with initial condition $X(0) = 0$,

By Equation (1.10),

$$E e^{iz\tilde{X}(t)} = \exp\left(\lambda \int_0^t g(e^{-\lambda(t-u)} z) du\right) ,$$

in which g denotes the cumulant of Z_1. Since Z has Lévy measure ρ we have $g(u) = \int_0^\infty (e^{iux} - 1)\rho(dx)$ and hence the exponent on the right-hand-side of the preceding display equals

$$\begin{aligned}
\lambda \int_0^t \int_0^\infty \left(e^{ie^{-\lambda(t-u)} zx} - 1\right) \rho(dx) du &= \lambda \int_0^\infty \int_0^t \left(e^{ie^{-\lambda(t-u)} zx} - 1\right) du \rho(dx) \\
&= \int_0^\infty \int_{e^{-\lambda t} x}^x (e^{iwz} - 1) w^{-1} dw \rho(dx) = \int_0^\infty (e^{iwz} - 1) w^{-1} dw \int_w^{e^{\lambda t} w} \rho(dx) \\
&= \int_0^\infty (e^{iwz} - 1) \kappa_t(dw).
\end{aligned}$$

Here

$$\kappa_t(B) = \int_B w^{-1} \rho(w, e^{\lambda t} w] dw \,, \quad B \in \mathcal{E} \,.$$

If we let $\kappa := \kappa_1$, the Lévy measure has the form as given in (1.17).

It remains to be shown that κ_t satisfies $\int_0^\infty (1 \wedge x)\kappa_t(dx) < \infty$ for each $t > 0$. This follows from

$$\int_0^1 x \kappa_t(dx) = \int_0^1 \rho(x, e^{\lambda t} x] dx \leq \int_0^{e^{\lambda t}} y \rho(dy) < \infty \,,$$

and

$$\begin{aligned}\int_1^\infty \kappa_t(dx) &= \kappa_t(1,\infty) = \int_1^\infty \int_{(1\vee e^{-\lambda t}y)}^y \frac{1}{w} dw\rho(dy) \\ &= \int_1^\infty \log\left(\frac{y}{(1\vee e^{-\lambda t}y)}\right)\rho(dy) < \infty \\ &= \int_1^{e^{\lambda t}} \log y\rho(dy) + \int_{e^{\lambda t}}^\infty \lambda t\rho(dy) < \infty.\end{aligned}$$

□

Lemma 1.28 *Under* (A),

$$I := \int_0^1 \frac{dz}{z}\exp\left(-\int_z^1 \frac{\lambda_\rho(y)}{y}dy\right) = +\infty\,, \tag{1.24}$$

where

$$\lambda_\rho(y) = \int_0^\infty (1-e^{-yx})\rho(dx)\,. \tag{1.25}$$

Proof Let $y \in (0,1)$. Since $1-e^{-u} \le \min(u,1)$ for $u>0$ we obtain

$$\begin{aligned}\lambda_\rho(y) &= \int_0^1 \ldots + \int_1^{1/\sqrt{y}} \ldots + \int_{1/\sqrt{y}}^\infty (1-e^{-yx})\rho(dx) \\ &\le y\int_0^1 x\rho(dx) + \int_1^{1/\sqrt{y}} \frac{y}{\sqrt{y}}\rho(dx) + \int_{1/\sqrt{y}}^\infty \frac{1-e^{-yx}}{\log x}\log x\rho(dx) \\ &\le c_1 y + c_2\sqrt{y} - \frac{2}{\log y}\int_{1/\sqrt{y}}^\infty \log x\rho(dx)\,,\end{aligned}$$

where $c_1 = \int_0^1 x\rho(dx)$ and $c_2 = \rho(1,\infty)$.

Choose $\alpha \in (0,1)$ such that $c_3 := 2\int_{1/\sqrt{\alpha}}^\infty \log x\rho(dx) < 1$, which is possible by (A). Since $y \mapsto \int_{1/\sqrt{y}}^\infty \log x\rho(dx)$ is increasing on $(0,1)$, we have

$$\lambda_\rho(y) \le c_1 y + c_2\sqrt{y} - c_3/\log y\,, \quad \text{if } y \in (0,\alpha)\,.$$

For $y \in (\alpha,1)$, we have the simple estimate $\lambda_\rho(y) \le c_1 y + c_2$. If $z \in (0,\alpha)$, then

$$\begin{aligned}\int_z^1 \frac{\lambda_\rho(y)}{y}dy &= \int_z^\alpha \frac{\lambda_\rho(y)}{y}dy + \int_\alpha^1 \frac{\lambda_\rho(y)}{y}dy \\ &\le c_1(\alpha - z) + 2c_2(\sqrt{\alpha}-\sqrt{z}) - c_3\int_z^\alpha \frac{1}{y\log y}dy + c_1(1-\alpha) - c_2\log\alpha \\ &= K_\alpha - c_1 z - 2c_2\sqrt{z} + c_3\log(-\log z)\,,\end{aligned}$$

where

$$K_\alpha = c_1 + c_2(2\sqrt{\alpha} - \log(\alpha)) - c_3\log(-\log\alpha) \in \mathbb{R}\,.$$

Using this inequality we get

$$I \geq \int_0^\alpha \frac{dz}{z} \exp\left(-\int_z^1 \frac{\lambda_\rho(y)}{y} dy \right) \geq e^{-K_\alpha} \int_0^\alpha e^{c_1 z + 2c_2\sqrt{z}} (-\log z)^{-c_3} \frac{dz}{z} \, .$$

The last integral exceeds

$$\int_0^\alpha \frac{1}{z(-\log z)^{c_3}} dz = \int_{-\log \alpha}^\infty \frac{1}{u^{c_3}} du = \infty \, ,$$

since α was chosen such that $c_3 < 1$. $\square$

Chapter 2

Nonparametric estimation for OU-processes driven by a subordinator

In Section 1.4 we discussed stationary OU-processes that are driven by a subordinator. In this chapter we turn attention to the statistical estimation for this type of processes in case we have discrete-time observations. By Theorem 1.16, the stationary distribution of a OU-process is parametrized by a decreasing function k on the positive half-line (the *canonical function*). Hence, together with the intensity parameter λ, the couple (k, λ) parametrizes the model. Most of the work in this chapter is concerned with estimating k nonparametrically.

In the next section we present our estimation framework and in Section 2.2 we demonstrate that this leads to consistent estimators. A sieved estimator is defined in Section 2.3, in which we also show that it can be computed numerically by a support-reduction-algorithm. In Section 2.4 we show that if we assume that the canonical function is convex or completely monotone, in addition to decreasing, all obtained results remain valid. We then move on to some examples and applications in Sections 2.5 and 2.6. A simple method to obtain the density corresponding to a canonical function is outlined in Section 2.7. The method of Figueroa-López and Houdré (2004) to estimate a Lévy density nonparametrically is briefly discussed in Section 2.9. We also explain how their method can be used to estimate k. In Section 2.8 we give a simple consistent estimator for λ. The final section contains some technical results with their proofs.

2.1 Definition of a cumulant M-estimator

Recall, that for a given $\lambda > 0$ a subordinator Z, the stationary process $X = (X(t), t \geq 0)$ defined as the solution to

$$dX(t) = -\lambda X(t)dt + dZ(\lambda t), \quad t \geq 0 \tag{2.1}$$

is called the Ornstein-Uhlenbeck process (OU-process), driven by Z. We assume throughout that the Lévy measure ρ of Z satisfies $\int_1^\infty \log x \rho(dx) < \infty$. Whence, by Theorem 1.16, the process X exists. We have the following setup:

Data It is assumed that X is observed at equally spaced times $t_n^k = k\Delta_n$ ($k = 0, \ldots, n-1$), so that the data consist of the time series $\{X_{k\Delta_n}\}_{k=0}^{n-1}$.

Statistical model The process X is parametrized by the couple (λ, k), where λ is the intensity parameter, and k the canonical function. The latter function is right-continuous, decreasing and satisfies the integrability condition $\int_0^\infty (x \wedge 1) x^{-1} k(x) dx < \infty$. The invariant probability distribution of X, denoted by π, has characteristic function ψ which can be expressed explicitly in terms of k in the following way

$$\psi_k(t) = \int e^{itx} \pi(dx) = \exp\left(\int_0^\infty (e^{itx} - 1) \frac{k(x)}{x} dx \right), \quad t \in \mathbb{R}. \tag{2.2}$$

The function k determines the process Z by the relation $k(x) = \rho(x, \infty)$ ($x > 0$). The special scaling in the defining equation for X, (2.1), allows for a separate estimation of λ and k. The former estimation problem is deferred to Section 2.8.

In this section we outline a new estimation method for the canonical function k. As pointed out in the introductory chapter, maximum likelihood techniques are hampered by the lack of a closed form expression for the density of π in terms of the "natural" parameter k (the density of a nondegenerate self-decomposable distribution always exists, see Sato (1999), Theorem 27.13). Even if such an expression would exist, we would in fact need more, namely, the transition densities of X, in terms of k. This motivates our choice for another type of M-estimator.

This estimator is constructed by firstly defining a preliminary estimator $\tilde{\psi}_n$ for ψ_0, which denotes the "true" underlying characteristic function. Any reference to the true distribution will be denoted by a subscript 0. For example, F_0 denotes the true underlying distribution function of $X(1)$ and k_0 denotes the true canonical function. Formally, in view of (2.2), we should write ψ_{k_0} in stead of ψ_0, but this should not cause confusion. In what follows, we choose $\tilde{\psi}_n$ such that

$$\text{for each } n, \ \tilde{\psi}_n \text{ is a ch.f. and } \forall\, t \in \mathbb{R} \quad \tilde{\psi}_n(t) \to \psi_0(t), \quad n \to \infty, \tag{2.3}$$

where the convergence is either almost surely, or in probability. Hence, $\tilde{\psi}_n$ estimates ψ_0 consistently pointwisely. In Section 2.2 we will see that this consistency carries over to the estimator for k_0 which we will define below. A natural preliminary estimator is given by the *empirical characteristic function*, which satisfies condition (2.3) under mixing conditions.

Given any preliminary estimator $\tilde{\psi}_n$ for ψ_0, a first idea to construct an estimator for k_0 would be to minimize some distance between ψ_k (as given in (2.2)) and $\tilde{\psi}_n$ over all canonical functions $k \in K$. Here K is defined as follows

$$K := \left\{ k \in \mathcal{L}^1(\mu) \ : \ k(x) \geq 0, \ k \text{ is decreasing and right-continuous} \right\},$$

for μ the measure defined by

$$\mu(dx) = \frac{1 \wedge x}{x} dx, \quad x \in (0, \infty)$$

and $\mathcal{L}^1(\mu)$ the space of μ-integrable functions on $(0, \infty)$ endowed with the semi-norm $\|\cdot\|_\mu = \int |\cdot| d\mu$. Note that the definition of the measure μ precisely suits the integrability condition on k, which can now be formulated by $\|k\|_\mu < \infty$. Hence, $K \subseteq \mathcal{L}^1(\mu)$ is the convex cone that contains precisely the canonical functions of all non-degenerate self-decomposable distributions on $\mathbb{R}_+$ and the degenerate distribution at 0.

As an example for constructing an estimator, we can take a weighted L^2 distance and define an estimator by

$$\hat{k}_n = \operatorname*{argmin}_{k \in K} \int |\psi_k(t) - \tilde{\psi}_n(t)|^2 w(t) dt,$$

for w a positive integrable compactly supported weight-function. Apart from the issue whether this estimator is well defined, one disadvantage of this estimating method is that the objective function is non-convex in k (convexity being desirable from a computational point of view). This problem can be avoided by comparing cumulants. The *cumulant* function is the distinguished logarithm of a characteristic function (for details on its definition we refer to the discussion just before the start of Section 1.1.1). If we assume that n is large enough such that $\tilde{\psi}_n$ has no zeroes on the support of w, then it admits a distinguished logarithm there and we can define an estimator for $\hat{k}_n$ by

$$\hat{k}_n = \operatorname*{argmin}_{k \in K} \int |\log \psi_k(t) - \log \tilde{\psi}_n(t)|^2 w(t) dt.$$

We call this estimator a *cumulant-M-estimator* (CME). For this criterion function we have the desired convexity (in k) and we will see in Section 2.3 that it can actually be computed. First, we make the presented approach more precise. For this, we introduce some notation.

Let G denote the set of cumulants corresponding to K. i.e.

$$G := \left\{ g : \mathbb{R} \to \mathbb{C} \mid g(t) = \int_0^\infty (e^{itx} - 1) \frac{k(x)}{x} dx, \text{ for some } k \in K \right\}.$$

The cumulant corresponding to a particular canonical function k_1, or characteristic function ψ_1 is denoted by g_1. In order to switch easily between canonical functions and cumulants, we define the mapping $L : K \to G$ by

$$[L(k)](t) = \int_0^\infty (e^{itx} - 1) \frac{k(x)}{x} dx, \quad t \in \mathbb{R}.$$

Let w be an integrable weight-function with compact support, denoted by S. Assume w is symmetric around the origin and w is strictly positive on a neighborhood of the origin. Lemma 2.26 in the appendix implies that k is determined by the values of $g = L(k)$ restricted to the set S.

Define the space of square integrable functions with respect to $w(t)dt$ by

$$L^2(w) := \left\{ f : \mathbb{R} \to \mathbb{C} \;\middle|\; f \text{ is measurable and } \int |f(t)|^2 w(t) dt < \infty \right\},$$

where we identify functions which are almost everywhere equal with respect to $w(t)dt$. We define an inner-product $\langle \cdot, \cdot \rangle_w$ on $L^2(w)$ by

$$\langle f, g \rangle_w = \Re \int f(t) \overline{g(t)} w(t) dt \,,$$

where the bar over g denotes complex conjugation and $\Re$ the operation of taking the real part of an element of $\mathbb{C}$. For $g \in L^2(w)$ define a norm by $\|g\|_w = \sqrt{\langle g, g\rangle_w}$. The space $(L^2(w), \langle\cdot,\cdot\rangle_w)$ is a Hilbert-space. Since elements of G are continuous and w is compactly supported, $G \subseteq L^2(w)$.

We fix the weight-function w, and assume the following:

Assumption on the sample For the given weight-function w, $\tilde{\psi}_n$ allows for a well-defined cumulant $\tilde{g}_n = \log \tilde{\psi}_n$ on S.

This assumption is satisfied with probability tending to 1 as $n \to \infty$ and enables us to define the mapping $\Gamma_n : G \to [0, \infty)$ by $\Gamma_n(g) = \|g - \tilde{g}_n\|_w^2$. If we write $\Gamma_n L$ for $\Gamma_n \circ L$, then the *cumulant M-estimator* $\hat{k}_n$ is defined as the minimizer of $\Gamma_n L$, where

$$[\Gamma_n L](k) := \|L(k) - \tilde{g}_n\|_w^2 = \int |[Lk](t) - \tilde{g}_n(t)|^2 w(t)dt, \quad k \in K,$$

over an appropriate subset of K, such that the estimator is well-defined. To find precise conditions, we will go along the following two steps:

(i) Use the fact from Hilbert-space theory that every non-empty, closed, convex set in $L^2(w)$ contains a unique element of smallest norm. Since Γ_n is a squared norm, it suffices to specify a closed, convex subset G' of G to obtain a unique minimizer of Γ_n over G'.

(ii) Show that L admits an inverse on G' so that $\hat{k}_n = L^{-1}(\hat{g}_n)$ exists. Next show that $\hat{k}_n$ minimizes Γ_n over $L^{-1}(G')$.

Ideally, we would like to take the whole set G to minimize over. The following remark shows why this is not appropriate in our context.

Remark 2.1 The set

$$G' := \Big\{g : \mathbb{R} \to \mathbb{C} \ : \ g(t) = \beta_0 it + \int_0^\infty \frac{e^{itx} - 1}{x} k(x)dx, \ \beta_0 \geq 0, k \in K\Big\},$$

is closed under uniform convergence on compact sets containing the origin. To see this, let S be such a compact set. If $\{g_n\}_n \in G'$ is such that $\sup_{t\in S} |g_n(t) - g(t)| \to 0$ for some g, then $\sup_{t\in S} |\psi_n(t) - \psi(t)| \to 0$ and then (by the same argument as in the proof of Lévy's continuity theorem) the random variables corresponding to $\{\psi_n\}$ are uniformly tight. Denote these r.v. by $\{X_n\}$. By Prohorov's theorem, there exists a subsequence n_l such that X_{n_l} converges weakly to a random variable X^*. Since X_n is a positive SD random variable, and the class of positive SD random variables is closed under weak convergence, X^* is positive SD. Let g^* be the cumulant of X^*. Then $g^* \in G'$ and $\sup_{t\in S} |g_{n_l}(t) - g^*(t)| \to 0$. Together with the continuity of g and g^* on S, this implies $g^* = g$ on S. Hence $g = g^* \in G'$.

However, the set G is not closed under uniform convergence on compact sets containing the origin. Let S again be such a set and define a sequence $\{k_n\}_{n\geq 1} \in K$ by $k_n(x) = n\mathbf{1}_{[0,1/n)}(x)$. Then for each $t \in \mathbb{R}$,

$$g_n(t) = [L(k_n)](t) = n \int_0^{1/n} \frac{e^{itx} - 1}{x} dx \to it\,, \quad \text{as } n \to \infty\,.$$

Let $g(t) = it$, then since each g_n and g are uniformly continuous on the compact set S, we have $\sup_{t\in S} |g_n(t) - g(t)| \to 0$. However $g \notin G$, since g can only correspond to a point mass at one. Returning to the set G', we see that this example is the canonical example that can preclude G being closed.

In view of dominated convergence, the above counter example also shows why the set G is not closed in $L^2(w)$. To obtain an appropriate *closed* subset of G, we define a set of envelope functions in K. Pick for each $R > 0$ a function $k_R \in K$ such that $\|k_R\|_\mu \leq R$ (for example $k_R(x) = R/(4\sqrt{x})$). The collection $\{k_R,\, R > 0\}$ defines a set of envelope functions. Now let

$$K_R := \{k \in K \mid k(x) \leq k_R(x) \text{ for } x \in (0, \infty)\}.$$

and put $G_R = L(K_R)$, the image of K_R under L.

Lemma 2.2 *Let $R > 0$. Then,*

(i) K_R is a compact, convex subset of $\mathcal{L}^1(\mu)$.

(ii) G_R is a compact, convex subset of $L^2(w)$.

Proof (i): Convexity of K_R is obvious.

Let $\{k_n\}$ be a sequence in K_R. Since each k_n is bounded on all strictly positive rational points, we can use a diagonalization argument to extract a subsequence n_j from n such that the sequence k_{n_j} converges to some function $\bar{k}$ on all strictly positive rationals. For $x \in (0, \infty)$ define

$$\tilde{k}(x) = \sup\{\bar{k}(q),\, x < q,\, q \in \mathbb{Q}\}\,.$$

This function is (by its definition) decreasing and right-continuous and satisfies $\tilde{k} \leq k_R$ on $(0, \infty)$. Thus $\tilde{k} \in K_R$. Furthermore, k_{n_j} converges pointwisely to $\tilde{k}$ at all continuity points of $\tilde{k}$. Since the number of discontinuity points of $\tilde{k}$ is at most countable, k_{n_j} converges to $\tilde{k}$ a.e. on $(0, \infty)$. Moreover, since $k_{n_j} \leq k_R$ on $(0, \infty)$ and $k_R \in \mathcal{L}^1(\mu)$, Lebesgue's dominated convergence theorem applies: $\|k_{n_j} - \tilde{k}\|_\mu \to 0$, as $n_j \to \infty$. Hence, K_R is sequentially compact.

(ii): Convexity of G_R follows from convexity of K_R. By the next lemma, $L : K \to G$ is continuous. Then G_R is compact since it is the image of the compact set K_R under the continuous mapping L. □

Lemma 2.3 *The mapping $L : (K, \|\cdot\|_\mu) \to (G, \|\cdot\|_w)$ is continuous, onto and one-to-one.*

Proof Let $\{k_n\}$ be a sequence in K converging to $k_0 \in K$, i.e. $\|k_n - k_0\|_\mu \to 0$, as $n \to \infty$.

For $t \in S$,

$$\begin{aligned} |L(k_n)(t) - L(k_0)(t)| &= \left|\int_0^\infty (e^{itx} - 1)\frac{k_n(x) - k_0(x)}{x} dx\right| \\ &\leq |t| \int_0^1 |k_n(x) - k_0(x)| dx + 2\int_1^\infty x^{-1}|k_n(x) - k_0(x)| dx \\ &\leq \max\{|t|, 2\}\|k_n - k_0\|_\mu\,, \end{aligned}$$

where we use the inequality $|e^{ix} - 1| \leq \min\{|x|, 2\}$. Thus $L(k_n) \to L(k_0)$ uniformly on S which implies $\|L(k_n) - L(k_0)\|_w \to 0$ $(n \to \infty)$. Hence, L is continuous.

The surjectivity is trivial by the definition of G. If $g_1, g_2 \in G$ and $\|g_1 - g_2\|_w = 0$, then (by continuity of elements in G), $g_1 = g_2$ on S. Then also $\psi_1 := e^{g_1} = e^{g_2} =: \psi_2$ on S. By Lemma 2.26, $\psi_1 \equiv \psi_2$, which implies $k_1 \equiv k_2$. □

Corollary 2.4 *The inverse operator of L, $L^{-1} : G_R \to K_R$ is continuous.*

Proof It is a standard result from topology that a continuous, bijective mapping, defined on a compact set, has a continuous inverse (see e.g. Corollary 9.12 in Jameson (1974)). Whence, the result is a direct consequence of Lemma 2.3. □

Since we want to define our objective function in terms of the canonical function, one last step is necessary. Since Γ_n has a unique minimizer over G_R and to each G_R belongs a unique member of K_R, there exists a unique minimizer of $\Gamma_n L$ over K_R. More precisely:

Lemma 2.5 *Let $\hat{g}_n = \mathrm{argmin}_{g \in G_R} \Gamma_n(g)$ (which is by now known to exist and to be unique). Then $\hat{k}_n = \mathrm{argmin}_{k \in K_R}[\Gamma_n L](k)$ exists. Moreover, $\hat{k}_n = L^{-1}(\hat{g}_n)$ and $\hat{k}_n$ is unique.*

Proof Since $L : K_R \to G_R$ is onto and one-to-one, to each $g \in G_R$ there corresponds a unique $k \in K_R$ such that $L(k) = g$. Thus

$$\beta := \min_{g \in G_R} \Gamma_n(g) = \min_{k \in K_R} [\Gamma_n L](k) .$$

Now define $\hat{k}_n = L^{-1}(\hat{g}_n)$ and choose an arbitrary $k \in K_R$ (but $k \neq \hat{k}_n$). Then $\hat{k}_n \in K_R$ and

$$[\Gamma_n L](\hat{k}_n) = \Gamma_n(\hat{g}_n) = \beta < [\Gamma_n L](k) ,$$

which shows that $\hat{k}_n$ is the unique minimizer of $\Gamma_n L$ over K_R. □

We summarize the obtained result.

Theorem 2.6 *Let w be an integrable, non-negative, symmetric, compactly supported function that is strictly positive in a neighborhood of the origin. Suppose n is large enough such that $\tilde{g}_n$ is well defined on the support of w. Then, for any $R > 0$,*

$$\hat{k}_n = \operatorname*{argmin}_{k \in K_R} [\Gamma_n L](k)$$

exists uniquely.

In Section 2.3 we will define a *sieved* estimator. The following result turns out to be useful to prove the existence of such an estimator.

Proposition 2.7 *Assume the conditions of Theorem 2.6 hold. Denote the support of the measure ξ by $\mathrm{supp}(\xi)$. If we define, for $\varepsilon > 0$*

$$K_\varepsilon = \Big\{ k \in K \mid k = \int \mathbf{1}_{[0,\theta)} \xi(d\theta),\ \mathrm{supp}(\xi) \subseteq [\varepsilon, \infty) \Big\}, \tag{2.4}$$

then the estimator

$$k_n^{(\varepsilon)} = \operatorname*{argmin}_{k \in K_\varepsilon} [\Gamma_n L](k)$$

exists uniquely.

Proof The minimization can be restricted to the set $K_\varepsilon^M = \{k \in K_\varepsilon \mid k(0) \le M\}$, for a positive constant M, to be constructed below. To see this, first note that if $k \in K_\varepsilon$

$$\begin{aligned}
\int_0^\infty \frac{1-\cos(tx)}{x} k(x)dx &= \int_0^\infty \frac{1-\cos(tx)}{x} \int_{[\varepsilon,\infty)} \mathbf{1}_{[0,\theta)}(x)\xi(d\theta)dx \\
&= \int_0^\infty \frac{1-\cos(tx)}{x} \xi((x,\infty)\cap[\varepsilon,\infty))dx \\
&\ge \xi([\varepsilon,\infty)) \int_0^\varepsilon \frac{1-\cos(tx)}{x} dx \\
&= k(0) \int_0^\varepsilon \frac{1-\cos(tx)}{x} dx.
\end{aligned}$$

This implies

$$|[Lk](t)|^2 \ge (\Re[Lk](t))^2 = \left(\int_0^\infty \frac{1-\cos(tx)}{x} k(x)dx\right)^2 \ge k(0)^2 \left(\int_0^\varepsilon \frac{1-\cos(tx)}{x} dx\right).$$

Therefore, for every $k \in K_\varepsilon$ there exists a positive constant C such that $(\Gamma_n L)(k) \ge Ck(0)^2$. Without loss of generality we can assume $\varepsilon < 1$. Let $\tilde{k}(\cdot) = \mathbf{1}_{[0,1]}(\cdot)$, then $\tilde{k} \in K_\varepsilon$ and $(\Gamma_n L)(\hat{k}_n^{(\varepsilon)}) \le (\Gamma_n L)(\tilde{k}) = D < \infty$ (say). Hence

$$\hat{k}_n^{(\varepsilon)}(0) \le \left(\frac{(\Gamma_n L)(\hat{k}_n^{(\varepsilon)})}{C}\right)^{1/2} \le \sqrt{D/C} < \infty,$$

which shows that we can take $M = \sqrt{D/C}$.

For R sufficiently large, $K_\varepsilon^M \subseteq K_R$. It is easily verified that K_ε^M is a closed subset of the compact set K_R and hence compact. This shows that we can follow the proof of Theorem 2.6 to prove that there exists a unique minimizer of $\Gamma_n L$ over K_ε^M. $\square$

2.2 Consistency

In this section we show the consistency of the cumulant M-estimator, which is a consequence of imposing condition (2.3) on the preliminary estimator. We start with two results, which strengthen the pointwise convergence in (2.3) to uniform convergence. The first is on convergence in probability, whereas the second result is on almost sure convergence. In the proof we use the following characterization of convergence in probability: $Y_n \xrightarrow{\text{p}} Y$ if and only if each subsequence of (Y_n) possesses a further subsequence that converges almost surely to Y.

Lemma 2.8 *Let $(\Omega, \mathcal{U}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose that $\varphi_n(\cdot,\omega)$ $(n = 1, 2, \dots)$ and φ are characteristic functions such that for each $t \in \mathbb{R}$, $\varphi_n(t,\cdot) \xrightarrow{\text{p}} \varphi(t)$, as $n \to \infty$. Then, as $n \to \infty$*

$$\sup_{t\in K} |\varphi_n(t,\cdot) - \varphi(t)| \xrightarrow{\text{p}} 0\,, \quad \textit{for every compact set } K \subseteq \mathbb{R}\,.$$

Proof Denote the distribution functions corresponding to $\varphi_n(\cdot,\omega)$ and φ by $F_n(\cdot,\omega)$ and F. The functions $x \mapsto e^{itx}$ for $t \in K$ are uniformly bounded and equicontinuous. Therefore (by the Arzelà-Ascoli theorem), if $F_n(\cdot,\omega) \rightsquigarrow F$ for some ω along some subsequence ($\rightsquigarrow$ denotes weak convergence), then $\sup_{t\in K}|\varphi_n(t,\cdot)-\varphi(t)| \to 0$ for this ω and subsequence. It follows that it suffices to show that for every subsequence of $\{n\}$ there exists a further subsequence $\{n'\}$ and a set $A \in \mathcal{U}$ with $\mathbb{P}(A) = 1$ such that $F_{n'}(\cdot,\omega) \rightsquigarrow F$, $\forall\,\omega \in A$, along the subsequence.

By assumption: for every t there exists a subsequence $\{n\}$ such that $\varphi_n(t,\omega) \xrightarrow{\text{a.s.}} \varphi(t)$. Denote $\mathbb{Q} = \{q_1, q_2, \dots\}$. There exists a subsequence $\{n^{(1)}\}$ of $\{n\}$ and a set $A^{(1)} \in \mathcal{U}$ with $\mathbb{P}(A^{(1)}) = 1$ such that $\varphi_{n^{(1)}}(q_1,\omega) \to \varphi(q_1)$, for all $\omega \in A^{(1)}$. There exists a subsequence $\{n^{(2)}\}$ of $\{n^{(1)}\}$ and a set $A^{(2)} \in \mathcal{U}$ with $\mathbb{P}(A^{(2)}) = 1$ such that $\varphi_{n^{(2)}}(q_2,\omega) \to \varphi(q_2)$, for all $\omega \in A^{(2)}$. Proceed iteratively in this way. Consider the diagonal sequence, obtained by $n_i := n_i^{(i)}$, and set $A = \cap_{i=1}^{\infty} A^{(i)}$, then $\mathbb{P}(A) = 1$ and

$$\varphi_n(q,\omega) \to \varphi(q)\,, \quad \forall\, q \in \mathbb{Q}\,, \quad \forall \omega \in A\,. \tag{2.5}$$

For every $\delta > 0$,

$$\int_{|x|>2/\delta} F_n(dx,\omega) \le \frac{1}{2\delta}\int_{-\delta}^{\delta} |1-\varphi_n(t,\omega)|dt =: a_n(\delta,\omega)\,, \tag{2.6}$$

by a well-known inequality (see for instance Chung (2001), Chapter 6, Section 3). Furthermore, with $a(\delta) := \frac{1}{2\delta}\int_{-\delta}^{\delta}|1-\varphi(t)|dt$, by Fubini's theorem

$$\mathbb{E}|a_n(\delta,\cdot) - a(\delta)| \le \frac{1}{2\delta}\int_{-\delta}^{\delta} \mathbb{E}\big||1-\varphi_n(t)| - |1-\varphi(t)|\big|dt \to 0\,, \quad n\to\infty\,,$$

by dominated convergence and the assumed convergence in probability. Thus, for every $\delta > 0$ there exists a further subsequence $\{n\}$ and a set $B \in \mathcal{U}$ with $\mathbb{P}(B) = 1$ such that $a_n(\delta,\omega) \to a(\delta)$ for all $\omega \in B$. By a diagonalization scheme we can find a further subsequence $\{n\}$ and a set $C \in \mathcal{U}$ with $\mathbb{P}(C) = 1$ such that

$$\lim_{n\to\infty} |a_n(\delta,\omega) - a(\delta)| = 0\,, \quad \forall\,\delta \in \mathbb{Q}\cap(0,\infty)\,, \quad \forall\,\omega \in C\,.$$

Combined with (2.6) this shows that

$$\limsup_{n\to\infty} \int_{\{|x|>2/\delta\}} F_n(dx,\omega) \le a(\delta)\,, \quad \forall\,\delta \in \mathbb{Q}\cap(0,\infty)\,, \quad \forall\,\omega \in C\,,$$

taking the limsup over the subsequence. Because $a(\delta) \downarrow 0$ as $\delta \downarrow 0$ we see that $\{F_n(\cdot,\omega)\}_{n=1}^{\infty}$ is tight for all $\omega \in C$.

If G is a limit point of $F_n(\cdot,\omega)$, then by (2.5)

$$\int e^{itx}dG(x) = \lim_{n\to\infty}\int e^{itx}F_n(dx,\omega) = \int e^{itx}dF(x)\,, \quad \forall\, t \in \mathbb{Q}\,, \quad \forall\,\omega \in A\,.$$

Hence $F = G$, and it follows that $\{F_n(\cdot,\omega)\}_n$ has only one limit point, whence $F_n(\cdot,\omega) \rightsquigarrow F$, for all $\omega \in A\cap C$, along the subsequence. $\square$

Lemma 2.9 *Let $(\Omega, \mathcal{U}, \mathbb{P})$ be a probability space. For each $\omega \in \Omega$, suppose that $\varphi_n(\cdot, \omega)$ $(n = 1, 2, \ldots)$ and φ are characteristic functions such that for each $t \in \mathbb{R}$, $\varphi_n(t, \cdot) \xrightarrow{\text{a.s.}} \varphi(t)$, as $n \to \infty$. Then, as $n \to \infty$*

$$\sup_{t \in K} |\varphi_n(t, \cdot) - \varphi(t)| \xrightarrow{\text{a.s.}} 0\,, \quad \textit{for every compact set } K \subseteq \mathbb{R}\,.$$

Proof It suffices to show that there exists an $A \in \mathcal{U}$ with $\mathbb{P}(A) = 1$ such that $F_n(\cdot, \omega) \rightsquigarrow F$, for all $\omega \in A$.

With a_n and a as in the proof of the previous lemma

$$\mathbb{E} \sup_{m \geq n} |a_m(\delta, \cdot) - a(\delta)| \leq \frac{1}{2\delta} \int_{-\delta}^{\delta} \mathbb{E} \left(\sup_{m \geq n} \Big| |1 - \varphi_m(t, \cdot)| - |1 - \varphi(t)| \Big| \right) dt \to 0\,, \quad n \to \infty\,.$$

This implies that $|a_n(\delta, \cdot) - a(\delta)| \xrightarrow{\text{a.s.}} 0$ $(n \to \infty)$, for all $\delta > 0$. Combined with (2.6) we see

$$\limsup_{n \to \infty} \int_{\{|x| > 2/\delta\}} F_n(dx, \omega) \leq a(\delta)\,, \quad \forall \delta \in \mathbb{Q}\,, \quad \omega \in A_1\,,$$

for some set $A_1 \in \mathcal{U}$ with $\mathbb{P}(A_1) = 1$. Thus, for $\omega \in A_1$ the whole sequence $\{F_n(\cdot, \omega)\}$ is tight.

Let $A_2 \in \mathcal{U}$ be a set of probability one such that $\varphi_n(t, \omega) \to \varphi(t)$, $\forall t \in \mathbb{Q}$ and $\forall \omega \in A_2$. Let $\omega \in A_2$, then (as in the end of the proof of Lemma 2.8), $F_n(\cdot, \omega)$ has only F as a limit point.

Hence for all $\omega \in A := A_1 \cap A_2$, $F_n(\cdot, \omega) \rightsquigarrow F$. □

Remark 2.10 In case φ_n is the empirical characteristic function of independent random variables with common distribution F, there is a large literature on results as in Lemma 2.9. We mention the final result by Csörgö and Totik (1983): if $\lim_{n \to \infty} (\log T_n)/n = 0$, then $\sup_{|t| \leq T_n} |\varphi_n(t) - \varphi(t)| \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$, where, $\varphi(t) = \int e^{itx} F(dx)$, $(t \in \mathbb{R})$. The rate $\log T_n = o(n)$ is the best possible in general for almost sure convergence.

Now we come to the consistency result, which will be applied in Section 2.5.

Theorem 2.11 *Assume that the sequence of preliminary estimators $\tilde{\psi}_n$ satisfies (2.3) for convergence almost surely. If $k_0 \in K_R$ for some $R > 0$, then the cumulant-M-estimator is consistent. That is,*

(i)

$$\|\hat{g}_n - g_0\|_w \to 0 \quad \textit{almost surely, as } n \to \infty\,. \tag{2.7}$$

(ii)

$$\|\hat{k}_n - k_0\|_\mu \to 0 \quad \textit{almost surely, as } n \to \infty\,. \tag{2.8}$$

The same results hold in probability, if we assume (2.3) holds in probability.

Proof We first prove the statement in the case where $\tilde{\psi}_n$ converges almost surely to ψ_0.

By Lemma 2.9, $\sup_{t \in S} |\tilde{\psi}_n(t, \cdot) - \psi_0(t)| \xrightarrow{\text{a.s.}} 0$. Let $A \subseteq \Omega$ be the set of probability one on which the convergence occurs. Fix $\omega \in A$. Since ψ_0 has no zeros, there exists an $\varepsilon > 0$ such that $\inf_{t \in S} |\psi_0(t)| > 2\varepsilon$. For this ε there exists an $N = N(\varepsilon, \omega) \in \mathbb{N}$ such that $\sup_{t\in S}|\tilde{\psi}_n(t,\omega)-\psi_0(t)|<\varepsilon$ for all $n\geq N$. Hence for all $n\geq N$ and for all $t\in S$, $|\tilde{\psi}_n(t,\omega)|\geq|\psi_0(t)|-|\tilde{\psi}_n(t,\omega)-\psi_0(t)|\geq\varepsilon>0$.

For $n\geq N$ we can define the distinguished logarithm of $\tilde{\psi}_n(\omega)$ on S. Denote this function by $\tilde{g}_n(\omega)$. Theorem 7.6.3 in Chung (2001) implies that the uniform convergence of $\tilde{\psi}_n(\omega)$ to ψ_0 on S carries over to uniform convergence of $\tilde{g}_n(\omega)$ to g_0 on S. By dominated convergence $\lim_{n\to\infty}\|\tilde{g}_n(\omega)-g_0\|_w=0$.

Since $\hat{g}_n(\cdot,\omega)$ minimizes Γ_n over G_R, we have

$$\|\hat{g}_n(\cdot,\omega)-g_0\|_w\leq\|\hat{g}_n(\cdot,\omega)-\tilde{g}_n(\cdot,\omega)\|_w+\|g_0-\tilde{g}_n(\cdot,\omega)\|_w\leq 2\|g_0-\tilde{g}_n(\cdot,\omega)\|_w\to 0\,,$$

as n tends to infinity. By Corollary 2.4 this implies

$$\|\hat{k}_n(\cdot,\omega)-k_0\|_\mu=\|L^{-1}(\hat{g}_n(\cdot,\omega))-L^{-1}(g_0)\|_\mu\to 0,\quad n\to\infty\,.$$

Hence, for all ω in a set with probability one, $\lim_{n\to\infty}\|\hat{k}_n(\cdot,\omega)-k_0\|_\mu\to 0$.

Next, we prove the corresponding statement for convergence in probability. By Lemma 2.8, $Y_n:=\sup_{t\in S}|\tilde{\psi}_n(t,\cdot)-\psi_0(t)|\xrightarrow{\text{p}}0$, as $n\to\infty$. Let (n_k) be an arbitrary increasing sequence of natural numbers, then $Y_{n_k}\xrightarrow{\text{p}}0$. Then there exists a subsequence (n_m) of (n_k) such that $Y_{n_m}\xrightarrow{\text{a.s.}}0$. Now we can apply the statement of the theorem for almost sure convergence, this gives $\|\hat{k}_{n_m}-k_0\|_\mu\xrightarrow{\text{a.s.}}0$. This in turn shows that $\|\hat{k}_n-k_0\|_\mu\xrightarrow{\text{p}}0$. $\square$

Corollary 2.12 *Assume the sequence of preliminary estimators $\tilde{\psi}_n$ satisfies (2.3) for convergence almost surely. Denote the distribution function corresponding to $\hat{\psi}_n(\cdot,\omega)$ by $\hat{F}_n(\cdot,\omega)$. If $k_0\in K_R$ for some $R>0$, then for all ω in a set of probability one,*

$$\|\hat{F}_n(\cdot,\omega)-F_0(\cdot)\|_\infty\to 0,\quad n\to\infty.$$

Here $\|\cdot\|_\infty$ denotes the supremum norm. If we assume (2.3) holds for convergence in probability, then

$$\|\hat{F}_n-F_0\|_\infty\xrightarrow{\text{p}}0,\quad n\to\infty.$$

Proof First assume (2.3) for convergence almost surely. Theorem 2.11 implies $\|\hat{k}_n-k_0\|_\mu\xrightarrow{\text{a.s.}}0$, as $n\to\infty$. Fix an arbitrary ω of the set on which the convergence takes place. From the proof of Lemma 2.3, we obtain that $\hat{g}_n(\cdot,\omega)$ converges uniformly on compacta to g_0. Then $\hat{\psi}_n(\cdot,\omega)$ also converges uniformly on compacta to ψ_0. By the continuity theorem (Chung (2001), Section 6.3), $\hat{F}_n(\cdot,\omega)\rightsquigarrow F_0(\cdot)$. Since F_0 is continuous, this is equivalent to $\|\hat{F}_n(\cdot,\omega)-F_0(\cdot)\|_\infty\to 0$, as $n\to\infty$.

The statement for convergence in probability follows by arguing along subsequences, as in the proof of Theorem 2.11. $\square$

Theorem 2.11 involves only functional analytic properties of various operators and sets. To fulfill the probabilistic assumption that the sequence of preliminary estimators satisfies a law of large numbers, we can use the β-mixing result from Section 1.5.

2.3 Computing the cumulant M-estimator

For numerical purposes we will approximate the convex cone K by a finite-dimensional subset. For $N \geq 1$, let $0 < \theta_1 < \theta_2 < \ldots < \theta_N$ be a fixed set of positive numbers and set $\Theta = \{\theta_1, \ldots, \theta_N\}$. For example, we can take an equidistant grid with grid points $\theta_j = jh$ $(1 \leq j \leq N)$, where h is the mesh. Define "basis functions" by

$$u_\theta(x) := \mathbf{1}_{[0,\theta)}(x), \quad x \geq 0,$$
$$z_\theta(t) = [Lu_\theta](t) = \int_0^{\theta t} \frac{e^{iu} - 1}{u} du, \quad t \in \mathbb{R},$$

and set $\mathcal{U}_\Theta := \{u_\theta,\, \theta \in \Theta\}$. Let K_Θ be the convex cone generated by $\mathcal{U}_\Theta$, i.e.

$$K_\Theta := \Big\{ k \in K \mid k = \sum_{i=1}^N \alpha_i u_{\theta_i},\ \alpha_i \in [0, \infty),\ 1 \leq i \leq N \Big\}.$$

Define a *sieved estimator* by

$$\check{k}_n = \operatorname*{argmin}_{k \in K_\Theta} \Gamma_n L(k) = \operatorname*{argmin}_{\alpha_1 \geq 0, \ldots, \alpha_N \geq 0} \Big\| \sum_{i=1}^N \alpha_i z_{\theta_i} - \tilde{g}_n \Big\|_w^2. \tag{2.9}$$

Theorem 2.13 *The sieved estimator $\check{k}_n$ is uniquely defined.*

Proof This is a corollary to Proposition 2.7. Simply take $\varepsilon = \theta_1$ and ξ a measure on Θ in (2.4). □

Firstly, we approximate $\check{k}_n$ on a fixed grid. By virtue of Proposition 2.7 we can subsequently try to locally optimize the location of the support points. This provides us with a numerical approximation of $k_n^{(\varepsilon)}$, as long as the left-most support point remains bounded away from zero. Thus, we have two steps in mind: *(i)* compute $\check{k}_n$ on a fixed grid, *(ii)* "fine-tune" in the support points.

We now explain how we can numerically deal with this first step. Since each $k \in K_\Theta$ is a finite positive mixture of basis-functions $u_\theta \in \mathcal{U}_\Theta$, our minimization problem fits precisely in the setup of Groeneboom et al. (2003). We will follow their approach to solve (2.9).

2.3.1 The support reduction algorithm

Define the directional derivative of $\Gamma_n L$ at $k_1 \in K$ in the direction of $k_2 \in K$ by

$$D_{\Gamma_n L}(k_2; k_1) := \lim_{\varepsilon \to 0} \varepsilon^{-1}([\Gamma_n L](k_1 + \varepsilon k_2) - [\Gamma_n L](k_1)).$$

This quantity exists (it may be infinite), since $\Gamma_n L$ is a convex functional on K (Γ_n, as an L^2-distance on a Hilbert-space, is a strictly convex functional on G, and L satisfies $L(k_1 + k_2) = L(k_1) + L(k_2)$).

The solution of $\check{k}_n$ of (2.9) can be characterized by variational inequalities. The following result is taken from Groeneboom et al. (2003), Lemma 3.1.

Lemma 2.14 *Suppose*

(a1) $\Gamma_n L$ *is convex on* K_Θ *and has a unique minimizer over* K_Θ*. Moreover, for each* k_1, $k_2 \in K_\Theta$ *where* $\Gamma_n L$ *is finite, the function* $\varepsilon \mapsto \Gamma_n L(k_1 + \varepsilon(k_2 - k_1))$ *is continuously differentiable for* $\varepsilon \in (0, 1)$*.*

(a2) $\Gamma_n L$ *has the property that for each* $k_1 \in K_\Theta$ *and* $k_2 = \sum_{j=1}^N \alpha_j u_{\theta_j}$*,*

$$D_{\Gamma_n L}(k_2; k_1) = \sum_{j=1}^N \alpha_j D_{\Gamma_n L}(u_{\theta_j}; k_1).$$

Write $\check{k}_n = \sum_{j\in J} \alpha_j u_{\theta_j}$*, where* $J := \{j \in \{1, \ldots, N\} \mid \alpha_j > 0\}$*. Then*

$$\check{k}_n \text{ minimizes } \Gamma_n L \text{ over } K_\Theta \quad \iff \quad D_{\Gamma_n L}(u_{\theta_j}; \check{k}_n) \begin{cases} \geq 0 & \forall\, j \in \{1, \ldots, N\} \\ = 0 & \forall\, j \in J \end{cases} . \tag{2.10}$$

We verify conditions *(a1)* and *(a2)*.

(a1) The first two conditions were already checked. Since each function in G is continuous and the weight-function w is assumed to be compactly supported, dominated convergence implies that the map $\varepsilon \mapsto (\Gamma_n L)(k_1 + \varepsilon(k_2 - k_1))$ is continuously differentiable for all $\varepsilon \in (0, 1)$ and $k_1, k_2 \in K$ for which $\Gamma_n L$ is finite.

(a2) For k_1, $k_2 \in K$ and $\varepsilon > 0$ we have

$$\begin{aligned} [\Gamma_n L](k_1 + \varepsilon k_2) - [\Gamma_n L](k_1) &= \|L(k_1 + \varepsilon k_2) - \tilde{g}_n\|_w^2 - \|Lk_1 - \tilde{g}_n\|_w^2 \\ &= 2\varepsilon \langle Lk_1 - \tilde{g}_n, Lk_2\rangle_w + \varepsilon^2 \|Lk_2\|_w^2 \end{aligned} \tag{2.11}$$

Dividing by ε and upon letting $\varepsilon \downarrow 0$ we get

$$D_{\Gamma_n L}(k_2, k_1) = 2\langle Lk_1 - \tilde{g}_n, Lk_2\rangle_w \,.$$

If $k_2 = \sum_{i=1}^N \alpha_i u_{\theta_i}$, then

$$\begin{aligned} D_{\Gamma_n L}(k_2, k_1) &= 2\Big\langle Lk_1 - \tilde{g}_n, L\big(\sum_{i=1}^N \alpha_i u_{\theta_i}\big)\Big\rangle_w = 2\Big\langle Lk_1 - \tilde{g}_n, \sum_{i=1}^N \alpha_i L(u_{\theta_i})\Big\rangle_w \\ &= 2\sum_{i=1}^N \alpha_i \langle Lk_1 - \tilde{g}_n, L(u_{\theta_i})\rangle_w = \sum_{i=1}^N D_{\Gamma_n L}(u_{\theta_i}, k_1). \end{aligned}$$

The second equality holds by linearity of L. The last step holds since $D_{\Gamma_n L}(k_1, k_2) = D_{\Gamma_n}(L(k_1), L(k_2))$.

Thus, we conclude Lemma 2.14 applies to our minimization problem. Based on the characterization of the solution provided by the above lemma, we can numerically approximate $\check{k}_n$ by the support-reduction algorithm. This is an iterative algorithm for solving (2.9). We discuss this algorithm briefly and refer for additional details to Section 3 of Groeneboom et al. (2003).

Suppose at each iteration we are given a "current iterate" $k^J \in K_\Theta$ that can be written as

$$k^J = \sum_{j\in J} \alpha_j u_{\theta_j}$$

(J refers to the index set of strictly positive α-weights). Relation (2.10) gives a criterion for checking whether k^J is optimal. As we will shortly see, each iterate k^J will satisfy the equality part of (2.10): $D_{\Gamma_n L}(u_{\theta_j}; k^J) = 0$, for all $j \in J$. This fact, together with (2.10) implies that if k^J is not optimal, then there is an $i \in \{1, \ldots, N\} \backslash J$ with $D_{\Gamma_n L}(u_{\theta_i}, k^J) < 0$. Thus u_{θ_i} provides a direction of descent for $\Gamma_n L$. In that case the algorithm prescribes two steps that have to be carried out:

Step (i). Determine a direction of descent for $\Gamma_n L$. Let

$$\Theta_< := \{\theta \in \Theta \ : \ D_{\Gamma_n L}(u_\theta, k^J) < 0\},$$

then $\Theta_<$ is non-empty. From $\Theta_<$ we choose a direction of descent. Suppose θ_{j^*} is this direction. (A particular choice is the direction of steepest descent, in which case $\theta_{j^*} := \operatorname{argmin}_{\theta \in \Theta_<} D_{\Gamma_n L}(u_\theta, k^J)$. This boils down to finding a minimum element in a vector of length at most N. We give an alternative choice below.)

Step (ii). Let the new iterate be given by

$$k^{J^*} = \sum_{j \in J^*} \beta_j u_{\theta_j}, \quad J^* := J \cup \{j^*\},$$

where $\{\beta_j\, , j \in J^*\}$ are (yet unknown) weights. The second step consists of first minimizing $\Gamma_n L(k^{J^*})$ with respect to $\{\beta_j\, , j \in J^*\}$, without positivity constraints. In our situation this is a (usually low-dimensional) quadratic unconstrained optimization problem.

If $\min\{\beta_j\, , j \in J^*\} \geq 0$, then $k^{J^*} \in K_\Theta$ and k^{J^*} satisfies the equality part of (2.10). In that case, we check the inequality part of (2.10) and possibly return to step *(i)*. Otherwise, we perform a support-reduction step. Since it can be shown that β_{j^*} is always positive, we can make a move from k^J towards k^{J^*} and stay within the cone K_Θ initially. As a next iterate, we take $k := k^J + \hat{c}(k^{J^*} - k^J)$, where

$$\begin{aligned}
\hat{c} &= \max\{c \in [0,1] \ : \ k^J + c(k^{J^*} - k^J) \in K_\Theta\} \\
&= \max\{c \in [0,1] \ : \ \sum_{j \in J} [c\beta_j + (1-c)\alpha_j]\, u_{\theta_j} + c\beta_{j^*} u_{\theta_{j^*}} \in K_\Theta\} \\
&= \max\{c \in [0,1] \ : \ c\beta_j + (1-c)\alpha_j \geq 0\, , \text{ for all } \beta_j \ (j \in J) \text{ with } \beta_j < 0\} \\
&= \min\{\alpha_j/(\alpha_j - \beta_j)\ , j \in J \text{ for which } \beta_j < 0\}. \qquad (2.12)
\end{aligned}$$

Then $k \in K_\Theta$. Let j^{**} be the index for which the minimum in (2.12) is attained, i.e. for which $\hat{c}\beta_{j^{**}} + (1-\hat{c})\alpha_{j^{**}} = 0$. Define $J^{**} := J^* \setminus \{j^{**}\}$, then k is supported on $\{\theta_j\, , j \in J^{**}\}$. That is, in the new iterate, the support point $\theta_{j^{**}}$ is removed. Next, set $k^{J^{**}} = \sum_{j \in J^{**}} \gamma_j u_{\theta_j}$ and compute optimal weights γ_j without the positivity constraints. If all weights γ_j are non-negative, the equality part of (2.10) is satisfied and we can check the inequality part of (2.10) and possibly return to step *(i)*. Otherwise, a new support-reduction step can be carried out, since all weights of k are positive. In the end, our iterate k will satisfy the equality part of (2.10).

To start the algorithm, we fix a starting value $\theta^{(0)} \in \Theta$. Then we determine the function $cu_{\theta^{(0)}}$ minimizing $\Gamma_n L$ as a function of $c > 0$. Once the algorithm has been initialized it starts iteratively adding and removing support points, while in between computing optimal weights.

The following lemma gives conditions to guarantee that the sequence of iterates indeed converges to the solution of our minimization problem (see Theorem 3.1 in Groeneboom et al. (2003)).

Lemma 2.15 *Suppose $\Gamma_n L$ satisfies (a1), (a2) and also*

(a3) For any starting function $k^{(0)} \in K_\Theta$ with $\Gamma_n L(k^{(0)}) < \infty$, there exists an $\bar{\varepsilon} \in (0,1)$ such that for all $k \in K_\Theta$ with $\Gamma_n L(k) \le \Gamma_n L(k^{(0)})$ and $\theta \in \Theta$, the following implication holds for all $\delta > 0$:

$$D_{\Gamma_n L}(u_\theta - k; k) \le -\delta \quad \Rightarrow \quad \Gamma_n L(k + \varepsilon(u_\theta - k)) - \Gamma_n L(k) \le -\frac{1}{2}\varepsilon\delta\,,\ \forall \varepsilon \in (0, \bar{\varepsilon}]\,.$$

then the sequence of iterates $k^{(i)}$, generated by the support reduction algorithm, satisfies $(\Gamma_n L)(k^{(i)}) \downarrow (\Gamma_n L)(\check{k}_n)$, as $i \to \infty$.

From the calculations under *(a2)* it follows easily that if $D_{\Gamma_n L}(u_\theta - k, k) \le -\delta$ ($\delta > 0$), then

$$\begin{aligned}&(\Gamma_n L)(k + \varepsilon(u_\theta - k)) - (\Gamma_n L)(k)\\ &= \varepsilon D_{\Gamma_n L}(u_\theta - k, k) + \varepsilon^2 \|L(u_\theta - k)\|_w^2 \le -\varepsilon\delta + \varepsilon^2\|L(u_\theta - k)\|_w^2.\end{aligned}$$

The second term on the right-hand-side can be bounded by $2(\Gamma_n L)(u_\theta) + 2(\Gamma_n L)(k)$. Taking the supremum over all $k \in \{k \in K_\Theta \ : \ \Gamma_n L(k) \le \Gamma_n L(k^{(0)})\}$ and $\theta \in \Theta$ this is bounded by a positive finite constant $C > 0$.

Next, we can pick $\bar{\varepsilon} \in (0,1]$ such that $C\varepsilon^2 \le \frac{1}{2}\varepsilon\delta$ for all $\varepsilon \in (0, \bar{\varepsilon}]$, and *(a3)* readily follows.

Corollary 2.16 *For any starting function $k^{(0)} \in K_\Theta$ with $\Gamma_n L(k^{(0)}) < \infty$, the sequence of iterates $k^{(i)}$, generated by the support reduction algorithm, satisfies $(\Gamma_n L)(k^{(i)}) \downarrow (\Gamma_n L)(\check{k}_n)$, as $i \to \infty$.*

2.3.2 Implementation details

We now work out the actual computations involved, when implementing the algorithm. Suppose $k = \sum_{j=1}^m \alpha_j u_{\theta_j}$.

Step (i). Given the "current iterate" k, we aim to add a function u_θ which provides a direction of descent for $\Gamma_n L$. By linearity of L,

$$\begin{aligned}[\Gamma_n L](k + \varepsilon u_\theta) - [\Gamma_n L](k) &= \|L(k + \varepsilon u_\theta) - \tilde{g}_n\|_w^2 - \|Lk - \tilde{g}_n\|_w^2\\ &= \varepsilon c_1(\theta, k) + \tfrac{1}{2}\varepsilon^2 c_2(\theta)\,,\end{aligned} \tag{2.13}$$

where $c_2(\theta) = 2\|Lu_\theta\|_w^2 = 2\|z_\theta\|_w^2 > 0$ and

$$c_1(\theta, k) = 2\langle Lk - \tilde{g}_n, Lu_\theta\rangle_w = 2\Big\langle \sum_{j=1}^m \alpha_j z_{\theta_j} - \tilde{g}_n, z_\theta\Big\rangle_w.$$

In order to find a direction of descent, we can pick any $\theta \in \Theta$ for which $c_1(\theta, k) < 0$. However, since the right-hand side of (2.13) is quadratic in ε, it can be minimized explicitly (and we choose to do so). If $c_1(\theta, k) < 0$, then

$$\operatorname*{argmin}_{\varepsilon > 0}(\varepsilon c_1(\theta, k) + \tfrac{1}{2}\varepsilon^2 c_2(\theta)) = -\frac{c_1(\theta, k)}{c_2(\theta)} =: \hat{\varepsilon}_\theta\,.$$

Minimizing $[\Gamma_n L](k + \hat{\varepsilon}_\theta u_\theta)$ over all points $\theta \in \Theta$ with $c_1(\theta, k) < 0$ gives

$$\hat{\theta} = \underset{\{\theta\in\Theta\,:\,c_1(\theta,k)<0\}}{\operatorname{argmin}} -\frac{c_1(\theta,k)^2}{2c_2(\theta)} = \underset{\theta\in\Theta}{\operatorname{argmin}} \frac{c_1(\theta,k)}{\sqrt{c_2(\theta)}}.$$

The argmin exists since Θ is a finite set.

Step (ii). Given a set of support points, we compute optimal weights. This is a standard least squares problem, that is solved by the normal equations. In our setup, these are obtained by differentiating $[\Gamma_n L](k)$ with respect to α_j ($j \in \{1, \ldots, m\}$) and setting the partial derivatives equal to zero. This gives the system $A\underline{\alpha} = \underline{b}$, where

$$A_{i,j} = \langle z_{\theta_i}, z_{\theta_j}\rangle_w, \quad i, j = 1, \ldots, m\,, \tag{2.14}$$

and

$$b_i = \langle z_{\theta_i}, \tilde{g}_n\rangle_w, \quad i = 1, \ldots, m\,.$$

The matrix A is easily seen to be symmetric. By the next lemma, A is non-singular, whence the system $A\underline{\alpha} = \underline{b}$ has a unique solution.

Lemma 2.17 *The matrix A, as defined in (2.14) is non-singular.*

Proof Denote by $\underline{a}_j$ the j-th column of A. Let $h_1, \ldots, h_m \in \mathbb{R}$. We aim to show that if $\sum_{i=1}^m h_i \underline{a}_i = \underline{0}$, then all h_j are zero. Now $\sum_{i=1}^m h_i \underline{a}_i = (\langle z_{\theta_1}, \varphi\rangle_w, \ldots, \langle z_{\theta_m}, \varphi\rangle_w)'$, where $\varphi \in L^2(w)$ is given by $\varphi := \sum_{i=1}^m h_i z_{\theta_i}$. Thus if $\sum_{i=1}^m h_i \underline{a}_i = \underline{0}$, then $\varphi \perp \operatorname{span}(z_{\theta_1}, \ldots, z_{\theta_m})$ in $L^2(w)$. Since also $\varphi \in \operatorname{span}(z_{\theta_1}, \ldots, z_{\theta_m})$, we must have $\varphi = 0$ a.e. w.r.t. Lebesgue measure on S. By continuity of $t \mapsto z_\theta(t)$, $\varphi = 0$ on S.

Now $\sum_{k=1}^m h_k \omega_{\theta_k}(t) = 0$ for all $t \in S$, where $\omega_\theta(t) = t\frac{d}{dt} z_\theta(t) = \cos(\theta t) - 1 + i\sin(\theta t)$. Then also

$$\frac{d^p}{dt^p} \sum_{k=1}^m h_k \omega_{\theta_k}(t)\Big|_{t=0} = \sum_{k=1}^m h_k (i\theta_k)^p = 0, \quad p = 1, \ldots m.$$

Rewriting this in a linear system we find that

$$\begin{pmatrix} i\theta_1 & i\theta_2 & \ldots & i\theta_m \\ -\theta_1^2 & -\theta_2^2 & \ldots & -\theta_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ (i\theta_1)^m & (i\theta_2)^m & \ldots & (i\theta_m)^m \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Since the matrix in this display is a generalized Vandermonde matrix, its determinant is unequal to zero. Therefore, $h_1 = h_2 = \ldots = h_m = 0$. $\square$

Remark 2.18

(i) Tucker (1967), Section 4.3, gives an explicit way to calculate the imaginary part of $\tilde{g}_n$.

(ii) An estimator $\check{f}_n$ of the density function can be obtained by inverting the ch.f. $\psi_{\check{k}_n}$. We use the method of Schorr (1975), which is briefly discussed in Section 2.7.

(iii) Since for any ch.f. ψ we have $\psi(-t) = \overline{\psi(t)}$ and consequently $g(-t) = \overline{g(t)}$ for its associated cumulant, we can write

$$\Gamma_n(g) = 2\int_0^\infty |g(t) - \tilde{g}_n(t)|^2 w(t)dt.$$

This halves the amount of numerical integration.

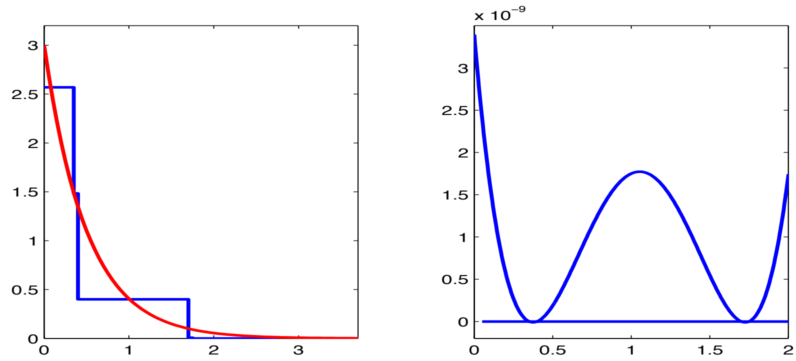

Figure 2.1: Cumulant estimator based on 1000 independent observations from a Gamma$(3, 2)$-distribution. The initial estimator is the empirical characteristic function (see Section 2.5). Left-hand: True (dashed) and estimated canonical function (solid). Right-hand: alternative directional derivative.

2.3.3 Local optimization of the support

The solution given by the support reduction algorithm typically contains support-points that are closely located. This is an unpleasant feature which is caused by the discreteness of the grid. A typical solution of the support reduction is given in Figure 2.1. The estimator was calculated with $w = \mathbf{1}_{[-1,1]}$, grid points $\theta_j = 0.05j$ $(j = 1, \ldots, 120)$ and the algorithm was run until the directional derivative did not drop below -10^{-15} on the grid. Looking at the right-hand picture of Figure 2.1, one can guess that the small steps near 0.40 and 1.70 can be removed. This is indeed the case.

At each iteration of the support-reduction algorithm, we can scan for two closely located points and, if available, determine the convex combination of the two that minimizes the objective function $\Gamma_n L$. If the replacement of the two closely located support points by a third one in between does not increase the objective function, this will not prevent the algorithm from converging (Groeneboom et al. (2003)). We implemented the minimization problem by a Fibonacci search algorithm. If the newly obtained solution decreases the objective function and is in the cone K_Θ (which will generally be the case), we have a successful join of points. As the addition of a replacement step at each iteration can make the algorithm converge slowly, we do this step only after the solution has been calculated on the grid, and not at each step of the algorithm. Slow convergence may happen for example if the algorithm finds the same two new directions of descent on and on, causing the algorithm to zigzag between these points. In that case, at each iteration there are closely located points and a Fibonacci search has to be carried out at each iteration. Besides this, a single interconnection at the end of the algorithm already removes (almost) all closely located points of the solution. For the example of Figure 2.1, Table 2.1 gives the support points and corresponding weights before and after the Fibonacci-step.

Remark 2.19 It is possible to add t a step which allows for "leaving the grid". Details on this can be found in Section 5 of Groeneboom et al. (2003). Since numerical tests have shown that this step has a negligible effect to the solution, we omit further discussion on

0.3500	1.0853
0.4000	1.0840
1.7000	0.3925
1.7500	0.0075

0.3757	2.1627
1.7004	0.4005

Table 2.1: Support points (left-hand column) with corresponding weights (right-hand column), before and after the Fibonacci-step.

this.

2.3.4 Testing the support reduction algorithm

We test the algorithm by taking as initial "estimator", the deterministic (i.e. non-random) ch.f. $\tilde{\psi}(t) = \exp(z_{2.02}(t) + 2z_3(t))$. Since this ch.f. corresponds to $k(x) = \mathbf{1}_{[0,2.02)}(x) + 2\mathbf{1}_{[0,3)}(x)$, we can easily verify whether the algorithm gives satisfactory results. We took as a grid $\theta_j = 0.05j$ $(j = 1, \ldots, 80)$ and ran the algorithm until the directional derivative on all grid points was above -10^{-15}. After that, we optimized the support locally, as described in the previous section. Table 2.2 shows the computed support points and their weights. The results look fine. We deliberately chose 2.02 as a support-point of the

2.02000000234422	1.00000000357463
3.00000000000000	1.99999996491368
3.05000000000000	0.00000003137456

Table 2.2: Support points (left-hand column) and corresponding weights (right-hand column).

true canonical function, to see how well the algorithm performs on leaving the grid by combining closely located points.

As a second step, we consider the effect of approximating a possibly smooth canonical function by a step-function. To get some feeling on how well our estimator works in this case, we took as an initial estimator the characteristic function of a Gamma random variable and computed the canonical function by the support-reduction-algorithm. As Figure 2.2 shows, a larger support of the weight-function improves the estimation-procedure (the estimator then makes more jumps). We come back to the choice of the weight-function in Section 2.5.2.

2.4 Extension to convex decreasing and completely monotone canonical functions

So far, we have assumed k is decreasing. If we make more assumptions on k, i.e. we restrict the set K, then we expect better results. Generally, there are two ways for restricting a nonparametric model: adding *smoothness assumptions* or adding *shape restrictions*. We consider the latter case. More specifically, we only deal with the the two cases where k

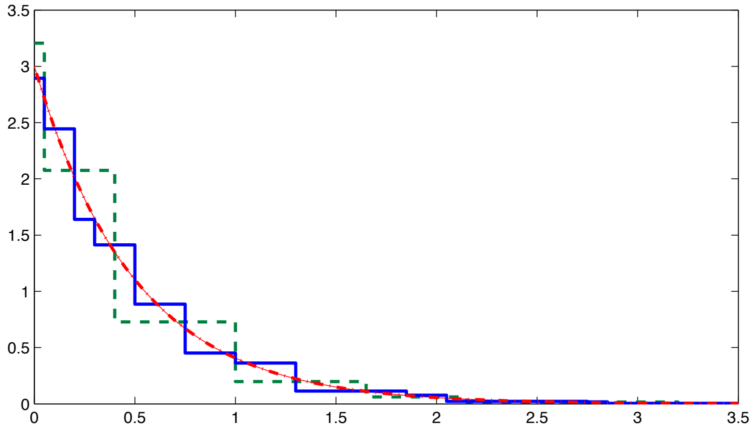

Figure 2.2: Test for the support-reduction-algorithm. The preliminary estimator is the ch.f. of a Gamma(3, 2) distribution, Dash-dotted: true canonical function. Dashed: estimator in case $w = \mathbf{1}_{[-1,1]}$. Solid: estimator in case $w = \mathbf{1}_{[-5,5]}$.

is assumed to be convex decreasing or completely monotone. We start with the first one. Let

$$K' = \{k \in \mathcal{L}^1(\mu) \ : \ k \text{ is convex and decreasing}\}$$

and define, for $R > 0$,

$$K'_R = \{k \in K' \ : \ k(x) \leq k_R(x), \text{ for } x \in (0, \infty)\}.$$

(The functions $(k_R, R > 0)$ were defined on page 37.) Note that $K'_R \subseteq K_R \subseteq \mathcal{L}^1(\mu)$.

Lemma 2.20 *The estimator* $\widehat{k'}_n = \operatorname{argmin}_{k \in K'_R} \Gamma_n L(k)$ *exists uniquely.*

Proof It suffices to prove that K'_R is a closed subset of $\mathcal{L}^1(\mu)$, since this implies that K'_R, as a closed subset of the compact set K_R, is compact. Subsequently, we can then follow the results of Section 2.1.

Let $\{k_n\}_n$ be a sequence in K'_R converging in $\mathcal{L}^1(\mu)$ to some k. Then there exists a subsequence $\{k_{n_j}\}_j$ such that $k_{n_j} \to k$ pointwisely μ-almost everywhere. However, since each k_{n_j} is convex (and hence continuous on the interior of its domain), $k_{n_j}(x) \to k(x)$ for all $x > 0$. Therefore, $k = \lim k_{n_j} \leq k_R$. Furthermore, the pointwise limit of a sequence of convex decreasing functions is convex decreasing (recall that by decreasing, we mean nonincreasing). Whence $k \in K'_R$ and K'_R is closed. $\square$

The sieved estimator can be defined analogously as before. We only need to replace the basis functions u_θ by

$$u'_\theta(x) = (\theta - x)_+ = \max(\theta - x, 0), \quad x > 0,$$

which implies that g_0 (the cumulant that we wish to estimate) is approximated by the cone that is generated by the functions $\{z'_\theta, \theta \in \Theta\}$ where

$$z'_\theta(t) = [Lu'_\theta](t) = \theta z_\theta(t) - \frac{\sin(\theta t)}{t} + \theta + i\frac{\cos(\theta t) - 1}{t}, \quad t \in \mathbb{R}.$$

Note that $\lim_{t\downarrow 0} z'_\theta(t) = 0$. The resulting sieved estimator is piecewise linear with increasing slope.

The completely monotone goes along similar lines. Let

$$K'' = \{k \in \mathcal{L}^1(\mu) \ : \ k \text{ is completely monotone}\}$$

and define, for $R > 0$,

$$K''_R = \{k \in K'' \ : \ k(x) \leq k_R(x), \text{ for } x \in (0, \infty)\}.$$

Lemma 2.21 *The estimator* $\widehat{k''}_n = \operatorname{argmin}_{k\in K''_R} \Gamma_n L(k)$ *exists uniquely.*

Proof The proof is analogous to the convex, decreasing case. It suffices to note that if $\{k_n\}_n$ is a sequence of completely monotone functions that converges pointwisely to a finite limit k, then k is also completely monotone (see Proposition A3.7*(iv)* in Van Harn and Steutel (2004)). $\square$

For the corresponding sieved estimator, the basis functions are given by

$$u''_\theta(x) = e^{-\theta x}, \quad x > 0,$$

which is a consequence of Bernstein's theorem on Laplace transforms (Theorem A3.6 in Van Harn and Steutel (2004)). The resulting generating functions for g_0 are given by $\{z''_\theta, \theta \in \Theta\}$ with

$$z''_\theta(t) = \int_0^\infty (e^{itx} - 1)\frac{e^{-\theta x}}{x} dx, \quad t \in \mathbb{R},$$

which is easily recognized as the cumulant to a *Gamma*$(1, \theta)$-distribution (see Example 1.8*(i)*). Therefore

$$z''_\theta(t) = -\log(1 - \theta^{-1}it) = -\tfrac{1}{2}\log(1 + t^2/\theta^2) + i \arctan(t/\theta), \quad t \in \mathbb{R}.$$

2.5 Applications and examples

We consider two observation schemes for (X_t):

(i) Observe (X_t) on a regularly spaced grid with fixed mesh-width Δ. Write $X_{k\Delta}$ for the observations $(k = 0, 1, \ldots)$.

(ii) As observation scheme *(i)*, but now suppose that the mesh-width Δ_n decreases as n increases. This gives, for each n, observations $(X_0, X_{\Delta_n}, X_{2\Delta_n}, \ldots, X_{(n-1)\Delta_n})$.

2.5.1 The empirical characteristic function as a preliminary estimator

In this section we prove that for both observation schemes the *empirical characteristic function* satisfies (2.3) and henceforth is an appropriate initial estimator.

For a certain stochastic process $(U_s, s \geq 0)$ the α-mixing "numbers" are defined by

$$\alpha_U(h) = 2 \sup_t \sup_{\substack{A\in\sigma(U_s,\, s\leq t)\\ B\in\sigma(U_s,\, s\geq t+h)}} |P(A \cap B) - P(A)P(B)|\,, \quad h > 0\,.$$

The process $(U_s\,, s \geq 0)$ is called α-mixing if $\alpha_U(h) \to 0$ as $h \to \infty$. As shown in e.g. Rio (2000), β-mixing (see Section 1.5) is a stronger property than α-mixing. In fact, for any process $(U_s\,, s \geq 0)$ we have $\alpha_U(h) \leq \beta_U(h)$ $(h > 0)$.

Proposition 2.22 *Suppose that for each $n \geq 1$, $X_0, X_{\Delta_n}, \ldots, X_{(n-1)\Delta_n}$ are n observations from an α-mixing stationary process $X = (X_t, t \geq 0)$. Let $\mathbb{F}_n$ denote the empirical distribution function of the observations. The empirical ch.f. is defined by*

$$\tilde{\psi}_n(t) := \int e^{itx} d\mathbb{F}_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} e^{itX_{j\Delta_n}}, \quad t \in \mathbb{R}.$$

(i) If $\Delta_n \equiv \Delta$, then $\tilde{\psi}_n$ satisfies (2.3) for convergence almost surely.

(ii) If $\Delta_n \downarrow 0$ and $n\Delta_n \to \infty$, then $\tilde{\psi}_n$ satisfies (2.3) for convergence in probability.

Proof (i): Since X is α-mixing, the chain $(X_{n\Delta}, n \geq 0)$ is *ergodic*. The first result now follows upon an application of Birkhoff's ergodic theorem (Krengel (1985), p. 9-10).

(ii): If we define for each fixed $u \in \mathbb{R}$ the continuous process $(Y^u_t,\, t \geq 0)$ by

$$Y^u_t = e^{iuX_t} - Ee^{iuX_t},$$

then

$$\tilde{\psi}_n(u) - \psi_0(u) = \frac{1}{n}\sum_{k=0}^{n-1} Y^u_{k\Delta_n}.$$

Denote the α-mixing numbers of $(|Y^u_t|\,, t \geq 0)$ by $\alpha_{|Y^u|}$, and similarly for $(|Y^u_{k\Delta_n}|)_k$ by $\alpha_{|Y^{u,n}|}$. Since $\sigma(|Y^u_t|,\, t \in T) \subseteq \sigma(X_t,\, t \in T)$ for any interval $T \subseteq [0, \infty)$, the definition of the α-mixing numbers implies that any $h > 0$, $\alpha_{|Y^u|}(h) \leq \alpha_X(h)$. In the same way one can verify that for $j \in \mathbb{N}$, $\alpha_{|Y^{u,n}|}(j) \leq \alpha_{|Y^u|}(j\Delta_n)$. Combining these inequalities gives: for $j \in \mathbb{N}$,

$$\alpha_{|Y^{u,n}|}(j) \leq \alpha_{|Y^u|}(j\Delta_n) \leq \alpha_X(j\Delta_n). \tag{2.15}$$

Lemma 2.28 implies that the following inequality holds: for each $h \in \mathbb{N}$

$$P\Big(\frac{1}{n}\sum_{k=0}^{n-1} |Y^u_{k\Delta_n}| \geq 2\varepsilon\Big) \leq \frac{2h}{n\varepsilon^2}\int_0^1 Q^2(1-w)dw + \frac{2}{\varepsilon}\int_0^{\alpha_{|Y^{u,n}|}(h)} Q(1-w)dw\,, \tag{2.16}$$

where $Q = F^{-1}_{|Y_1|}$. Since $P(|Y_1| \leq y) = 1$ if $y \geq 2$, we have $Q(u) \leq 2$ for all $u \in (0,1)$. Hence, for all $\varepsilon > 0$,

$$\begin{aligned} P\Big(\Big|\frac{1}{n}\sum_{k=0}^{n-1} Y^u_{k\Delta_n}\Big| \geq 2\varepsilon\Big) &\leq P\Big(\frac{1}{n}\sum_{k=0}^{n-1} |Y^u_{k\Delta_n}| \geq 2\varepsilon\Big) \\ &\leq \frac{8h}{n\varepsilon^2} + \frac{4}{\varepsilon}\alpha_{|Y^{u,n}|}(h) \leq \frac{8h}{n\varepsilon^2} + \frac{4}{\varepsilon}\alpha_X(h\Delta_n)\,, \end{aligned} \tag{2.17}$$

where the last inequality follows from (2.15).

Take $h = h_n = \sqrt{n/\Delta_n}$, then $h_n/n \to 0$ and $h_n\Delta_n \to \infty$ $(n \to \infty)$. Hence both terms in (2.17) can be made arbitrarily small by letting $n \to \infty$. $\square$

For the rest of this chapter, we assume the preliminary estimator is the empirical ch.f. Since a stationary OU-process is β-mixing, and hence α-mixing, we can apply Proposition 2.22. Combining this with Theorem 2.11, we see that for observation scheme *(i)*: $\|\hat{k}_n - k_0\|_\mu \xrightarrow{\text{a.s.}} 0$ $(n \to \infty)$, provided that $k_0 \in K_R$ for some $R > 0$. A similar statement holds for observation scheme *(ii)* with convergence in probability.

2.5.2 Data-adaptive choice of the weight-function

Up till now, the weight-function has been fixed in advance of the estimation procedure. One of our working hypotheses is that, for a sufficiently large number of observations, the logarithm of the preliminary estimator can be defined on the support of w. In this section we explain a data-adaptive procedure to ensure this assumption is satisfied. Let, for a small positive number η, and for $0 < m < M < \infty$

$$t_n = \inf\{t \in [m, M] \ : \ |\tilde{\psi}_n(t)| < \eta\},$$

denote the smallest number $t \in [m, M]$ such that the absolute value of the preliminary estimator at t drops below η. If the set on the right-hand-side is empty, we define $t_n = M$. Define a new weight-function $\tilde{w}_n$ by $\tilde{w}_n = w\mathbf{1}_{[-t_n,t_n]}$, for w the original weight-function. Hence, we restrict the support of the weight-function to the interval $\mathbf{1}_{[-t_n,t_n]}$, on which we can define the logarithm of the preliminary estimator.

For the random weight-function $\tilde{w}_n$ we now prove that the cumulant M-estimator is still well-defined and consistent. Define

$$\Gamma_{n,t_n} = \|(g - \tilde{g}_n)\sqrt{\tilde{w}_n}\|_2 = \|(g - \tilde{g}_n)\sqrt{w}w\mathbf{1}_{[-t_n,t_n]}\|_2,$$

in which $\|\cdot\|_2$ is the ordinary L^2-norm. Let

$$\hat{g}_{n,t_n} = \operatorname*{argmin}_{g\in G_R} \Gamma_{n,t_n}(g) \ \text{ and } \ \hat{k}_{n,t_n} = L^{-1}(\hat{g}_{n,t_n}).$$

Examining the steps that led to Theorem 2.6, we see that we only need to check if Lemma 2.3 goes through. The mapping L, with w replaced by $\tilde{w}_n$, is almost surely continuous and one-to-one due to the fact that $t_n \leq M$ and $t_n \geq m$ respectively. Hence, the upper bound M ensures continuity, whereas the lower bound m ensures identifiability. This suffices, since we can condition on the observed data and use the compactness of K_R.

For consistency, note that we can adapt the given proof (see Theorem 2.11) by the inequalities

$$\|(\hat{g}_{n,t_n} - g_0)\sqrt{\tilde{w}_n}\|_2 \leq 2\|(g_0 - \tilde{g}_n)\sqrt{\tilde{w}_n}\|_2 \leq \|(g_0 - \tilde{g}_n)\sqrt{w}\mathbf{1}_{[-M,M]}\|_2.$$

The right-hand side of this display tends to zero almost surely, by the argument given in the proof of Theorem 2.11. Since the inverse mapping L^{-1} is continuous, we therefore obtain that $\|\hat{k}_{n,t_n} - k_0\|_\mu \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$.

In the examples of the following section we will specify η, m, M and w, and calculate estimators using $\tilde{w}_n$.

2.5.3 A simple example

Suppose we observe a stationary OU-process under observation scheme *(i)* with $\Delta = 0.5$ and that the true invariant law of the process has $Gamma(1/3, 2)$ distribution. Figure shows the simulated process for $\lambda = 1$ on the interval $[0, 1000)$. Since $\Delta = 0.5$, we have $n = 2000$ observations. For the weight-function we took $\tilde{w}_n(t) = e^{-t/2}\mathbf{1}_{[-t_n,t_n]}(t)$ (t_n determined for $m = 0.1$, $M = 10$ and $\eta = 0.01$, it turned out that $t_n = 10$ for the data under consideration) and points $\theta_i = 0.05i$ for Θ. Some numerical results are plotted in Figure 2.4. We make the following remarks

- The estimate for the density function shows a much better fit than for the canonical function.
- The estimated value at zero equals $\check{k}_n(0) \approx 0.3005$, whereas $k_0(0) = 1/3$.
- The overlap of the estimated and empirical cumulant, indicate that it is much harder to estimate the canonical function than the cumulant function.
- We ran the algorithm until the solution k satisfied

$$\min_{\theta\in\Theta} c_1(\theta, k) > -10^{-15}. \tag{2.18}$$

  Using this accuracy, it may happen that we have to stop the iterations before the stopping criterion (2.18) is reached. This occurs if for a given iteration k, supported on points $\{\theta_j, j \in J\}$ ($J \subseteq \{1, \ldots, N\}$), the alternative directional derivative attains its minimum value at one of the support points θ_j ($j \in J$). In that case, the prescribed algorithm would pick a "new" support point which is already in the support of the solution. A finer grid may prevent this phenomenon to appear. On the other hand, a very fine grid can cause the matrix A in (2.14) to get close to singularity. Therefore, we first run the algorithm on a course grid and, once (2.18) has been reached, carry out the step as described in Section 2.3.3.
- In the lower-right plot in Figure 2.4, the value of the alternative derivative at zero is calculated by its limiting value

$$\lim_{\theta\downarrow 0} \frac{c_1(\theta, k)}{\sqrt{c_2(\theta)}} = \lim_{\theta\downarrow 0} \sqrt{2}\frac{\langle Lk - \tilde{g}_n, z_\theta\rangle_w}{\|z_\theta\|_w} = \sqrt{2}\frac{\langle Lk - \tilde{g}_n, \tau\rangle_w}{\|\tau\|_w},$$

  for τ the mapping defined by $t \mapsto it$. Although in the example at hand this value is positive, it may happen that the optimal solution yields a negative value of the alternative directional derivative near zero. This indicates that the optimal solution seeks for a point in $(0, \theta_1)$ to add to its support. Refining the grid near zero may be a solution to this. However, in advance of the numerical procedure, we should pick a small positive ε and take care that there are no support points smaller than ε (see remark 2.1 and Proposition 2.7).

As numerical experiments show, these observations hold in greater generality. In particular, there is not much special about the choice of a Gamma-distribution.

2.5.4 Example: data from the OU-process. Observation scheme *(i)*

For an Inverse-Gaussian($\delta > 0, \gamma \geq 0$) distribution, the canonical function is given by $k(x) = \frac{\delta}{\sqrt{2\pi x}} e^{-\gamma^2 x/2}\mathbf{1}_{(0,\infty)}(x)$. Since both $x \mapsto e^{-x}$ and $x \mapsto 1/\sqrt{x}$ are completely

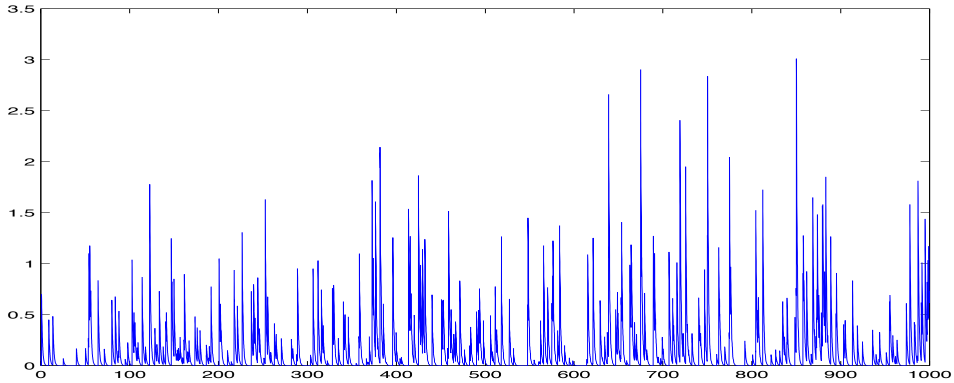

Figure 2.3: The sample path of a simulated stationary OU-process with Gamma$(1/3, 2)$ marginal distribution.

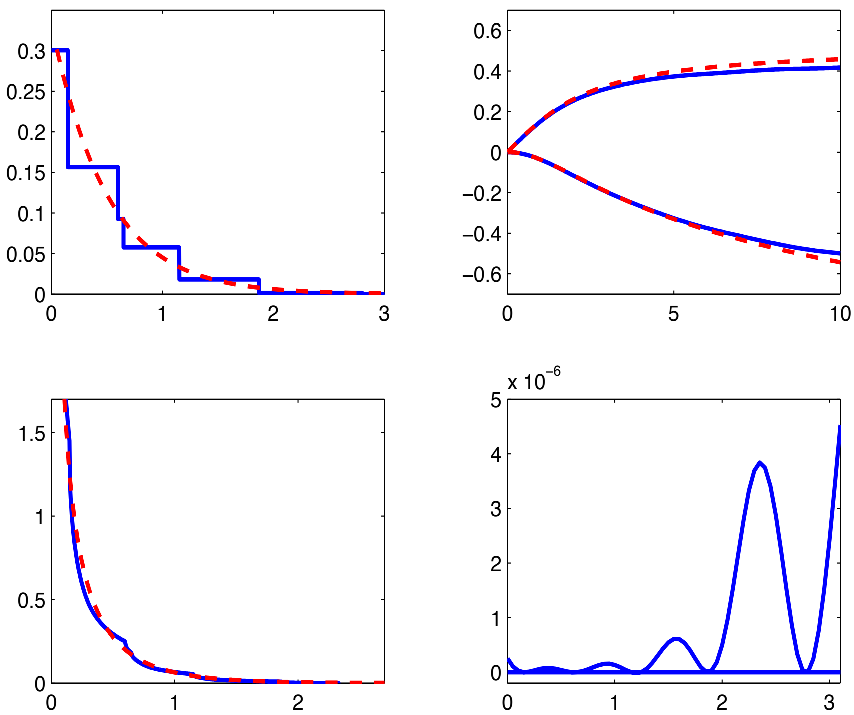

Figure 2.4: Computational results. Top-left: true (dashed) and estimated (solid) canonical function. Top-right: true (dashed), estimated (solid) and empirical (indistinguishable from the estimated) cumulant function. Bottom-left: true (dashed) and estimated (solid) density function. Bottom-right: alternative directional derivative at the optimal solution.

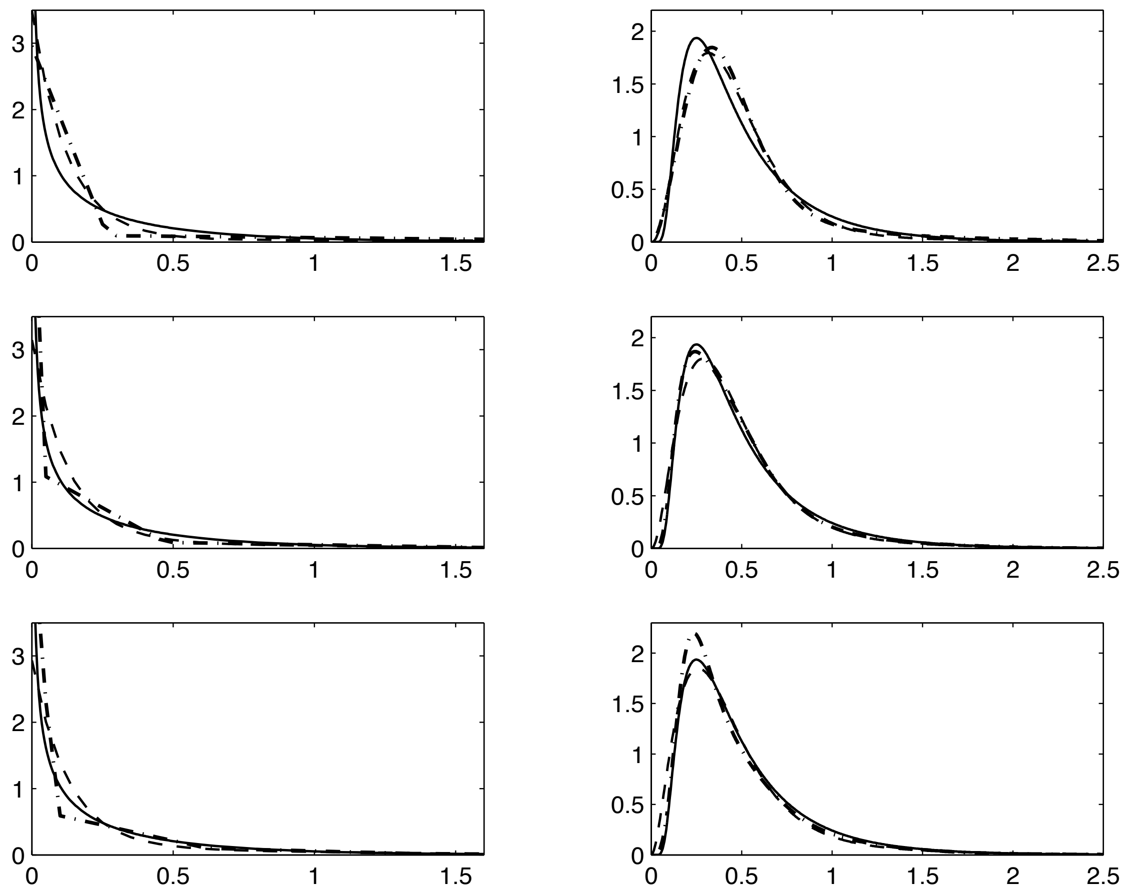

Figure 2.5: Inverse-Gaussian$(1, 2)$ distribution, Top: $n = 50$, middle: $n = 500$, bottom: $n = 3000$. Solid: k_0, dash-dotted: $\widehat{k'}_n$, dashed: $\widehat{k''}_n$. Left: canonical function, right: density function.

monotone (c.m.), and the product of c.m. functions on $(0, \infty)$ is c.m. (Van Harn and Steutel (2004), Proposition A.3.7*(iii)*), it follows that k is c.m. For samples of size $n = 50$, $n = 500$ and $n = 3000$ we compute the cumulant estimators $\widehat{k'}_n$ and $\widehat{k''}_n$. These estimators are respectively convex decreasing and c.m. We use the weight-function $\tilde{w}_n(t) = e^{-t/2}\mathbf{1}_{[-t_n,t_n]}$ with t_n determined by $m = 0.1$, $M = 10$ and $\eta = 0.01$. Furthermore, Θ consists of points $\theta_j = 0.05j$ $(j = 1, \ldots, 160)$ and we stopped the iteration as soon as the alternative directional derivative was above -10^{-15} for all θ_j. Figure 2.5 shows some numerical results.

2.5.5 Example: estimating a positive self-decomposable Lévy density

Suppose we observe a Lévy process $Z = (Z(t), t \geq 0)$ at discrete times $\{0, \Delta_n, \ldots, n\Delta_n\}$. Assume Z is positive self-decomposable with canonical function k. Since Z has stationary and independent increments, the scaled differences

$$Y_j^n := (\Delta_n)^{-1}(Z_{j\Delta_n} - Z_{(j-1)\Delta_n})$$

are independent and identically distributed with Lévy density $x \mapsto x^{-1}k(x)$. Hence, the distribution of Y_1^n is independent of n. Similar to Section 2.5.1, we obtain (by the strong law)

$$\frac{1}{n}\sum_{j=1}^{n} e^{itY_j^n} \xrightarrow{\text{a.s.}} Ee^{itY_1} = \exp\left(\int_0^\infty (e^{itx}-1)\frac{k_0(x)}{x}dx\right), \quad t \in \mathbb{R}.$$

Therefore, $\tilde{\psi}_n$ defined by $\tilde{\psi}_n(t) = \frac{1}{n}\sum_{j=1}^n e^{itY_j^n}$, converges pointwisely almost surely to ψ_0, as n tends to infinity. Consistency of $\hat{k}_n$ follows from Theorem 2.11.

A related problem is the estimation of a self-decomposable density, based on i.i.d. data. Let $Y_j = Z_j - Z_{j-1}$, then $\{Y_j\}_j$ is an i.i.d. sequence, say with canonical function k_0 and density function f_0. In general, a closed form expression for the density function f_0 in terms of k_0 is not known, This hampers the use of nonparametric maximum likelihood techniques for the estimation problem at hand. However, given $\hat{k}_n$, we can calculate $\hat{\psi}_n = \psi_{\hat{k}_n}$, and then numerically invert this function to obtain a non-parametric estimator $\hat{f}_n$ for the density f_0. In contrast to other nonparametric estimators (such as kernel estimators), our estimator $\hat{f}_n$ is guaranteed to be of the correct type (i.e. self-decomposable). Alternative preliminary estimators are also possible. For example, suppose we know, in addition to the assumptions already made, that the density of f_0 is decreasing. Then we can take as a preliminary estimator $\tilde{\psi}_n(t) = \int e^{itx}F_{n,Gren}(dx)$ $(t \in \mathbb{R})$, where $F_{n,Gren}$ is the Grenander estimator, which is defined as the least concave majorant of the empirical distribution function $\mathbb{F}_n$. Using similar arguments as before, we can show that the estimator for k based on $F_{n,Gren}$ is consistent. As another example, one may also take the maximum likelihood estimator for a unimodal density as a preliminary estimator for f_0. This makes sense, since every self-decomposable density is unimodal (Sato (1999), Theorem 53.1).

2.6 Some additional remarks

- The estimation framework as outlined in Section 2.1 is applicable in a more general context. Let $Y = (Y(t), t \geq 0)$ be a stationary process with invariant law π satisfying
$$\int e^{itx}d\pi(x) = \exp\left(\int_0^\infty (e^{itx}-1)a(x)dx\right), \quad t \in \mathbb{R},$$
for a function $a : [0,\infty) \to [0,\infty)$ satisfying $\int_0^\infty (x \wedge 1)a(x)dx < \infty$. That is, the marginal law of Y is positive infinitely divisible with Lévy measure that admits a density. If a satisfies a shape restriction (for example a is decreasing), then all our previous results can be restated by redefining the measure μ by $\mu(dx) = (x \wedge 1)dx$ $(x \geq 0)$.
- More generally, suppose π satisfies
$$\int e^{itx}d\pi(x) = \exp\left(\int (e^{itx}-1-itx\mathbf{1}_{[-1,1]}(x))a(x)dx\right), \quad t \in \mathbb{R}, \tag{2.19}$$
for a function $a : \mathbb{R}_0 \to [0,\infty)$ satisfying $\int_0^\infty (x^2 \wedge 1)a(x)dx < \infty$. In that case, we cannot apply Lemma 2.26. This lemma is important, since it ensures that the

parameter that we wish to estimate, a, is determined by the values of its corresponding ch.f. on a neighborhood of the origin. By Lemma 2.27, this property can be restored by restricting the set of infinitely divisible distributions to those for which $\int e^{r|x|}\pi(dx) < \infty$ for some $r > 0$. Since $\int e^{r|x|}\pi(dx) < \infty$ if and only if $\int_{|x|>1} e^{r|x|}a(x)dx < \infty$ (Sato (1999), Theorem 2.5.3), we consider the set of infinitely divisible distributions as in (2.19) with Lévy density in the class

$$\begin{aligned} A &= \Big\{ a : \mathbb{R}_0 \to [0,\infty) \mid a \text{ satisfies} \\ &\int_{|x|\leq 1} x^2 a(x)dx + \int_{|x|>1} e^{r|x|}a(x)dx < \infty, \text{ for some } r > 0 \Big\}. \end{aligned} \tag{2.20}$$

For this set of infinitely divisible distributions we can extend the estimation framework from Section 2.1.

- A popular model in mathematical finance is the CGMY-model (Carr et al.(2002)), in which log-returns of stock price are modeled by a pure-jump Lévy process. More precisely, this model postulates that the Lévy measure of the Lévy process is of the form
$$\nu(dx) = \frac{C}{|x|^{Y+1}}\big(e^{-G|x|}\mathbf{1}_{(-\infty,0)}(x) + e^{-Mx}\mathbf{1}_{(0,\infty)}(x)\big)dx$$
where $C > 0$, $G \geq 0$, $M \geq 0$ and $Y < 2$. If $Y \in (-1, 2)$, the Lévy density a is completely monotone and hence in the set A as defined in (2.20). We could replace the four-parameter Lévy density by a density in the class
$$\Big\{ a \text{ is a Lévy density on } \mathbb{R}_0 \text{ for which } a(x) = \int_0^\infty e^{-u|x|}\xi(du) \Big\}.$$
Here ξ is a measure on the positive halfline. For this model nonparametric estimation is feasible by our methods.
- From the nonparametric cumulant-M-estimator we can construct a parametric estimator by regression of k_θ (θ denotes the unknown parameter) to $\hat{k}_n$. Choose design-points $\{x_1, \ldots, x_p\}$ and define the parametric estimator as
$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{p} \big(k_\theta(x_i) - \hat{k}_n(x_i)\big).$$
The design points can be chosen for example as the mid-points of the support-points of $\hat{k}_n$. In the next chapter, we define another parametric estimator that is based on the same criterion function as the nonparametric estimator.

2.7 Computing the density from the canonical function

If $\hat{k}_n$ is an estimator for k_0, then we can invert the ch.f. $\psi_{\hat{k}_n}$ to obtain an estimator $\hat{f}_n$ for the density function, which is known to exists (a property shared by all nondegenerate self-decomposable laws). Moreover, by Theorem 28.4 in Sato (1999), this density is continuous on $(0,\infty)$. In this section we show how we can compute $\hat{f}_n$.

Let ψ be the ch.f. of a probability measure with density f that vanishes on the negative half-line. If $\int |\psi(t)|dt < \infty$ then, by the inversion formula for characteristic functions, $f(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} \psi(u) e^{-iux} du$ ($x \in \mathbb{R}$). However, the ch.f. of a self-decomposable distribution need not be integrable, a counterexample being given by the exponential distribution. Therefore, we can in general not apply the inversion formula and have to resort to different methods. One such method is the inversion method of Schorr (1975), which we will now explain briefly, without going into details. We first symmetrize the density: define

$$\tilde{f}(x) = \frac{1}{2} f(|x|), \quad x \in \mathbb{R}.$$

Then $\tilde{\psi}(t) = \int_{\mathbb{R}} e^{itx} \tilde{f}(x) dx = \Re\psi(t)$; the real part of ψ. Inversion of $\tilde{\psi}$ will then give $\tilde{f}$ and hence f on $(0, \infty)$. The advantage of this approach is that we only need to consider real-valued characteristic functions.

Define, for $T > 0$,

$$h(x) = \tilde{f}(x) \mathbf{1}_{[-T,T]}(x), \quad x \in \mathbb{R}.$$

We expand h in a Fourier-cosine series:

$$h(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi x}{T}\right), \tag{2.21}$$

where

$$a_k = \frac{1}{T} \int_{-T}^{T} h(x) \cos\left(\frac{k\pi x}{T}\right) dx = \frac{2}{T} \int_{0}^{T} h(x) \cos\left(\frac{k\pi x}{T}\right) dx, \quad k = 0, 1, \ldots.$$

An easy calculation shows that

$$\frac{1}{T} \tilde{\psi}\left(\frac{k\pi}{T}\right) = a_k + \frac{1}{T} \int_{T}^{\infty} f(x) \cos\left(\frac{k\pi x}{T}\right) dx, \quad k = 0, 1, \ldots.$$

Now we suppose that it is possible to choose T such that we can ignore the second term on the right-hand side of the preceding display. Substituting the above expression for a_k into (2.21) gives,

$$\tilde{f}(x) \approx \frac{1}{2T} + \frac{1}{T} \sum_{k=1}^{\infty} \tilde{\psi}\left(\frac{k\pi}{T}\right) \cos\left(\frac{k\pi x}{T}\right), \quad x \in [-T, T].$$

In the actual implementation we have to truncate the summation in the above expression. Let $M - 1$ be the truncation point. Let $x_j = \frac{2(j-1)T}{M}$, $j = 1, \ldots, M/2 + 1$, then

$$\begin{aligned} f(x_j) &\approx \frac{1}{T} + \frac{2}{T} \sum_{k=1}^{M-1} \tilde{\psi}\left(\frac{k\pi}{T}\right) \cos\left(\frac{2\pi(j-1)k}{M}\right) \\ &= -\frac{1}{T} + \frac{2}{T} \sum_{k=1}^{M} \tilde{\psi}\left(\frac{(k-1)\pi}{T}\right) \cos\left(\frac{2\pi(k-1)(j-1)}{M}\right). \end{aligned} \tag{2.22}$$

Based on $\check{k}_n$ we calculate $\tilde{\psi}$ on a grid $0 < \pi/T < 2\pi/T < \ldots < (M-1)\pi/T$. The vector containing these values will be denoted by $\widetilde{\Psi}$. The *fast Fourier transform* of a (complex-valued) vector $(y_1, \ldots, y_M)$ is given by

$$[FFT(y)](j) := \sum_{k=1}^{M} y(k) \exp(-2\pi i(k-1)(j-1)/M), \quad j = 1, \ldots, M.$$

From (2.22) and the definition of the FFT we get

$$f(x_j) \approx \frac{-1 + 2\Re[FFT(\widetilde{\Psi})]_j}{T}, \quad j = 1, \ldots, M/2+1.$$

2.8 Estimation of the intensity parameter λ

Fix $\Delta > 0$ and suppose $X_0, X_\Delta, \ldots, X_{n\Delta}$ are discrete-time observations from the stationary OU-process. In this section, we define an estimator for λ. For ease of notation we write $X_i = X_{i\Delta}$ $(i = 0, \ldots, n)$. By Proposition 1.20, for $n \geq 1$, $X_n = e^{-\lambda} X_{n-1} + W_n(\lambda)$, where $\{W_n(\lambda)\}_{n\geq 1}$ is a sequence of independent random variables with common infinitely divisible distribution. Since $(X_n)_{n\geq 0}$ is stationary, $X_0 \sim \pi_0$, where π_0 has Lévy density $x \mapsto \rho(x, \infty)/x$.

Let $\theta = e^{-\lambda}$ and denote the true parameter by θ_0. Since $W_n(\lambda) \geq 0$ for each $n \geq 1$, we easily obtain the bound $\theta_0 \leq \min_{n\geq 1} \frac{X_n}{X_{n-1}}$. Define the estimator

$$\hat{\theta}_n = \min_{1\leq k\leq n} \frac{X_k}{X_{k-1}}.$$

Then $\hat{\theta}_n(\omega) \geq \theta_0$, for each ω. Hence $\hat{\theta}_n$ is always biased upwards. However, we have

Lemma 2.23 *The estimator $\hat{\theta}_n$ is consistent: $\hat{\theta}_n \xrightarrow{\mathrm{p}} \theta_0$, as n tends to infinity.*

Proof Let $\varepsilon > 0$. Since $\{|\hat{\theta}_n - \theta_0| > \varepsilon\} = \{\hat{\theta}_n > \theta_0 + \varepsilon\}$,

$$\begin{aligned} p(n,\varepsilon) := P(|\hat{\theta}_n - \theta_0| > \varepsilon) &= P(X_k/X_{k-1} > \theta_0 + \varepsilon, \forall\, k \in \{1, \ldots, n\}) \\ &= P(W_k(\lambda) > \varepsilon X_{k-1}\,, \forall\, k \in \{1, \ldots, n\}) \end{aligned}$$

Define $N_n := \sum_{k=1}^{n} \mathbf{1}\{X_{k-1} > 1\}$, then

$$\begin{aligned} p(n,\varepsilon) &= \sum_{j=0}^{n} P(W_k(\lambda) > \varepsilon X_{k-1}\,, \forall\, k \in \{1, \ldots, n\} \,|\, N_n = j)\, P(N_n = j) \\ &\leq \sum_{j=0}^{n} (P(W_1(\lambda) > \varepsilon))^j\, P(N_n = j), \end{aligned}$$

where the inequality holds since $\{W_k(\lambda)\}_{k\geq 1}$ is an i.i.d. sequence. Since $W_1(\lambda)$ has support $[0, \infty)$ (Sato (1999), Corollary 24.8), $\alpha_\varepsilon := P(W_1(\lambda) > \varepsilon) \in [0, 1)$. This gives

$$p(n,\varepsilon) \leq \sum_{j=0}^{\infty} \alpha_\varepsilon^j\, P(N_n = j).$$

By dominated convergence, $\lim_{n\to\infty} p(n,\varepsilon) \leq \sum_{j=0}^{\infty} \alpha_\varepsilon^j [\lim_{n\to\infty} P(N_n = j)]$. We are done, once we have proved that $\lim_{n\to\infty} P(N_n = j) = 0$.

We claim $N_n \xrightarrow{\text{a.s.}} \infty$, as $n \to \infty$. From Section 1.5 we know that $\lim_{n\to\infty} \|P^n(x,\cdot) - \pi_0\|_{TV} = 0$, for all $x \in E$. By Proposition 6.3 in Nummelin (1984), this implies that the chain $(X_n)_n$ is positive Harris-recurrent and a-periodic. Now Harris recurrence implies that the set $(1,\infty)$ is visited infinitely many times by $(X_n)_n$, almost surely. Therefore, the claim holds and we conclude that $p(n,\varepsilon) \to 0$ $(n \to \infty)$. $\square$

By the continuous mapping theorem we have

Corollary 2.24 *Define* $\hat{\lambda}_n = -\log \hat{\theta}_n$, *then* $\hat{\lambda}_n \xrightarrow{\text{p}} \lambda_0$, *as* $n \to \infty$, *where* λ_0 *denotes the true value of* λ.

Remark 2.25 *(i)* If the innovations $W_n(\lambda)$ are exponentially distributed, $\hat{\theta}_n$ equals the maximum likelihood estimator for the model. A detailed asymptotic analysis for this model is given in Nielsen and Shephard (2003).

(ii) If the innovations $W_n(\lambda)$ have a finite second moment, θ_0 can also be estimated by the Yule-Walker estimator for the first order auto-correlation coefficient. This estimator is given by $(\sum_{i=1}^{n-1}(X_i - \bar{X}_n)(X_{i+1} - \bar{X}_n))/(\sum_{i=1}^{n}(X_i - \bar{X}_n)^2)$.

2.9 An alternative estimation method

An alternative approach towards the estimation of a Lévy density of a Lévy process is given in Figueroa-López and Houdré (2004). We now sketch their estimation method and how this method may also be used to estimate the canonical function of an OU-process nonparametrically. For the moment, suppose we continuously observe a Lévy process $(Z(t), t \geq 0)$, characterized by its Lévy measure ν, on an interval $[0,T]$. We wish to estimate the Lévy density $a(x) = \frac{\nu(dx)}{dx}$ (which we assume to exist) on a fixed interval D. Suppose ν is absolutely continuous with respect to a known measure η and that the Radon-Nikodym derivative

$$\frac{d\nu}{d\eta}(x) = s(x)$$

is positive, bounded and satisfies $\int_D s^2(x)\eta(dx) < \infty$. Then s is called the *regularized Lévy density*, and is an element of the Hilbert space $L^2 \equiv L^2(D,\eta)$.

Let S be a finite dimensional subspace of L^2. The projection estimator of s on S is defined by

$$\hat{s}(x) = \sum_{i=1}^{d} \hat{\beta}_i \varphi_i(x),$$

where $\{\varphi_1, \ldots, \varphi_d\}$ is an arbitrary orthonormal basis of S and

$$\hat{\beta}_i = \frac{1}{T} \int\int_{[0,T]\times D} \varphi_i(x) N(dt,dx).$$

Here N is the Poisson random measure that determines the jumps of Z (see Section 1.2). In fact, $\hat{s}$ is the unique minimizer of the random functional

$$\gamma_D(f) = -\frac{2}{T} \int\int_{[0,T]\times D} f(x) N(dt,dx) + \int_D f^2(x)\eta(dx).$$

The estimator $\hat{s}$ is an unbiased estimator for the projection of s on S, which we denote by $s^{\perp}$.

Suppose we have a sequence of sets $\{S_m\}_{m\geq 1}$ with good approximation properties for S. In the estimation procedure for a particular set S_m, we can decompose the mean square error (MSE) as follows

$$E\|\hat{s}_m - s\|^2 = \|s - s^{\perp}\|^2 + E\|\hat{s}_m - s^{\perp}\|^2.$$

The main accomplishment in Figueroa-López and Houdré (2004) is to choose the model that balances the MSE in an optimal way. This is achieved by adding a penalty term to the objective function. We do not go into further details here.

To compute the proposed estimator, we need to calculate the functional

$$I(f) = \int\int_{[0,T]\times D} f(x)N(dt, dx) = \sum_{t\leq T} f(\Delta Z(t)),$$

which is possible, if one can observe all jumps of the Lévy process. If this is not the case, an *approximate projection estimator* can be defined, by approximating I by

$$I_n(f) = \sum_{t\leq T} f(\Delta Z^n(t)) = \sum_{k=1}^{n} f(Z(t_k^n) - Z(t_{k-1}^n)), \tag{2.23}$$

where $t_k^n = \frac{k}{n}T$. By small-time distributional properties of Z, one can prove that $I_n(f)$ converges weakly to $I(f)$ if f satisfies certain conditions. Two sufficient conditions are: *(i)* $f(x) = \mathbf{1}_{(a,b]}(x)h(x)$ for an interval $(a,b] \subseteq \mathbb{R}_0$ and a continuous function h, *(ii)* f is continuous on $\mathbb{R}_0$ and $\lim_{|x|\to 0} |x|^{-2}f(x) = 0$. Thus, Figueroa-López and Houdré (2004) provide an estimation method for the Lévy density of a Lévy process, based on high-frequency data.

We now explain that this method can be extended to OU-processes, driven by a subordinator. Let X be a stationary OU-process, driven by the Lévy process Z. Assume that the intensity parameter of the OU-process equals one, so that the jumps of X and Z coincide. Consider

$$X^n(t) = X(T)\mathbf{1}_{[T,\infty)}(t) + \sum_{k=1}^{n} X(t_{k-1}^n)\mathbf{1}_{[t_{k-1}^n, t_k^n)}(t), \quad t \geq 0,$$

then X^n converges to X in the Skorohod topology (details on the Skorohod topology can be found in Pollard (1984) or Billingsley (1968)). This is a consequence of lemmas 4 and 5 of Chapter VI in Pollard (1984). By Proposition VI.3.16 in Jacod and Shiryaev(1987), this implies that for any continuous function f vanishing in a neighborhood of zero, the process $\sum_{t\leq \cdot} f(\Delta X^n(t))$ converges to the process $\sum_{t\leq \cdot} f(\Delta X(t))$ in the Skorohod topology. Thus,

$$\sum_{t\leq T} f(\Delta X^n(t)) = \sum_{k=1}^{n} f(X(t_k^n) - X(t_{k-1}^n)) \rightsquigarrow \sum_{t\leq T} f(\Delta X(t)) = \sum_{t\leq T} f(\Delta Z(t)).$$

The left-hand side of this expression can be used for $I_n(f)$ as defined in (2.23). Subsequently, we can apply the method of Figueroa-López and Houdré (2004).

2.10 Some technical results

The following result extends the uniqueness theorem for characteristic functions.

Lemma 2.26 *Let X be a positive random variable with characteristic function ψ. If for $\delta > 0$, ψ_δ is the restriction of ψ to an interval $(-\delta, \delta)$, then ψ_δ determines ψ.*

Proof Since X is positive, the definition of ψ can be extended to the upper half of the complex plane: if $A := \{z \in \mathbb{C} \mid Im(z) > 0\}$, then for $z = s + it \in A$ we have

$$|\psi(z)| = |Ee^{izX}| = |Ee^{isX-tX}| \le Ee^{-tX} < \infty\,.$$

On the real line, ψ is (uniformly) continuous and on A, ψ is analytic. To see the analyticity, first note that by dominated convergence, for each $m > 0$ and $z \in A$, the mapping $z \mapsto \psi^{(m)}(z) := \int_{[0,m]} e^{izx}\mu(dx)$ is analytic. Here μ denotes the distribution of X. Since

$$\left|\int_{[0,m]} e^{izx}\mu(dx) - \int e^{izx}\mu(dx)\right| \le \mu(m,\infty) \to 0 \quad (m \to \infty),$$

independently of z, $\psi^{(m)}$ converges uniformly to ψ on A ($m \to \infty$). By Theorem V.1.5 in Lang (1985), a uniform limit of analytic functions is analytic.

By Schwarz' reflection principle (Lang (1985), Theorem VIII.1.1*(ii)*), there exists a unique analytic continuation of ψ to the lower half of the complex plane. Put $\bar{A} := \{z \in \mathbb{C} \mid Im(z) < 0\}$ and define for $z \in \bar{A}$, $\psi(z) = \overline{\psi(\bar{z})}$, where "$\bar{z}$" denotes the complex conjugate of a complex number z. Similarly, the function ψ_δ has a unique analytic continuation on $\{z \in \mathbb{C} \mid -\delta < \Re z < \delta\}$. Since ψ and ψ_δ agree on a disc centered at the origin, they are equal (Theorem II.3.2 in Lang (1985)). Whence, ψ_δ determines ψ. □

Lemma 2.27 *Let X be a random variable with characteristic function ψ and distribution μ. If $\int e^{r|x|}\mu(dx) < \infty$ for some $r > 0$, then ψ is uniquely determined by its values on a neighborhood of the origin.*

Proof By the integrability condition, ψ is analytic on $\{z \in \mathbb{C} : |z| < r\}$. By Theorem 1.7.7 in Ushakov (1999), an analytic ch.f. is determined by its values on a neighborhood of the origin. □

The statement and proof of the following lemma are similar to Theorem 3.2 in Rio (2000).

Lemma 2.28 *For any mean zero time series X_t with α-mixing numbers $\alpha(h)$, every $x > 0$ and every $h \in \mathbb{N}$, with $Q_t = F^{-1}_{|X_t|}$ the quantile function of $|X_t|$,*

$$P(\bar{X}_n \ge 2x) \le \frac{2}{nx^2}\int_0^1 \big(\alpha^{-1}(u) \wedge h\big)\,\frac{1}{n}\sum_{t=1}^n Q_t^2(1-u)\,du + \frac{2}{x}\int_0^{\alpha(h)} \frac{1}{n}\sum_{t=1}^n Q_t(1-u)\,du.$$

Proof The quantile function of the variable $|X_t|/(xn)$ is equal to $u \mapsto F^{-1}_{|X_t|}(u)/(nx)$. Therefore, by a rescaling argument we can see that it suffices to bound the probability $P\big(\sum_{t=1}^n X_t \ge 2\big)$ by the right side of the lemma, but with the factors $2/(nx^2)$ and $2/x$

replaced by 2 and the factor n^{-1} in front of $\sum Q_t^2$ and $\sum Q_t$ dropped. For ease of notation set $S_0 = 0$ and $S_n = \sum_{t=1}^n X_t$.

Define the function $g : \mathbb{R} \to \mathbb{R}$ to be 0 on the interval $(-\infty, 0]$, to be $x \mapsto \frac{1}{2}x^2$ on $[0, 1]$, to be $x \mapsto 1 - \frac{1}{2}(x-2)^2$ on $[1, 2]$, and to be 1 on $(2, \infty)$. Then g is continuously differentiable with uniformly Lipschitz derivative. By Taylor's theorem it follows that $\bigl|g(x) - g(y) - g'(x)(x-y)\bigr| \le \frac{1}{2}|x-y|^2$ for every $x, y \in \mathbb{R}$. Because $1_{[2,\infty)} \le g$ and $S_t - S_{t-1} = X_t$,

$$P(S_n \ge 2) \le Eg(S_n) = \sum_{t=1}^n E\bigl(g(S_t) - g(S_{t-1})\bigr) \le \sum_{t=1}^n E\bigl|g'(S_{t-1})X_t\bigr| + \frac{1}{2}\sum_{t=1}^n EX_t^2.$$

The last term on the right can be written $\frac{1}{2}\sum_{t=1}^n \int_0^1 Q_t^2(1-u)\,du$, which is bounded by $\sum_{t=1}^n \int_0^{\alpha(0)} Q_t^2(1-u)\,du$, because $\alpha(0) = \frac{1}{2}$ and $u \mapsto Q_t(1-u)$ is decreasing.

For $i \ge 1$ the variable $g'(S_{t-i}) - g'(S_{t-i-1})$ is measurable relative to $\sigma(X_s : s \le t-i)$ and is bounded in absolute value by $|X_{t-i}|$. Therefore, Lemma 2.29 yields the inequality

$$\bigl|E\bigl(g'(S_{t-i}) - g'(S_{t-i-1})\bigr)X_t\bigr| \le 2\int_0^{\alpha(i)} Q_{t-i}(1-u)Q_t(1-u)\,du.$$

For $t \le h$ we can write $g'(S_{t-1}) = \sum_{i=1}^{t-1}\bigl(g'(S_{t-i}) - g(S_{t-i-1})\bigr)$. Substituting this in the left side of the following display and applying the preceding display, we find that

$$\sum_{t=1}^h E\bigl|g'(S_{t-1})X_t\bigr| \le 2\sum_{t=1}^h\sum_{i=1}^{t-1}\int_0^{\alpha(i)} Q_{t-i}(1-u)Q_t(1-u)\,du.$$

For $t > h$ we can write $g'(S_{t-1}) = g'(S_{t-h}) + \sum_{i=1}^{h-1}\bigl(g'(S_{t-i}) - g(S_{t-i-1})\bigr)$. By a similar argument, this time also using that the function $|g'|$ is uniformly bounded by 1, we find

$$\sum_{t=h+1}^n E\bigl|g'(S_{t-1})X_t\bigr| \le 2\sum_{t=h+1}^n\int_0^{\alpha(h)} Q_t(1-u)\,du + 2\sum_{t=h+1}^n\sum_{i=1}^{h-1}\int_0^{\alpha(i)} Q_{t-i}(1-u)Q_t(1-u)\,du.$$

Combining the preceding displays we obtain that $P(S_n \ge 2)$ is bounded above by

$$2\sum_{t=1}^n\int_0^{\alpha(h)} Q_t(1-u)\,du + 2\sum_{t=1}^n\sum_{i=1}^{t\wedge h-1}\int_0^{\alpha(i)} Q_{t-i}(1-u)Q_t(1-u)\,du + \frac{1}{2}\sum_{t=1}^n EX_t^2.$$

In the second term we can bound $2Q_{t-i}(1-u)Q_t(1-u)$ by $Q_{t-i}^2(1-u) + Q_t^2(1-u)$ and next change the order of summation to $\sum_{i=1}^{h-1}\sum_{t=i+1}^n$. Because $\sum_{t=i+1}^n (Q_{t-i}^2 + Q_t^2) \le 2\sum_{t=1}^n Q_t^2$ this term is bounded by $2\sum_{i=1}^{h-1}\int_0^{\alpha(i)}\sum_{t=1}^n Q_t^2(1-u)\,du$. Together with the third term on the right this gives rise to the first sum on the right of the lemma, as $\sum_{i=0}^{h-1} \mathbf{1}_{\{u \le \alpha(i)\}} = \alpha^{-1}(u) \wedge h$. ∎

Lemma 2.29 *Let X_t be a time series with α-mixing coefficients $\alpha(h)$ and let Y and Z be random variables that are measurable relative to $\sigma(\ldots, X_{-1}, X_0)$ and $\sigma(X_h, X_{h+1}, \ldots)$, respectively, for a given $h \ge 0$. Then,*

$$\bigl|\mathrm{cov}(Y, Z)\bigr| \le 2\int_0^{\alpha(h)} F_{|Y|}^{-1}(1-u)F_{|Z|}^{-1}(1-u)\,du.$$

Proof By the definition of the mixing coefficients, we have, for every $y, z > 0$,

$$\big|\mathrm{cov}(\mathbf{1}_{\{Y^+>y\}}, \mathbf{1}_{\{Z^+>z\}})\big| \leq \frac{1}{2}\alpha(h).$$

The same inequality is valid with Y^+ and/or Z^+ replaced by Y^- and/or Z^-. It follows that

$$\big|\mathrm{cov}(\mathbf{1}_{\{Y^+>y\}} - \mathbf{1}_{\{Y^->y\}}, \mathbf{1}_{\{Z^+>z\}} - \mathbf{1}_{\{Z^->z\}})\big| \leq 2\alpha(h).$$

Because $\big|\mathrm{cov}(U,V)\big| \leq 2(E|U|)\|V\|_\infty$ for any pair of random variables U, V (the simplest Hölder inequality), we obtain that the covariance on the left side of the preceding display is also bounded by $2\big(P(Y^+ > y) + P(Y^- > y)\big)$. Yet another bound for the covariance is obtained by interchanging the roles of Y and Z. Combining the three inequalities, we see that, for any $y, z > 0$,

$$\begin{aligned}\big|\mathrm{cov}(\mathbf{1}_{\{Y^+>y\}} - \mathbf{1}_{\{Y^->y\}}, \mathbf{1}_{\{Z^+>z\}} - \mathbf{1}_{\{Z^->z\}})\big| &\leq 2\alpha(h) \wedge 2P(|Y| > y) \wedge 2P(|Z| > z) \\ &= 2\int_0^{\alpha(h)} \mathbf{1}_{\{1-F_{|Y|}(y)>u\}}\, \mathbf{1}_{\{1-F_{|Z|}(z)>u\}}\, du.\end{aligned}$$

Next we write $Y = Y^+ - Y^- = \int_0^\infty (\mathbf{1}_{\{Y^+>y\}} - \mathbf{1}_{\{Y^->y\}})\, dy$ and similarly for Z, to obtain, by Fubini's theorem,

$$\begin{aligned}\big|\mathrm{cov}(Y,Z)\big| &= \Big|\int_0^\infty\int_0^\infty \mathrm{cov}(\mathbf{1}_{\{Y^+>y\}} - \mathbf{1}_{\{Y^->y\}}, \mathbf{1}_{\{Z^+>z\}} - \mathbf{1}_{\{Z^->z\}})\, dy\, dz\Big| \\ &\leq 2\int_0^\infty\int_0^\infty\int_0^{\alpha(h)} \mathbf{1}_{\{F_{|Y|}(y)<1-u\}}\, \mathbf{1}_{\{F_{|Z|}(z)<1-u\}}\, du\, dy\, dz.\end{aligned}$$

Any pair of a distribution and a quantile function satisfies $F_X(x) < u$ if and only if $x < F_X^{-1}(u)$, for every x and u. We can conclude the proof of the lemma by another application of Fubini's theorem. □

2.11 Bibliographic notes

Rubin and Tucker (1959) have considered nonparametric estimation for general Lévy processes, based on both continuous and discrete time observations, and Basawa and Brockwell (1982) have considered estimation for the subclass of continuously observed increasing Lévy processes. In this chapter we have considered indirect estimation through the observation of the Ornstein-Uhlenbeck process X at discrete time instants. Thus we have dealt with an inverse problem, and correspondingly our estimation techniques are quite different from the ones in these papers. Another paper on estimation for OU-processes is Roberts et al. (2004), in which Bayesian estimation for parametric models is considered.

The idea of estimating the canonical function of a positive self-decomposable random variable has its origin in Korsholm (1999), Chapter 5.

Chapter 3

Parametric estimation for subordinators and induced OU-processes

In the previous chapter we introduced a nonparametric estimation technique for estimating the Lévy density of the invariant law of a stationary Ornstein-Uhlenbeck process that is driven by a subordinator. Although we were able to prove consistency, we did not address convergence rates, which seems to be a difficult problem. By restricting our model to a parametric one, this problem should become feasible. Indeed, in this chapter we will show that if the Lévy density is parametrized by a Euclidean parameter, the cumulant M-estimator converges at rate $n^{-1/2}$. Moreover, the difference between the estimator $\hat{\theta}_n$ and the true parameter, multiplied by $\sqrt{n}$, converges to a Normal distribution.

The fact that the canonical function is decreasing does not play a significant role for the parametric estimator. Therefore, we can state the estimation problem in a slightly broader context, without introducing extra difficulties. We assume that we have observations (possibly dependent, as in the case of discrete-time observations of an OU-process) from a probability distribution π which is infinitely divisible on $\mathbb{R}_+$ with Lévy measure that admits a density a_θ with respect to Lebesgue measure. Here $\theta \in \Theta \subseteq \mathbb{R}^k$ denotes an unknown parameter, which we wish to estimate.

Here is an outline of the contents of this chapter. In the next section, we start with a precise definition of the model, observation scheme and estimator. Sufficient conditions for consistency are given in Section 3.2. In Section 3.3 we analyze the asymptotic behavior of the cumulant M-estimator (CME), in case the preliminary estimator is the empirical characteristic function and all observations are independent. The results are illustrated in Section 3.4 for a discretely observed Gamma Lévy process. In Section 3.5 we move to the case of dependent observations from a Lévy driven OU-process. To illustrate the results, we work out the computations involved for the Inverse-Gaussian OU-process. Some technical results are collected in the appendix.

3.1 Definition of the estimator

We have the following setup:

Data Let Z be a subordinator without drift term. It is assumed that Z is observed at equally spaced times $t_n^k = k\Delta$ ($k = 0, \dots, n-1$), so that the data consist of the time series $\{Z_{k\Delta}\}_{k=0}^{n-1}$. Equivalently, since Z has stationary and independent increments, we observe an i.i.d. sample $X_i = Z_{i\Delta} - Z_{(i-1)\Delta}$. Without loss of generality we take $\Delta = 1$.

Statistical model The process Z is parametrized by its Lévy measure ν_{θ_0} which admits a density a_{θ_0} with respect to Lebesgue measure. Here $\theta_0 \in \Theta \subseteq \mathbb{R}^k$ is an unknown parameter, which we wish to estimate. The invariant probability distribution of X_1, denoted by π_{θ_0}, has characteristic function ψ_{θ_0}, which takes the form

$$\psi_{\theta_0}(t) = \int e^{itx}\pi_{\theta_0}(dx) = \exp\left(\int_0^\infty (e^{itx} - 1)a_{\theta_0}(x)dx\right), \quad t \in \mathbb{R}.$$

In Section 3.5, we extend this setup to allow for dependent data. To get a well-defined cumulant M-estimator, we assume (as before)

(i) there exists a preliminary estimator $\tilde\psi_n$ for ψ_0 such that (2.3) holds.

(ii) w is an integrable, symmetric, compactly supported weight function that is strictly positive in a neighborhood of the origin.

(iii) $\tilde g_n = \log \tilde\psi_n$ is well defined on the support of w.

Let $\mathcal{L}^2(w)$ be the space of square integrable functions w.r.t. $w(t)dt$ with semi inner-product $\langle \cdot, \cdot \rangle_w$ given by $\langle f, g\rangle_w = \Re \int f(t)\overline{g(t)}w(t)dt$ ($f, g \in \mathcal{L}^2(w)$). Denote the cumulant function corresponding to ψ_θ by g_θ. Given $\tilde g_n$, we define an estimator for θ_0 as the minimizer of

$$\Gamma_n(\theta) := \|g_\theta - \tilde g_n\|_w^2 = \int |g_\theta(t) - \tilde g_n(t)|^2 w(t)dt$$

over $\Theta \subseteq \mathbb{R}^k$. Put $M_n(\theta) := \sqrt{\Gamma_n(\theta)}$. Recall that S denotes the support of the weight-function w. In the following, we write $\|\cdot\|_\infty$ for the supremum norm on S, i.e. for $f : \mathbb{R} \to \mathbb{C}$ bounded, we define $\|f\|_\infty = \sup_{t\in S} |f(t)|$.

Proposition 3.1 *Assume the parameter-set $\Theta \subseteq \mathbb{R}^k$ is compact. If for every sequence $(\theta_m) \in \Theta$*

$$\theta_m \to \theta \quad \Rightarrow \int_0^\infty (x \wedge 1) \cdot |a_{\theta_m}(x) - a_\theta(x)|dx \to 0, \quad (m \to \infty) \tag{3.1}$$

then there is a $\hat\theta_n \in \Theta$ such that $M_n(\hat\theta_n) = \min_{\theta\in\Theta} M_n(\theta)$.

Proof It suffices to prove that the mapping $\theta \mapsto M_n(\theta)$ is continuous. Let $\theta, \xi \in \Theta$. For $t \in S$,

$$\begin{aligned}
|g_\theta(t) - g_\xi(t)| &= \Big|\int_0^\infty (e^{itx} - 1)(a_\theta(x) - a_\xi(x))dx\Big| \\
&\le |t| \int_0^1 x\,|a_\theta(x) - a_\xi(x)|dx + 2\int_1^\infty |a_\theta(x) - a_\xi(x)|dx \\
&\le \max\{|t|, 2\} \int_0^\infty (x \wedge 1)|a_\theta(x) - a_\xi(x)|dx,
\end{aligned}$$

where we use the inequality $|e^{ix}-1| \leq \min\{|x|,2\}$. Next, by the triangle inequality, we find that

$$\begin{aligned}
|M_n(\theta)-M_n(\xi)| &= \big|\|g_\theta-\tilde{g}_n\|_w-\|g_\xi-\tilde{g}_n\|_w\big| \leq \|g_\theta-g_\xi\|_w \\
&\leq \left(\|g_\theta-g_\xi\|_\infty^2 \int w(t)dt\right)^{1/2} \\
&\leq C\int_0^\infty (x\wedge 1)\cdot|a_\theta(x)-a_\xi(x)|dx,
\end{aligned}$$

for some constant $C>0$. Now, continuity of M_n on Θ is a consequence of assumption (3.1). $\square$

By a dominated convergence argument, it is often easy to check (3.1).

Corollary 3.2 *Assume the parameter-set $\Theta \subseteq \mathbb{R}^k$ is compact. Suppose there exists a function $A:(0,\infty)\to[0,\infty)$ satisfying $\int_0^\infty (x\wedge 1)A(x)dx<\infty$ (e.g. $A(x)=1/(x\sqrt{x})$) such that $\sup_{\theta\in\Theta} a_\theta(x)\leq A(x)$, for all $x>0$. If the mapping $\theta\mapsto a_\theta(x)$ is continuous for all $x>0$, then $\hat{\theta}_n=\operatorname{argmin}_{\theta\in\Theta}\Gamma_n(\theta)$ exists.*

Remark 3.3 Even if the Lévy measure ν does not admit a density with respect to Lebesgue measure we may still prove existence of a CME. For example, let Z be a Poisson process with jumps of size $\theta\in\Theta\subseteq\mathbb{R}$. Then ν_θ is a point-mass at θ and hence

$$|g_\theta(t)-g_\xi(t)| = |e^{i\theta t}-e^{i\xi t}| \leq |t|\cdot|\theta-\xi|,$$

which implies that we can adapt the proof of Proposition 3.1 to this case.

3.2 Consistency

Under appropriate conditions, we expect the random criterion functions M_n to converge to a deterministic function $M:\Theta\to[0,\infty]$. Since roughly, $\tilde{g}_n \approx g_{\theta_0}$ for n large, we expect, under θ_0, that

$$M_n(\theta) \xrightarrow{\text{a.s.}} M(\theta) := \|g_\theta - g_{\theta_0}\|_w.$$

The latter deterministic map is easily seen to be minimized for $\theta=\theta_0$.

Although $\hat{\theta}_n$ is possibly not uniquely defined, the next theorem shows that any choice of $\hat{\theta}_n$ (as a minimizer of M_n) is a consistent estimator for θ_0 under mild conditions. We write d for Euclidean distance on $\mathbb{R}^k$.

Theorem 3.4 *Suppose $\Theta\subseteq\mathbb{R}^k$ is compact and (3.1) holds. Assume the sequence of preliminary estimators is such that (2.3) holds (for convergence almost surely). If $\pi_\theta\neq\pi_{\theta_0}$ whenever $\theta\neq\theta_0$, then, $\hat{\theta}_n\to\theta_0$, almost surely, as n tends to infinity.*

Proof By Lemma 2.8, we can strengthen the convergence in (2.3) to $\sup_{t\in S}|\tilde{\psi}_n(t)-\psi_{\theta_0}(t)| \xrightarrow{\text{a.s.}} 0$ under θ_0. Theorem 7.6.3 in Chung (2001) implies that the uniform convergence of $\tilde{\psi}_n$ to ψ_{θ_0} on S carries over to uniform convergence of $\tilde{g}_n$ to g_{θ_0} on S. That is,

$\|\tilde{g}_n - g_{\theta_0}\|_\infty \xrightarrow{\text{a.s.}} 0$. By the triangle inequality, we have

$$\begin{aligned}\sup_{\theta\in\Theta}|M_n(\theta) - M(\theta)| &= \sup_{\theta\in\Theta}|\|g_\theta - \tilde{g}_n\|_w - \|g_\theta - g_{\theta_0}\|_w| \\ &\leq \|\tilde{g}_n - g_{\theta_0}\|_w \leq \|\tilde{g}_n - g_{\theta_0}\|_\infty \Big(\int w(t)dt\Big)^{1/2} \xrightarrow{\text{a.s.}} 0.\end{aligned}$$

Furthermore, since $\hat{\theta}_n$ minimizes M_n over Θ,

$$M_n(\hat{\theta}_n) \leq M_n(\theta_0) = \|g_{\theta_0} - \tilde{g}_n\|_w \xrightarrow{\text{a.s.}} 0.$$

Combining both assertions, we conclude that $M(\hat{\theta}_n) \xrightarrow{\text{a.s.}} 0$ $(n \to \infty)$. Once we have proved

$$\forall\, \varepsilon > 0 \qquad \inf_{\theta\,:\,d(\theta,\theta_0)\geq\varepsilon} M(\theta) > M(\theta_0) = 0, \tag{3.2}$$

this implies that $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$, as $n \to \infty$.

We now prove (3.2). Suppose $\theta \neq \theta_0$, then by identifiability $\pi_\theta \neq \pi_{\theta_0}$. Using Lemma 2.26, this implies that there exists a neighborhood $U \subseteq S$ around zero, such that $\psi_\theta \not\equiv \psi_{\theta_0}$ on U. Therefore $g_\theta \not\equiv g_{\theta_0}$ on U and hence, by the continuity of the cumulant function, $M(\theta) = \|g_\theta - g_{\theta_0}\|_w > 0$, since we have $U \subseteq S$. Thus $M(\theta) > M(\theta_0) = 0$. Therefore M has a *unique* minimizer on Θ.

By similar inequalities as in the proof of Lemma 3.1, for $\theta, \xi \in \Theta$,

$$|M(\theta) - M(\xi)| \leq \|g_\theta - g_\xi\|_w \leq C\int_0^\infty (x \wedge 1)|a_\theta - a_\xi|dx,$$

for some positive constant C. Hence, by (3.1), the mapping $\theta \mapsto M(\theta)$ is continuous on Θ. Let $\varepsilon > 0$. The set $V_{\theta_0} := \{\theta \in \Theta \ : \ d(\theta, \theta_0) \geq \varepsilon\}$ is a closed subset of the compact set Θ and hence compact. By continuity, $M(V_{\theta_0})$, the image of V_{θ_0} under M, is compact. Since θ_0 is the *unique* minimizer of M on Θ and $M(\theta_0) = 0$, $0 \notin M(V_{\theta_0})$. Therefore, there exists an $\eta > 0$ such that $M(\theta) \geq \eta$ on V_{θ_0}. This proves (3.2). $\square$

In case we assume (2.3) holds with convergence in probability instead of convergence almost surely and retain the other assumptions of Theorem 3.4, the proof of the previous theorem can easily be adapted to show that in that case $\sup_{\theta\in\Theta}|M_n(\theta) - M(\theta)| \xrightarrow{\text{p}} 0$. Since (3.2) remains true, we can apply Theorem 5.7 in Van der Vaart (1998), which then asserts that any sequence of estimators $\bar{\theta}_n$ with $M_n(\bar{\theta}_n) \leq M_n(\theta_0) + o_P(1)$ converges in probability to θ_0. Now

$$|M_n(\hat{\theta}_n) - M_n(\theta_0)| \leq M_n(\hat{\theta}_n) + |M(\theta_0) - M_n(\theta_0)| = o_P(1) + o_P(1),$$

where the two $o_P(1)$ terms follow from the inequalities in the first two displays of the proof of Theorem 3.4. Therefore, $\hat{\theta}_n \xrightarrow{\text{p}} \theta_0$.

Corollary 3.5 *Suppose $\Theta \subseteq \mathbb{R}^k$ is compact and (3.1) holds. Assume the sequence of preliminary estimators satisfies (2.3) for convergence in probability. If $\pi_\theta \neq \pi_{\theta_0}$ whenever $\theta \neq \theta_0$, then, $\hat{\theta}_n \xrightarrow{\text{p}} \theta_0$, as n tends to infinity.*

3.3 Asymptotic behavior of the estimator

To derive asymptotic distribution results for the CME, we go through the following steps:

(I) Derive weak convergence of the process $(\sqrt{n}(\tilde{\psi}_n(t) - \psi_{\theta_0}(t)), t \in S)$.

(II) Show Hadamard differentiability of the mapping that attaches to a characteristic function its cumulant. Subsequently, apply the functional Delta method to obtain weak convergence of the process $(\sqrt{n}(\tilde{g}_n(t) - g_{\theta_0}(t)), t \in S)$

(III) Use results from the theory of M- and Z-estimators to derive the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Details will be given along the way.

To simplify results a bit, we assume in this section that our initial estimator is the empirical characteristic function, which we denote by $\tilde{\psi}_n$ (see however Remark 3.16). Throughout this section, we assume all observations are independent, with common distribution π_{θ_0}.

Step *(I)*

Define for $t \in \mathbb{R}$

$$Y_n(t) = \sqrt{n}(\tilde{\psi}_n(t) - \psi_{\theta_0}(t)).$$

(We suppress the dependence of Y_n on θ_0 for ease of notation.)

By the multivariate central limit theorem it follows that for $t_1 < \cdots < t_k$,

$$(Y_n(t_1), \dots, Y_n(t_k)) \rightsquigarrow N_k(0, \Sigma^{\theta_0}), \quad n \to \infty,$$

where, if X is a random variable with distribution π_{θ_0},

$$\Sigma^{\theta_0}|_{j,k} = E(e^{it_jX} - \psi_{\theta_0}(t_j))\overline{(e^{it_kX} - \psi_{\theta_0}(t_k))} = \psi_{\theta_0}(t_j - t_k) - \psi_{\theta_0}(t_j)\psi_{\theta_0}(-t_k). \tag{3.3}$$

We aim to prove that there exists a centered Gaussian process Y with covariance as in (3.3) such that Y_n converges weakly to Y in the space $\ell^\infty(S)$, where

$$\ell^\infty(S) = \{z : S \to \mathbb{C} \ : \ \|z\|_\infty = \sup_{t \in S} |z(t)| < \infty\},$$

is the space of bounded complex-valued functions on S, equipped with the supremum norm. By weak convergence we mean

$$E^* h(Y_n) \to Eh(Y), \quad n \to \infty$$

for every bounded and continuous function $h : \ell^\infty(S) \to \mathbb{R}$. (We use outer expectations, since elements of $\ell^\infty(S)$ may not be Borel-measurable. See for instance Van der Vaart (1998), Chapter 18.)

The result below is taken from Giné and Zinn (1986). The ε-covering number of a set A for a semi-metric ρ, denoted by $N(\varepsilon, A, \rho)$, is defined as the minimal number of ρ-balls of radius ε needed to cover the set A.

Theorem 3.6 (Giné and Zinn (1986), Chapter 4, Theorem 6.1) *Let X be a random variable with distribution π_{θ_0}. Define for $s, t \in \mathbb{R}$*

$$\sigma(s,t) := \Big(E|e^{itX} - e^{isX}|^2\Big)^{1/2} = 2\Big(E\big[\sin^2 \tfrac{1}{2}(t-s)X\big]\Big)^{1/2}.$$

If

$$\int_0^{\varepsilon_0} (\log N(\varepsilon, S, \sigma))^{1/2} d\varepsilon < \infty, \tag{3.4}$$

for some small $\varepsilon_0 > 0$, then Y_n converges weakly in the space $\ell^\infty(S)$ to a complex-valued centered Gaussian process Y' with covariance structure

$$cov(Y'_t, Y'_s) = \psi_{\theta_0}(t-s) - \psi_{\theta_0}(t)\psi_{\theta_0}(-s). \tag{3.5}$$

Let B denote a standard Brownian Bridge and define a process B^0 by

$$B^0(x) = B(\pi_{\theta_0}([0,x])), \quad x \geq 0. \tag{3.6}$$

The limit process Y' in the previous theorem can be identified as the process Y, defined by

$$Y(t) := \int e^{itx} dB^0(x), \quad t \in S. \tag{3.7}$$

To see this, note that Y is a centered Gaussian process with the same covariance function as Y. Since weak limits are unique in a distributional sense, it follows that Y_n converges weakly in $\ell^\infty(S)$ to Y.

Since $F := (C(S), \|\cdot\|_\infty)$ is a closed subset of $(\ell^\infty(S), \|\cdot\|_\infty)$, we have by the Portmanteau theorem

$$P(Y \in F) \geq \limsup_n P(Y_n \in F) = 1.$$

Therefore Y is almost surely continuous and by Lemma 18.13 in Van der Vaart (1998) the weak convergence of Y_n of Y in space $\ell^\infty(S)$ carries over to the space $C(S)$. Hence, $Y_n \rightsquigarrow Y$ in $(C(S), \|\cdot\|_\infty)$. We summarize this result in a corollary.

Corollary 3.7 *If the entropy condition (3.4) holds true, then Y_n converges weakly in $(C(S), \|\cdot\|_\infty)$ to the centered Gaussian process Y, as defined in (3.7).*

A simple moment condition suffices for (3.4).

Lemma 3.8 *If for some $\alpha > 0$*

$$\int |x|^\alpha \pi_{\theta_0}(dx) < \infty, \tag{3.8}$$

then the entropy condition (3.4) is satisfied and the conclusion of Theorem 3.6 holds.

A proof is given in the appendix to this chapter, Section 3.6.

Step *(II)*

Next, we consider the weak convergence of the empirical cumulant process, defined by

$$U_n(t) := \sqrt{n}(\tilde{g}_n(t) - g_{\theta_0}(t)).$$

We use the functional Delta method (see e.g. Van der Vaart (1998), Chapter 20). Let Ψ denote the set of infinitely divisible characteristic functions and G denote the corresponding set of cumulants. Recall that $\|\cdot\|_\infty$ stands for the supremum norm on S.

Proposition 3.9 *The mapping $T : (\Psi, \|\cdot\|_\infty) \to (G, \|\cdot\|_\infty)$, mapping $\psi \in \Psi$ to its cumulant, is Hadamard differentiable and its Hadamard derivative at ψ is given by $T'_\psi(\varphi) = \varphi/\psi$ for $\varphi \in C^0 := \{f : \mathbb{R} \to \mathbb{C} \mid f$ is continuous and $f(0) = 0\}$.*

Proof Fix $\psi \in \Psi$ and $\varphi \in C^0$. Let φ_ε be such that $\psi + \varepsilon\varphi_\varepsilon \in \Psi$ for all small $\varepsilon > 0$ and such that $\|\varphi_\varepsilon - \varphi\|_\infty \to 0$, as $\varepsilon \downarrow 0$. By the triangle inequality

$$\left\|\frac{T(\psi+\varepsilon\varphi_\varepsilon)-T(\psi)}{\varepsilon} - \frac{\varphi}{\psi}\right\|_\infty \leq \left\|\frac{T(\psi+\varepsilon\varphi_\varepsilon)-T(\psi)}{\varepsilon} - \frac{\varphi_\varepsilon}{\psi}\right\|_\infty + \left\|\frac{\varphi-\varphi_\varepsilon}{\psi}\right\|_\infty. \tag{3.9}$$

By the defining property of the distinguished logarithm, we have for any $\psi_1, \psi_2 \in \Psi$ that $T(\psi_1) - T(\psi_2) = T(\psi_1/\psi_2)$. Therefore, the first term on the right-hand-side of the preceding display equals

$$\frac{1}{\varepsilon}\left\|T\left(1+\varepsilon\frac{\varphi_\varepsilon}{\psi}\right) - \varepsilon\frac{\varphi_\varepsilon}{\psi}\right\|_\infty = \frac{1}{\varepsilon}\sup_{t\in S}\left|T\left(1+\varepsilon\frac{\varphi_\varepsilon}{\psi}\right)(t) - \varepsilon\frac{\varphi_\varepsilon(t)}{\psi(t)}\right|. \tag{3.10}$$

Define $u_\varepsilon(t) := \frac{\varphi_\varepsilon(t)}{\psi(t)}$. Since $\{|u_\varepsilon(t)|, t \in S\}$ is compact, we can choose ε small enough such that $\varepsilon u_\varepsilon(t) \in B_{1/2}(0) = \{z \in \mathbb{C} \ : \ |z| \leq 1/2\}$ for all $t \in S$. Within $B_{1/2}(0)$ the following inequality holds: $|\log(1+w) - w| \leq \frac{1}{2}|w|^2/(1-|w|)$ (see Remmert (1991), Section 5.4). Using this inequality we see that (3.10) is bounded by

$$\frac{1}{\varepsilon}\sup_{t\in S}\left|\frac{1}{2}\frac{\varepsilon^2(u_\varepsilon(t))^2}{1-\varepsilon u_\varepsilon(t)}\right| \leq \frac{1}{2\varepsilon}\frac{\varepsilon^2 M_\varepsilon^2}{1-\varepsilon M_\varepsilon} \leq \frac{1}{2}M_\varepsilon^2\varepsilon,$$

where $M_\varepsilon = \|u_\varepsilon\|_\infty$. Now $M_\varepsilon \leq \left\|\frac{\varphi-\varphi_\varepsilon}{\psi}\right\|_\infty + \left\|\frac{\varphi}{\psi}\right\|_\infty$. We conclude that once we have proved that $\left\|\frac{\varphi-\varphi_\varepsilon}{\psi}\right\|_\infty \to 0$ as $\varepsilon \downarrow 0$ and $\left\|\frac{\varphi}{\psi}\right\|_\infty < \infty$, then both terms on the right-hand-side of (3.9) can be made arbitrarily small by letting ε tend to zero.

Since an infinitely divisible characteristic function has no zeros, every $\psi \in \Psi$ is bounded away from zero on compacta. Therefore $\|\frac{1}{\psi}\|_\infty = \sup_{t\in S}\left|\frac{1}{\psi(t)}\right| \leq C$ for some positive constant C and

$$\left\|\frac{\varphi-\varphi_\varepsilon}{\psi}\right\|_\infty \leq \left\|\frac{1}{\psi}\right\|_\infty \|\varphi_\varepsilon - \varphi\|_\infty \leq C\|\varphi_\varepsilon - \varphi\|_\infty,$$

which tends to zero as ε tends to zero. The same argument shows that $\left\|\frac{\varphi}{\psi}\right\|_\infty \leq C\|\varphi\|_\infty < \infty$. Hence, the left-hand-side of (3.9) can be made arbitrarily small, which means that T is Hadamard differentiable at ψ with T'_ψ as stated. $\square$

By the functional Delta method (Van der Vaart (1998), Theorem 20.8) we now obtain

Corollary 3.10 *If $Y_n \rightsquigarrow Y$ in $(C(S), \|\cdot\|_\infty)$, then $U_n \rightsquigarrow U$ in $(C(S), \|\cdot\|_\infty)$, where*

$$U = T'_\psi(Y) = \frac{Y}{\psi_{\theta_0}}.$$

The process U is centered Gaussian with covariance function

$$cov(U(t), U(s)) = [\psi_{\theta_0}(t-s) - \psi_{\theta_0}(t)\psi_{\theta_0}(-s)]/[\psi_{\theta_0}(t)\psi_{\theta_0}(-s)], \quad t, s \in S.$$

In the following U_n and U are always assumed to depend on θ_0, though this dependence is suppressed in the notation.

Step *(III)*

To derive the asymptotic behavior of $\hat{\theta}_n$, we define conditions which enable us to define $\hat{\theta}_n$ as a Z-estimator. By this we mean that we define conditions under which $\hat{\theta}_n$ is a point for which all partial derivatives of $\theta \mapsto \Gamma_n(\theta)$ are (nearly) zero. The following assumption suits our purposes well. We denote by θ_i the i-th coordinate of a vector $\theta \in \Theta \subseteq \mathbb{R}^k$.

Assumption 3.11

(i) Condition (3.4) is satisfied, so that the process Y_n converges weakly and, by Corollary 3.10, the process U_n as well.

(ii) For each $i = 1, \ldots, k$ the partial derivative $\theta \mapsto g^i_\theta := \frac{\partial}{\partial \theta_i} g_\theta$ exists. Moreover, the mapping $(\theta, t) \mapsto g^i_\theta(t)$ is jointly continuous on $\Theta \times S$.

Combining the second part of this assumption with dominated convergence we get

$$\Psi^i_n(\theta) := \frac{\partial}{\partial \theta_i} \Gamma_n(\theta) = 2\langle g^i_\theta, g_\theta - \tilde{g}_n \rangle_w, \quad 1 \le i \le k.$$

Hence, $\hat{\theta}_n$ is a (near) zero of the random criterion function

$$\Psi_n(\theta) := (\Psi^1_n(\theta), \ldots, \Psi^k_n(\theta))',$$

Since $\tilde{g}_n$ converges almost surely uniformly on S to g_{θ_0}, we expect that $\hat{\theta}_n$ converges to a zero of

$$\Psi(\theta) := (2\langle g^1_\theta, g_\theta - g_{\theta_0} \rangle_w, \ldots, 2\langle g^k_\theta, g_\theta - g_{\theta_0} \rangle_w)'.$$

Note that θ_0 is such a zero. Define $\mathbb{H}_n : \Theta \subseteq \mathbb{R}^k \to \mathbb{R}^k$ by

$$\begin{aligned} \mathbb{H}_n(\theta) :&= \sqrt{n}(\Psi_n(\theta) - \Psi(\theta)) \\ &= -2(\langle g^1_\theta, \sqrt{n}(\tilde{g}_n - g_{\theta_0}) \rangle_w, \ldots, -2\langle g^k_\theta, \sqrt{n}(\tilde{g}_n - g_{\theta_0}) \rangle_w)' \\ &= -2(\langle g^1_\theta, U_n \rangle_w, \ldots, \langle g^k_\theta, U_n \rangle_w)'. \end{aligned}$$

Denote the i-th coordinate of $\mathbb{H}_n$ by $\mathbb{H}^i_n$. For $\theta \in \Theta$, define $z_\theta : \mathbb{R}^2 \to \mathbb{C}$ by

$$z_\theta(t, x) := \frac{e^{itx}}{\psi_\theta(t)}. \tag{3.11}$$

Lemma 3.12 *For each $\theta \in \Theta$ the sequence of random vectors $\{\mathbb{H}_n(\theta)\}_n$ converges weakly in $\mathbb{R}^k$ to the random vector*

$$\mathbb{H}(\theta) := -2(\langle g_\theta^1, U\rangle_w, \ldots, \langle g_\theta^k, U\rangle_w)'.$$

Moreover, the vector $\mathbb{H}(\theta)$ has a mean zero multivariate Normal distribution with covariance matrix Σ^θ that is given by

$$\begin{aligned}\Sigma^\theta|_{i,j} &= 4\int \langle g_\theta^i, z_{\theta_0}(\cdot, x)\rangle_w \langle g_\theta^j, z_{\theta_0}(\cdot, x)\rangle_w d\pi_{\theta_0}(x) \\ &\quad -4\left(\int \langle g_\theta^i, z_{\theta_0}(\cdot, x)\rangle_w d\pi_{\theta_0}(x)\right)\left(\int \langle g_\theta^j, z_{\theta_0}(\cdot, x)\rangle_w d\pi_{\theta_0}(x)\right). \end{aligned} \tag{3.12}$$

The proof is given in Section 3.6.

In the following, we use stochastic order symbols: for a sequence of random vectors X_n and a given sequence of random variables R_n we write $X_n = o_P(R_n)$ if $X_n = Y_n R_n$ for a sequence of random vectors Y_n tending to zero in probability.

Theorem 3.13 *Assume the mapping $\theta \mapsto \Psi(\theta)$ is differentiable at θ_0 with non-singular derivative matrix $\dot\Psi(\theta_0)$. Assume*

$$\mathbb{H}_n(\hat\theta_n) - \mathbb{H}_n(\theta_0) \xrightarrow{\text{p}} 0. \tag{3.13}$$

If $\hat\theta_n \xrightarrow{\text{p}} \theta_0$ and $\Psi_n(\hat\theta_n) = o_P(n^{-1/2})$ ($\hat\theta_n$ is a near zero of Ψ_n), then

$$\sqrt{n}(\hat\theta_n - \theta_0) = -\dot\Psi^{-1}(\theta_0)\mathbb{H}_n(\theta_0) + o_P(1). \tag{3.14}$$

Moreover, the sequence $\sqrt{n}(\hat\theta_n - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $\dot\Psi(\theta_0)^{-1}\Sigma^{\theta_0}(\dot\Psi(\theta_0)^{-1})'$, where Σ^{θ_0} is given in (3.12).

Proof The proof follows the line of thought as in Theorem 5.21 of Van der Vaart (1998). First note that

$$\begin{aligned}\mathbb{H}_n(\hat\theta_n) &= \sqrt{n}\Psi_n(\hat\theta_n) + \sqrt{n}(\Psi(\theta_0) - \Psi(\hat\theta_n)) - \sqrt{n}\Psi(\theta_0) \\ &= \sqrt{n}((\Psi(\theta_0) - \Psi(\hat\theta_n)) + o_P(1). \end{aligned} \tag{3.15}$$

Since $\theta \mapsto \Psi(\theta)$ is differentiable near θ_0, we have for $h \in \mathbb{R}^k$, that $R(h) := \Psi(\theta_0 + h) - \Psi(\theta_0) - \dot\Psi(h)$ satisfies $R(h) = o(\|h\|)$, as $h \to 0$. Since $\hat\theta_n - \theta_0 = o_P(1)$ this implies (Van der Vaart (1998), Lemma 2.12)

$$\sqrt{n}(\Psi(\hat\theta_n) - \Psi(\theta_0)) - \dot\Psi(\theta_0)\sqrt{n}(\hat\theta_n - \theta_0) = \sqrt{n}o_P(\|\hat\theta_n - \theta_0\|).$$

Inserting (3.15) into this expression and using (3.13) gives

$$\sqrt{n}\dot\Psi(\theta_0)(\hat\theta_n - \theta_0) + \sqrt{n}o_P(\|\hat\theta_n - \theta_0\|) = -\mathbb{H}_n(\hat\theta_n) + o_P(1) = -\mathbb{H}_n(\theta_0) + o_P(1). \tag{3.16}$$

By Lemma 3.12, the sequence $\{\mathbb{H}_n(\theta_0)\}_n$ is tight. By invertibility of the matrix $\dot\Psi_{\theta_0}$ we obtain from (3.16)

$$\sqrt{n}\|\hat\theta_n - \theta_0\| \leq \|\dot\Psi^{-1}\|\sqrt{n}\|\hat\theta_n - \theta_0\| = O_P(1) + o_P(\sqrt{n}\|\hat\theta_n - \theta_0\|).$$

This implies that $\hat{\theta}_n$ is $\sqrt{n}$-consistent. Inserting this in (3.16), we obtain that $\sqrt{n}\dot{\Psi}(\hat{\theta}_n - \theta_0) = -\mathbb{H}_n(\theta_0) + o_P(1)$. Now multiplication of both sides with $\dot{\Psi}^{-1}$ gives (3.14). (The remainder term still converges to zero in probability, since matrix multiplication is a continuous operation.)

The second assertion follows directly from Lemma 3.12.

□

Remark 3.14 Under regularity conditions,

$$\dot{\Psi}(\theta)|_{i,j} = \frac{\partial}{\partial \theta_j}\Psi^i(\theta) = 2\langle g_\theta^{ji}, g_\theta - g_{\theta_0}\rangle_w + 2\langle g_\theta^i, g_\theta^j\rangle_w,$$

where $g_\theta^{ji} = \frac{\partial}{\partial\theta_j\partial\theta_i} g_\theta$. In that case $\dot{\Psi}(\theta_0)|_{i,j} = 2\langle g_{\theta_0}^i, g_{\theta_0}^j\rangle_w$.

It remains to give conditions under which (3.13) holds.

Lemma 3.15 *Assume Θ is compact. If $\hat{\theta}_n$ is a sequence in Θ such that $\hat{\theta}_n \xrightarrow{\text{p}} \theta_0$, then $\mathbb{H}_n(\hat{\theta}_n) - \mathbb{H}_n(\theta_0) \xrightarrow{\text{p}} 0$.*

Proof Since convergence in probability of a vector is equivalent to convergence in probability of each of its components, it suffices to prove $\mathbb{H}_n^i(\hat{\theta}_n) - \mathbb{H}_n^i(\theta_0) \xrightarrow{\text{p}} 0$ for each $i \in \{1, \dots, k\}$. By Lemma 3.23 it is enough to show that almost all sample paths $\theta \mapsto \mathbb{H}^i(\theta)$ are continuous on Θ and that $\mathbb{H}_n^i \rightsquigarrow \mathbb{H}^i$ in $\ell^\infty(\Theta)$.

Let (θ_m) be a sequence in Θ converging to θ. By Assumption 3.11*(ii)*, and compactness of S and Θ, dominated convergence gives that $\|g_{\theta_m}^i - g_\theta^i\|_w \to 0$ $(m \to \infty)$. By the Cauchy-Schwartz inequality

$$|\mathbb{H}^i(\theta_m) - \mathbb{H}^i(\theta)| \leq 2\|g_{\theta_m}^i - g_\theta^i\|_w \|U\|.$$

The right-hand-side of this expression tends to zero, as $m \to \infty$, since almost all sample paths $t \mapsto U(t)$ are continuous on S. This shows continuity of $\theta \mapsto \mathbb{H}^i(\theta)$, a.s.

Let $\|\cdot\|_\Theta$ denote the supremum norm on Θ. Define $\Lambda : C(S, \|\cdot\|_\infty) \to C(\Theta, \|\cdot\|_\Theta)$ by $[\Lambda(f)](\theta) = \langle g_\theta^i, f\rangle_w$. If $f_n \to f$ in $C(S)$, then

$$\|\Lambda(f_n) - \Lambda(f)\|_\Theta = \sup_{\theta\in\Theta} |\langle g_\theta^i, f_n - f\rangle_w| \leq \sup_{\theta\in\Theta} \|g_\theta^i\|_w \|f_n - f\|_w \to 0,$$

which proves continuity of Λ. Since $U_n \rightsquigarrow U$ in $C(S)$, the continuous mapping theorem implies $\Lambda(U_n) \rightsquigarrow \Lambda(U)$ in $C(\Theta)$. That is, $\mathbb{H}_n^i \rightsquigarrow \mathbb{H}^i$ in $C(\Theta)$. □

Remark 3.16 The line of proof given in this section holds for more general preliminary estimators than the empirical characteristic function. We now point out at which places in the proof we use that the preliminary estimator is the empirical characteristic function. Firstly, by Theorem 3.6 it was easy to show that $Y_n \rightsquigarrow Y$. Now suppose we have another sequence of preliminary estimators satisfying $Y_n \rightsquigarrow Y$. By Theorem 3.13, we then obtain that $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}(\theta_0)^{-1}\mathbb{H}(\theta_0)$. For the empirical ch.f. the distribution of $\mathbb{H}(\theta)$ can be found due to the expression for Y in terms of a Brownian bridge, for other preliminary estimators such an expression may not exist. Consequently, the derivation of the distribution of $\mathbb{H}(\theta)$ may be much harder in this case.

Remark 3.17 Ideally, we would like to choose the weight-function w such that the asymptotic variance of $\sqrt{n}\hat{\theta}_n$ (as given in Theorem 3.13) is minimal. However, even if the parameter is one-dimensional, this problem is difficult by the complicated form of the asymptotic variance. Moreover, this optimal weight-function will depend on the unknown parameter that needs to be estimated.

We now point out some related estimation procedures that are based on characteristic functions and cumulants, where similar problems occur. The K-L-procedure of Feuerverger and McDunnough (1981) runs as follows:

(i) Choose a set of points $T = \{t_1, \ldots, t_q\}$. Set

$$\begin{aligned} z_\theta &= (\Re\psi_\theta(t_1), \ldots, \Re\psi_\theta(t_q), \Im\psi_\theta(t_1), \ldots, \Im\psi_\theta(t_q))' \\ z_n &= (\Re\tilde{\psi}_n(t_1), \ldots, \Re\tilde{\psi}_n(t_q), \Im\tilde{\psi}_n(t_1), \ldots, \Im\tilde{\psi}_n(t_q))', \end{aligned}$$

for $\tilde{\psi}_n$ the empirical characteristic function.

(ii) Under conditions, $\sqrt{n}(z_n - z_{\theta_0})$ converges weakly to a Normal distribution with mean zero and covariance-matrix B^{θ_0}. Define $\hat{\theta}_n$ as the maximum-likelihood estimator for this Normal distribution, with B^{θ_0} replaced by a consistent estimator, say B^n. Thus, $\hat{\theta}_n$ is the minimizer of $\theta \mapsto (z_n - z_\theta)'(B^n)^{-1}(z_n - z_\theta)$.

For the one-dimensional case, Feuerverger and McDunnough (1981) point out that the variance of $\sqrt{n}\hat{\theta}_n$ can be made arbitrarily close to the Cramer-Rao-bound, by choosing the grid points in T sufficiently fine and extended. For the multivariate case, which is considered in Feuerverger and McDunnough (1981a) this is not so clear. Most arguments are given in a heuristic way. If we take a regularly spaced grid, i.e. $t_j = j\tau$ for some $\tau > 0$, it is proposed to take τ as the minimizer of the determinant of B^{θ_0}.

Knight and Satchell (1997) follow exactly the same approach as in Feuerverger and McDunnough (1981a), but with the characteristic function replaced by the cumulant function. The same remarks hold for the choice of the weights.

To conclude, since there is no universal rule for choosing the grid-points in the K-L-procedure, as well as the weight-function for the CME, a fair numerical comparison of both models is hard.

3.4 Example: discrete observations from a Gamma-process

In this section we give an example in which we consider existence, consistency and asymptotic behavior of the CME. Furthermore, we explain how the estimator can be approximated numerically.

Suppose we discretely observe a Gamma process. Statistically, this is equivalent to observing a sample $X_1, \ldots, X_n$ with common law $\pi_\theta \sim \text{Gamma}(c, \alpha)$, where $\theta = (c, \alpha) \in (0, \infty)^2$. The cumulant of π_θ is given by

$$g_\theta(t) = -c \log(1 - \alpha^{-1} it).$$

The Lévy density is given by $a_\theta(x) = cx^{-1}e^{-\alpha x}\mathbf{1}_{\{x \geq 0\}}$, which is continuous in θ for all $x > 0$. Let $\Theta \subseteq (0, \infty)^2$ be compact and suppose the true value of the parameter θ_0

is in Θ. If we take a sequence of preliminary estimators which satisfies (2.3), then, by Corollary 3.2, $\hat{\theta}_n = \operatorname{argmin}_{\theta\in\Theta}\Gamma_n(\theta)$ exists. From Theorem 3.4 we obtain consistency: $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$.

Next, we turn attention to the asymptotics of $\hat{\theta}_n$. For the rest of this example, we use the empirical characteristic function as a preliminary estimator. Clearly, condition (3.8) is satisfied. Whence the empirical cumulant process U_n converges weakly. Both partial derivatives of g_θ exist and are given by

$$g^1_\theta(t) = \frac{\partial}{\partial c} g_\theta(t) = -\log(1 - \alpha^{-1} it)$$
$$g^2_\theta(t) = \frac{\partial}{\partial \alpha} g_\theta(t) = -c\frac{-t^2 + i\alpha t}{\alpha(\alpha^2 + t^2)}.$$

Joint continuity of the partial derivatives in (t, c, α) is easily seen. We conclude that assumption (3.11) is fulfilled. Now, we want to apply Theorem 3.13 to obtain asymptotic normality of $\hat{\theta}_n$.

By existence and continuity of all partial derivatives of $\theta \mapsto g_\theta$ it follows that the mapping $\theta \mapsto \Psi(\theta)$ is differentiable near θ_0, for $\theta_0 \in \Theta$. Since also all second order partial derivatives exist and are continuous, we have $\dot{\Psi}(\theta_0)|_{i,j} = 2\langle g^i_{\theta_0}, g^j_{\theta_0}\rangle_w$ (see Remark 3.14). The matrix $\dot{\Psi}(\theta_0)$ is non-singular. To see this: $\det(\dot{\Psi}(\theta_0)) = 4(\|g^1_{\theta_0}\|^2_w \, \|g^2_{\theta_0}\|^2_w - \langle g^1_{\theta_0}, g^2_{\theta_0}\rangle^2_w)$ and this expression equals zero if and only if $g^1_{\theta_0} = a g^2_{\theta_0}$ almost everywhere w.r.t. $w(t)dt$ for some $a \in \mathbb{R}$. By Lemma 3.15, condition (3.13) is easily satisfied. Thus,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N((0,0)', \dot{\Psi}(\theta_0)^{-1}\Sigma^{\theta_0}(\dot{\Psi}(\theta_0)^{-1})'), \tag{3.17}$$

with Σ^{θ_0} as specified in (3.12).

The estimator can easily be computed numerically. First note the linearity property $g_\theta(\cdot) = g_{(c,\alpha)}(\cdot) = c g_{(1,\alpha)}(\cdot)$. This property makes the numerical optimization problem relatively easy. The objective function can be written as

$$\Gamma_n(\theta) \equiv \Gamma_n(g_\theta) = \|g_\theta - \tilde{g}_n\|^2_w = \|c g_{(1,\alpha)} - \tilde{g}_n\|^2_w.$$

First we minimize $\Gamma_n(\theta)$ for fixed α with respect to c. This is easy, since $c \mapsto \Gamma_n((c,\alpha))$ is quadratic. Taking the partial derivative with respect to c gives

$$\frac{\partial}{\partial c}\Gamma_n((c,\alpha)) = 2c\|g_{(1,\alpha)}\|^2_w - 2\langle g_{(1,\alpha)}, \tilde{g}_n\rangle_w$$

and by equating this expression to zero we obtain

$$c_n(\alpha) = \frac{\langle g_{(1,\alpha)}, \tilde{g}_n\rangle_w}{\|g_{(1,\alpha)}\|^2_w}.$$

as a minimizer for fixed α. Now we can minimize $\alpha \mapsto \Gamma_n((c_n(\alpha), \alpha))$ numerically by using a Fibonacci search algorithm. Denote the minimizer by $\hat{\alpha}_n$, then $\hat{c}_n = c_n(\hat{\alpha}_n)$.

The asymptotic covariance can be obtained from Theorem 3.13 by numerical integration. We took $w(\cdot) = \mathbf{1}_{[-2,2]}(\cdot)$ and $\theta_0 = (3, 2)$. Applying Simpson's rule gives that

$$A^{\theta_0} := \dot{\Psi}(\theta_0)^{-1}\Sigma^{\theta_0}(\dot{\Psi}(\theta_0)^{-1})' \approx \begin{pmatrix} 25.5691 & 17.3076 \\ 17.3076 & 13.6966 \end{pmatrix}. \tag{3.18}$$

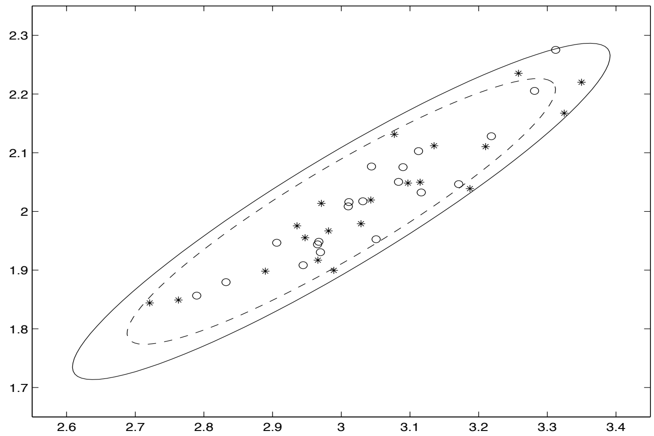

Figure 3.1: Gamma(3, 2) distribution, Scatterplot of parameter estimates for 20 i.i.d. data sets of sample-size $n = 1000$. Starred: CME, circled: MLE. Solid ellipse: 95%-contour plot for the CME. Dashed ellipse: 95%-contour plot for the MLE.

This value can be compared with the asymptotic variance we would find if we would apply maximum likelihood. This value is given by the inverse of the Fisher information matrix, which equals

$$I_{\theta_0}^{-1} = \frac{\alpha_0^2}{c_0\Upsilon(1,c_0)-1}\begin{pmatrix} c_0/\alpha_0^2 & 1/\alpha_0 \\ 1/\alpha_0 & \Upsilon(1,c_0)\end{pmatrix} \approx \begin{pmatrix} 16.2336 & 10.8224 \\ 10.8224 & 8.5483\end{pmatrix}. \tag{3.19}$$

Here Υ denotes the derivative of the logarithm of the Gamma-function. This shows that the CME is less efficient than the MLE. However, there is a gain in robustness, by the use of the empirical characteristic function, see Feuerverger and McDunnough (1981a).

For a sample of size $n = 1000$ we computed both the CME and MLE 20 times. The resulting scatterplot is given in Figure 3.1. By (3.17), for n large, the CME $\hat{\theta}_n$ has approximately a $N_2(\theta_0, \frac{1}{n}A^{\theta_0})$ distribution, where A^{θ_0} is given in (3.18). If we define the ellipse

$$\mathcal{E}_{n,\alpha} := \left\{\theta \in \mathbb{R}^2 \;:\; (\theta-\theta_0)'(A^{\theta_0})^{-1}(\theta-\theta_0) \leq \frac{\chi_{2,\alpha}}{n}\right\},$$

then, for n large, $P(\hat{\theta}_n \in \mathcal{E}_{n,\alpha}) \approx 1-\alpha$. Here $\chi_{2,\alpha}$ denotes the upper α-quantile of the χ^2-distribution with two degrees of freedom. In Figure 3.1 we added a contour plot of $\mathcal{E}_{20,0.05}$ for both the CME and the MLE.

Figure 3.2 shows the estimated canonical function and density in case we have respectively $n = 100$ (dashed curve) and $n = 1000$ (solid curve) observations. The true canonical function and true density are also plotted (dotted line).

For $n = 100$ we computed the CME 100 times. Denote the estimates by $q_{100} = (\hat{\theta}_{100}^{(1)}, \ldots, \hat{\theta}_{100}^{(100)})$, then

$$\text{mean}(q_{100}) = \begin{pmatrix} 3.1228 \\ 2.0836\end{pmatrix} \qquad \text{and} \qquad 100 * \text{cov}(q_{100}) = \begin{pmatrix} 23.4063 & 16.1694 \\ 16.1694 & 12.6988\end{pmatrix}.$$

Similarly, for $n = 1000$ we got

$$\text{mean}(q_{1000}) = \begin{pmatrix} 3.0038 \\ 2.0048\end{pmatrix} \qquad \text{and} \qquad 1000 * \text{cov}(q_{1000}) = \begin{pmatrix} 25.6449 & 16.2889 \\ 16.2889 & 12.3042\end{pmatrix}.$$

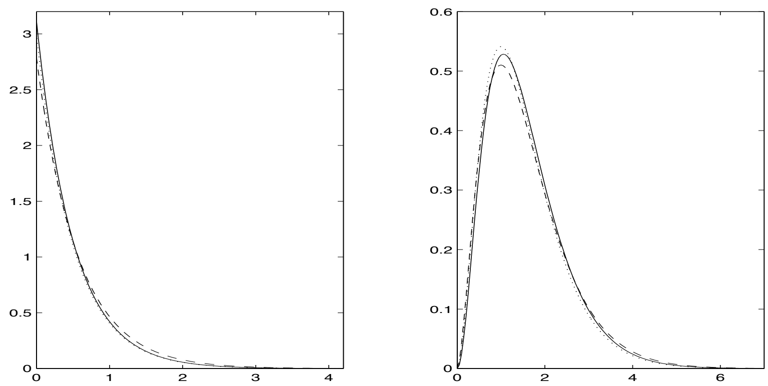

Figure 3.2: Gamma(3, 2) distribution, Left: $x \mapsto xa(x)$. Right: density function. Dashed: $n = 100$, solid: $n = 1000$, dotted: true.

Remark 3.18 The asymptotic results are only valid for a fixed (non-random) weight-function. We did not use the data-dependent weight-function as proposed in Section 2.5.2, since then our comparison of asymptotic variances between CME and MLE is not valid as stated.

3.5 Parametric estimation for OU-processes driven by a subordinator

In this section we extend the asymptotic results of Section 3.3 to OU-processes. Hence, assume X is a stationary OU-process, driven by a subordinator Z. We assume that ρ, the Lévy measure of Z, is parametrized by $\theta \in \Theta \subseteq \mathbb{R}^k$ and that we have discrete-time observations $X_0, X_\Delta, \ldots, X_{(n-1)\Delta}$ $(\Delta > 0)$ from X. The marginal law of X is characterized by its Lévy measure a_θ. We aim to estimate θ_0. This problem is the same as the problem considered so far, except for the fact that the observations are not independent. Nevertheless, the results on existence and consistency of the CME of sections 3.1 and 3.2 remain valid. In Section 3.5.1 we consider asymptotic normality for the CME. In Section 3.5.2 we apply the results to an Inverse-Gaussian OU-process.

3.5.1 Adaptations to the proof of asymptotic normality of the CME

Part **I** of Section 3.3 needs some adaptations, which we shall now work out. We start with a result due to Rio (1998). Denote the L_2-norm with respect to a measure Q by $\|\cdot\|_{2,Q}$. For the definition of an image admissible class we refer to the appendix (Section 3.6), Definition 3.24.

Theorem 3.19 (Rio (2000), Theorem 1) *Suppose $(X_n, n \in \mathbb{Z})$ is a stationary time-series with β-mixing coefficients β_n satisfying $\sum_{n>0} \beta_n < \infty$. Let $\mathcal{F}$ be a class of image admissible functions. Suppose*

(i) *there exists an envelope function F for the class $\mathcal{F}$ for which*

$$\int_0^1 \beta^{-1}(u) Q_F^2(1-u)du < \infty. \tag{3.20}$$

Here Q_F denotes the quantile function of $|F(X_0)|$ and $\beta^{-1}(u) = \inf\{k \in \mathbb{N} : \beta_k \leq u\}$.

(ii)

$$\int_0^1 \sqrt{H_2^F(\varepsilon, \mathcal{F}) \log(1/\varepsilon)} d\varepsilon < \infty, \tag{3.21}$$

where

$$H_2^F(\varepsilon, \mathcal{F}) := \log(\sup_Q N(\varepsilon \|F\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}) \vee 2),$$

and the supremum is taken over all finite discrete measures Q on $\mathbb{R}$ (recall the definition of N given before Theorem 3.6).

If we let P denote the common law of the observations and P_n the empirical measure of the first n observations, then $\sqrt{n}(P_n - P)$ converges weakly in $\ell^\infty(\mathcal{F})$ to a Gaussian process with covariance function Γ.

We verify the conditions of this theorem for $X_n \equiv X_{n\Delta}$ the discretely observed OU-process and $\mathcal{F}$ the class of functions defined by

$$\mathcal{F} := \{f_t \mid f_t(x) = e^{itx}, x \in \mathbb{R}, t \in S\}, \tag{3.22}$$

Since

$$\ell^\infty(\mathcal{F}) := \{z : \mathcal{F} \to \mathbb{R} \ : \ \|z\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |z(f)| = \sup_{t \in S} |z(f_t)| < \infty\},$$

it is natural to identify f_t with $t \in S$ and the space $\ell^\infty(\mathcal{F})$ with $\ell^\infty(S)$.

Corollary 3.20 *Let $X_0, X_\Delta, \ldots, X_{(n-1)\Delta}$ be observations from the stationary OU-process. If (3.8) holds, then the stochastic process $(Y_n(t), t \in S)$ converges weakly in the space $\ell^\infty(S)$ to a centered Gaussian process Y.*

Proof First note that $\mathcal{F}$ has envelope function $F \equiv 1$. By Lemma 3.25 in Section 3.6, $\mathcal{F}$ is image admissible. If (3.8) holds, then the OU-process is geometrically ergodic (Masuda (2004), Theorem 4.3). That is, there exists a constant c such that $\beta_n \leq e^{-c\Delta n}$. Therefore,

$$\beta^{-1}(u) \leq \inf\{k \in \mathbb{N} : e^{-c\Delta k} \leq u\} = \lceil -\tfrac{1}{c\Delta} \log u \rceil \leq -\tfrac{1}{c\Delta} \log u + 1.$$

Furthermore, for all $x \in (0,1)$, $Q_F(1-x) = \inf\{u : P(1 \leq u) \geq 1-x\} = 1$. These combined results show that condition (3.20) is satisfied if (3.8) holds.

The class $\mathcal{F}$ is Lipschitz in the parameter, in the sense that $|f_t(x) - f_s(x)| \leq d(s,t)R(x)$, with $d(s,t) = |s-t|$ and $R(x) = |x|$. We have

$$N(\varepsilon \|R\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}) \leq N(\varepsilon, S, d) = O(\varepsilon^{-1}). \tag{3.23}$$

The inequality follows as in Theorem 2.7.11 Van der Vaart and Wellner (1996). The proof is easy: let $t_1, \ldots, t_p$ be an ε-net for d for S, then for each $t \in S$ there is a t_i such that $d(t_i, t) \leq \varepsilon$. Now

$$\|f_t - f_{t_i}\|_{2,Q} \leq d(t, t_i) \|R\|_{2,Q} \leq \varepsilon \|R\|_{2,Q}.$$

Taking the supremum over all finite discrete measures Q in (3.23) shows that $\sup_Q N(\varepsilon\|R\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q})$ is bounded by a function of order $1/\varepsilon$. Thus (3.21) is satisfied and the above theorem applies to our case. □

Remark 3.21 To gain some insight to the conditions involved in Theorem 3.19, we compute for a fixed $t \in \mathbb{R}$, $v_t^n := \mathrm{var} Y_n(t)$. Define random variables (W_j) by $W_j = e^{itX_{j\Delta}} - E_\theta e^{itX_{j\Delta}}$. Then

$$v_t^n = \mathrm{var}(\sqrt{n}\,\overline{W}_n) = \sum_{h=-n}^{n} \left(\frac{n-|h|}{n}\right) \gamma_W(h),$$

where $\gamma_W(h) = \mathrm{cov}(W_0, W_h)$ denotes the auto-covariance function of the stationary process (W_j). Therefore, we can bound v_t^n by

$$\begin{aligned} v_t^n &\leq \gamma_W(0) + 2\sum_{h=1}^{n} \left(\frac{n-|h|}{n}\right) |\gamma_W(h)| \\ &\leq \gamma_W(0) + 4\int_0^1 \sum_{h=1}^{n} \mathbf{1}_{\{u<\beta_W(h)\}} Q^2_{|W_0|}(1-u)du \\ &\leq \gamma_W(0) + 4\int_0^1 \beta_X^{-1}(u) Q^2_{|W_0|}(1-u)du, \end{aligned}$$

where the second inequality follows from Theorem 1.1 in Rio (2000). The last inequality follows since the β-mixing numbers of W are smaller than the β-mixing numbers of X.

Thus condition (3.20) ensures that the variance of the limiting process is finite. Furthermore, the entropy condition measures the size of the class $\mathcal{F}$, which should not be too large.

Next, we go through steps ***(I)*** up to ***(III)*** in case we have dependent observations by pointing out where the differences with the case of independent observations occur. For ***(I)***, we see that the limit process Y is still Gaussian, but with covariance function Γ, for which we do not have a nice closed form expression. Accordingly, the limit process U in part ***(II)*** remains a Gaussian process, but with a different (more complicated) covariance structure as for the case with independent observations.

From Theorem 3.13 we still obtain that $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}(\theta_0)\mathbb{H}(\theta_0)$. In this case it is much harder to find the distribution of $\mathbb{H}(\theta_0)$.

3.5.2 Example: discrete observations from an Inverse Gaussian-OU-process

Suppose we have observations $X_0, \ldots, X_{(n-1)\Delta}$ from a stationary OU-process with Inverse Gaussian marginal law, which we denote by π_θ. Hence, for $\theta = (\delta, \gamma) \in (0,\infty) \times [0,\infty)$, $\pi_\theta \sim \mathrm{IG}(\delta, \gamma)$. The density of π_θ is given by

$$f_{(\delta,\gamma)}(x) = \frac{1}{\sqrt{2\pi}} \delta e^{\delta\gamma} x^{-3/2} \exp(-(\delta^2 x^{-1} + \gamma^2 x)/2)\mathbf{1}_{\{x>0\}},$$

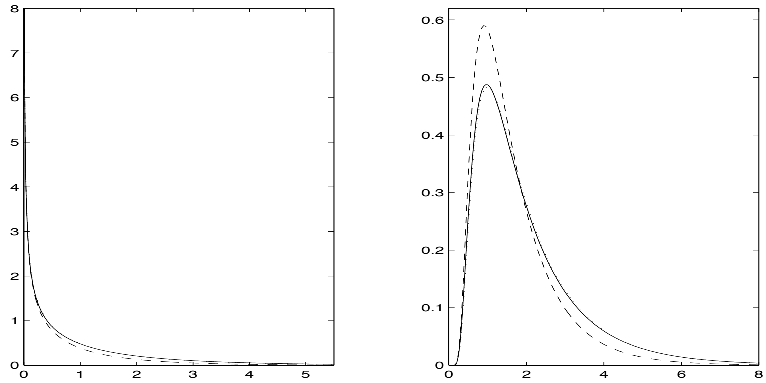

Figure 3.3: Inverse Gaussian(2, 1) distribution, Data from the OU-process. Left: canonical function. Right: density function. Dashed: $n = 100$, solid: $n = 1000$, dotted: true.

and if $\gamma > 0$ the cumulant function is given by

$$g_{(\delta,\gamma)}(t) = \delta\gamma - \delta|\gamma^2 - 2it|^{1/2} \exp(-\tfrac{i}{2}|\arctan(2t/\gamma^2)|).$$

Let $\Theta \subset (0,\infty) \times [0,\infty)$ be compact and suppose the true value of the parameter θ_0 is in Θ. The Lévy density is given by $a_\theta(x) = x^{-1}k_\theta(x)$, where the canonical function is given by $k_\theta(x) = \frac{1}{\sqrt{2\pi}}\delta x^{-1/2}\exp(-\gamma^2 x/2)$. It is easy to see that $\theta \mapsto a_\theta(x)$ is continuous for each $x > 0$. Assume the sequence of preliminary estimators satisfies (2.3) for convergence almost surely. By Corollary 3.2, $\hat{\theta}_n = \operatorname{argmin}_{\theta\in\Theta}\Gamma_n(\theta)$ exists. From Theorem 3.4 we obtain consistency: $\hat{\theta}_n \xrightarrow{\text{a.s.}} \theta_0$. If we take the empirical characteristic function as a preliminary estimator, we can verify the conditions for $\sqrt{n}(\hat{\theta}_n - \theta_0)$ to converge weakly to a 2-dimensional Normal distribution. Since this is basically the same problem as in the Gamma-example of Section 3.4, we omit discussion on this.

The numerical issues are also similar to the Gamma-case. This time we have the linearity property $g_\theta(\cdot) = g_{(\delta,\gamma)}(\cdot) = \delta g_{(1,\gamma)}(\cdot)$. Therefore, the objective function can be written as $\Gamma_n(\theta) = \|\delta g_{(1,\gamma)} - \tilde{g}_n\|_w^2$. Minimizing $\Gamma_n(\theta)$ for fixed γ with respect to δ gives

$$\delta_n(\gamma) = \frac{\langle g_{(1,\gamma)}, \tilde{g}_n\rangle_w}{\|g_{(1,\gamma)}\|_w^2}.$$

Now we can minimize $\gamma \mapsto \Gamma_n(\delta_n(\gamma), \gamma)$ numericallly by using a Fibonacci search algorithm. Denote the minimizer by $\hat{\gamma}_n$, then $\hat{\theta}_n = (\delta_n(\hat{\gamma}_n), \hat{\gamma}_n)$.

We simulated a stationary Inverse-Gaussian OU-process with $(\delta, \gamma) = (2, 1)$ and intensity parameter $\lambda = 2$ (by this we mean that Z is constructed such that $\pi \sim$IG$(2, 1)$). We took observations at $t = 0, \ldots, 999$ (i.e. $\Delta = 1$) and computed the CME based on the first 100 and all 1000 observations. We took $w(\cdot) = \mathbf{1}_{[-2,2]}(\cdot)$. Figure 3.3 shows estimated canonical function and density in both cases. (dashed curve: $n = 100$, solid curve: $n = 1000$). The true canonical function and true density are also plotted (dotted line). The estimated parameters corresponding to this figure are

$$\begin{pmatrix}\hat{\delta}_{100}\\ \hat{\gamma}_{100}\end{pmatrix} = \begin{pmatrix}1.9864\\ 1.2039\end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix}\hat{\delta}_{1000}\\ \hat{\gamma}_{1000}\end{pmatrix} = \begin{pmatrix}1.9600\\ 0.9877\end{pmatrix}.$$

So the main reason for the bad fit of the density-function in case $n = 100$ is due to a relatively bad estimation of γ. This has less influence on the fit of the canonical function, since in the expression for k, γ appears only in the exponent.

3.6 Appendix

Proof of Lemma 3.8 Let $\varepsilon > 0$ and $k_\alpha = E|X|^\alpha$. For each $\alpha \in (0, 1]$ and $x \geq 0$ we have $\sin(x) \leq x^\alpha$. Therefore $E \sin^2 |\varepsilon x| \leq E|\varepsilon X|^{2\alpha} = \varepsilon^{2\alpha} k_{2\alpha}$. Taking $\varepsilon = |t - s|/2$ gives

$$\sigma(s,t) \leq 2 \left|\frac{t-s}{2}\right|^\alpha \sqrt{k_{2\alpha}} = 2^{1-\alpha}|t-s|^\alpha \sqrt{k_{2\alpha}} =: C_\alpha |t-s|^\alpha.$$

Put $d_\alpha(s,t) = |s - t|^\alpha$, then $N(C_\alpha \varepsilon, \sigma, S) \leq N(\varepsilon, d_\alpha, S) \leq K\varepsilon^{-1/\alpha}$, for some positive constant K, depending on S. Since the right hand side of this display satisfies the entropy condition (3.4), this suffices. □

Proof of Lemma 3.12 We first show that for each $\theta \in \Theta$ the mapping $\Lambda_\theta : (C(S), \|\cdot\|_\infty) \to \mathbb{R}^k$, defined by

$$\Lambda_\theta(f) = (-2\langle g_\theta^1, f\rangle_w, \ldots, -2\langle g_\theta^d, f\rangle_w)', \quad f \in C(S),$$

is continuous.

By the Cauchy-Schwarz inequality, we have

$$\|\Lambda_\theta(f)\|_2^2 = \sum_{i=1}^k |-2\langle g_\theta^i, f\rangle_w|^2 \leq 4 \sum_{i=1}^k \|g_\theta^i\|_w^2 \|f\|_w^2.$$

Now if $\|f_n - f\|_\infty \to 0$, then (since $\|f_n - f\|_w^2 \leq \|f_n - f\|_\infty^2 \int w(t)dt$) also $\|f_n - f\|_w \to 0$ which implies (by the inequality in the preceding display) that $\Lambda_\theta(f_n - f) \to 0$ in $\mathbb{R}^k$. Finally, by linearity of Λ_θ it follows that $\Lambda_\theta(f_n) \to \Lambda_\theta(f)$.

Since $U_n \rightsquigarrow U$ in $C(S)$ we have by the continuous mapping theorem

$$\mathbb{H}_n(\theta) = \Lambda_\theta(U_n) \rightsquigarrow \Lambda_\theta(U) = \mathbb{H}(\theta).$$

To derive the distribution of $\mathbb{H}(\theta)$ note that

$$\begin{aligned}
&\int |\langle g_\theta^i, z_{\theta_0}(\cdot, x)\rangle_w|^2 d\pi_{\theta_0}(x) \\
&= \int \left|\Re \int g_\theta^i(t)\overline{z_{\theta_0}(t,x)} w(t)dt\right|^2 d\pi_{\theta_0}(x) \leq \int \int \left|g_\theta^i(t)\overline{z_{\theta_0}(t,x)} w(t)dt\right|^2 d\pi_{\theta_0}(x) \\
&\leq \int \int |e^{itx}|^2 d\pi_{\theta_0}(x)\, |g_\theta^i(t)|^2 \left|\frac{1}{\psi_{\theta_0}(t)}\right|^2 w(t)dt \leq \int |g_\theta^i(t)|^2 \left|\frac{1}{\psi_{\theta_0}(t)}\right|^2 w(t)dt.
\end{aligned}$$

This is finite, since, by Assumption 3.11*(ii)*, $t \mapsto g_\theta^i(t)$ is continuous, ψ_{θ_0} is continuous, ψ_{θ_0} is bounded away from zero on S, and S is compact.

This enables us to apply Lemma 3.22 below, which yields that the distribution of $\mathbb{H}(\theta)$ is multivariate normal with mean zero and covariance Σ^θ as specified in the statement of the theorem. □

For the next lemma, recall the definition of B^0 as given in (3.6).

Lemma 3.22 *(i) Let U be defined as in Corollary 3.10. That is, $U(t) = (\psi_{\theta_0}(t))^{-1} \int e^{itx} dB^0(x)$ $(t \in S)$. For $f \in C(\mathbb{R}, \mathbb{C})$,*

$$\langle f, U \rangle_w \stackrel{\text{a.s.}}{=} \int \langle f, z_{\theta_0}(\cdot, x) \rangle_w dB^0(x),$$

where z_{θ_0} is as defined in (3.11).

(ii) Let $f_i \in C(\mathbb{R}, \mathbb{R})$ $(i = 1, \ldots, k)$ be such that $\int f_i(x)^2 d\pi_{\theta_0}(x) < \infty$. The vector $V = (V_1, \ldots, V_k)' \in \mathbb{R}^k$ with $V_i = \int f_i(x) dB^0(x)$ possesses a $N_k(0, \Xi)$ distribution, where

$$\Xi|_{i,j} = cov(V_i, V_j) = \int f_i(x) f_j(x) d\pi_{\theta_0}(x) - \left(\int f_i(x) \pi_{\theta_0}(x) \right) \left(\int f_j(x) \pi_{\theta_0}(x) \right). \tag{3.24}$$

Proof The proof of *(i)* is straightforward:

$$\begin{aligned} \langle f, U \rangle_w &= \Re \int f(t) \overline{U(t)} w(t) dt = \Re \int f(t) \int \overline{z_{\theta_0}(t, x)} dB^0(x) w(t) dt \\ &\stackrel{\text{a.s.}}{=} \int \Re \int f(t) \overline{z_{\theta_0}(t, x)} w(t) dt \, dB^0(x) = \int \langle f, z_{\theta_0}(\cdot, x) \rangle_w dB^0(x), \end{aligned}$$

where we use the stochastic Fubini Theorem (Protter (2004), Theorem 64 of Chapter IV) at the third equality sign. The second assertion follows by first noting that for $a \in \mathbb{R}^k$,

$$a'V = \int \left(\sum_{i=1}^k a_i f_i(x) \right) dB^0(x).$$

For $f \in C(\mathbb{R}, \mathbb{R})$ satisfying $\int |f(x)|^2 d\pi_{\theta_0}(x) < \infty$, we have that $\int f(x) dB^0(x)$ is normally distributed with mean zero and variance

$$\int f^2(x) d\pi_{\theta_0}(x) - \left(\int f(x) d\pi_{\theta_0}(x) \right)^2,$$

which follows from elementary rules for Itô-integrals. Therefore, $a'V$ is normally distributed with mean zero and variance

$$\int \left(\sum_{i=1}^k a_i f_i(x) \right)^2 d\pi_{\theta_0}(x) - \left(\int \left(\sum_{i=1}^k a_i f_i(x) \right) d\pi_{\theta_0}(x) \right)^2 = a' \Xi a,$$

with Ξ as in (3.24). $\square$

The next result is similar to Lemma 19.24 in Van der Vaart (1998).

Lemma 3.23 *Suppose that $H_n \rightsquigarrow H$ in the space $\ell^\infty(\Theta)$. Assume almost all sample paths $\theta \mapsto H(\theta)$ are continuous on Θ. If $\hat{\theta}_n$ is a sequence in Θ such that $\hat{\theta}_n \xrightarrow{\text{p}} \theta_0$, then $H_n(\hat{\theta}_n) - H_n(\theta_0) \xrightarrow{\text{p}} 0$.*

Proof Since $\hat{\theta}_n \xrightarrow{\text{P}} \theta_0$ in Θ and $H_n \rightsquigarrow H$ in $\ell^\infty(\Theta)$ we have $(H_n, \hat{\theta}_n) \rightsquigarrow (H, \theta_0)$ in $\ell^\infty(\Theta) \times \Theta$.

Define a mapping $\varphi : \ell^\infty(\Theta) \times \Theta \to \mathbb{R}$ by $\varphi(z, \theta) = z(\theta) - z(\theta_0)$. The function φ is continuous with respect to the product semimetric at every point (z, θ) such that $\eta \mapsto z(\eta)$ is continuous at θ. If $(z_n, \theta_n) \to (z, \theta)$, then $z_n \to z$ uniformly on Θ and hence $z_n(\theta_n) = z(\theta_n) + o(1) \to z(\theta)$, if z is continuous at θ.

Since H is assumed to be continuous on Θ a.s., we have by the continuous mapping theorem that $\varphi(H_n, \hat{\theta}_n) \rightsquigarrow \varphi(H, \theta_0)$. This means that $H_n(\hat{\theta}_n) - H_n(\theta_0) \rightsquigarrow H(\theta_0) - H(\theta_0) = 0$. The lemma follows, since convergence in distribution and convergence in probability are the same for a degenerate limit. □

Definition 3.24 Let $\mathcal{F}$ be a class of functions on a measurable space $(X, \mathcal{B}(X))$. $\mathcal{F}$ is called image-admissible if there exist a locally compact space with countable base Y, with its Borel-σ-algebra $\mathcal{B}(Y)$, and a surjective mapping $\tilde{T} : Y \to \mathcal{F}$, for which

$$T : (X \times Y, \mathcal{B}(X) \otimes \mathcal{B}(Y)) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$$

defined by $T(x, y) = [\tilde{T}(y)](x)$, is measurable.

Lemma 3.25 *If we take $X = \mathbb{R}$ and $Y = S$ and $T(x, y) = e^{iyx}$, then $\mathcal{F}$, as defined in (3.22) is image-admissible.*

Proof If we define $\tilde{T} : S \to \mathcal{F}$ by the mapping $x \mapsto [\tilde{T}(y)](x)$, then $\tilde{T}$ is surjective. The mapping $T : (\mathbb{R} \times S, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(S)) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ defined by $T(x, y) = [\tilde{T}(y)](x)$ is continuous in the product topology and hence measurable with respect to the product σ-algebra $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(S)$. □

Chapter 4

Convergence rates of posterior distributions for Brownian semimartingale models

4.1 Introduction

Suppose that we observe the stochastic process $X^n = (X_t^n, 0 \le t \le T_n)$ defined through the stochastic differential equation (SDE)

$$dX_t^n = \beta^{\theta,n}(t, X^n)\, dt + \sigma^n(t, X^n)\, dB_t^n, \quad t \in [0, T_n], \quad X_0^n = X_0, \tag{4.1}$$

where B^n is a standard Brownian motion. Based on a realization of X^n we wish to make inference on the parameter θ that determines the shape of the "drift coefficient" $\beta^{\theta,n}$. The "diffusion coefficient" σ^n is considered to be known, as it can be determined without error from continuous observations of the process. The natural number $n \in \mathbb{N}$ serves as an indexing parameter for our asymptotic setup, in which n tends to infinity. The endpoint T_n of the observational interval may be fixed or tend to infinity. This *Brownian semimartingale model* contains the *diffusion model* as the special case in which $\beta^{\theta,n}$ and σ^n are measurable functions of the process X_t^n at time t only.

To set this up more formally we assume that $\beta^{\theta,n}$ and σ^n are measurable functions that satisfy regularity conditions that ensure that the SDE (4.1) has a unique weak solution. We then let $P^{\theta,n}$ be the induced distribution on the Borel sets of the space $C[0, T_n]$ of continuous functions on $[0, T_n]$ of a solution $X^n = (X_t^n, 0 \le t \le T_n)$, and consider the statistical experiment $(P^{\theta,n} : \theta \in \Theta^n)$ for a given parameter set Θ^n. We are mostly interested in the case that the parameter set Θ^n is infinite-dimensional, but our results also apply to parametric models.

The Bayesian approach to statistical inference consists of putting a prior distribution Π^n on the parameter set Θ^n and making inference based on the posterior distribution $\Pi^n(\cdot|X^n)$. The latter is the conditional distribution of the parameter θ given the observation X^n. In this chapter we adopt the Bayesian framework to define the posterior distribution, but study the properties of the posterior distribution from a frequentist point of view. This entails that we assume that the observation X^n is generated from a measure $P^{\theta_0,n}$ in the model, where the value θ_0 is referred to as the "true value" of the parameter.

We are interested in the asymptotic behaviour of the posterior distributions, as $n \to \infty$. If the priors Π^n do not exclude θ_0 as a possible value of θ, then we may expect posterior consistency, meaning that the sequence of random measures $\Pi^n(\cdot|X^n)$ converges weakly to the degenerate measure at θ_0. In this chapter we focus on the rate of this convergence. More precisely, we focus on the maximal rate at which we can shrink balls around θ_0 while still capturing almost all posterior probability. Our main result is a characterization of this rate through a measure of the amount of prior mass near θ_0 and a measure of the complexity of the parameter set Θ^n relative to the SDE model.

Earlier work on versions of this problem include Ibragimov and Has'minskii (1981), Kutoyants (1994), Kutoyants (2004), Zhao (2000), Shen and Wasserman (2001) and Ghosal and Van der Vaart (2004). The last paper relates the problem to general Bayesian inference, and we refer to this paper for further references and an overview of the literature on Bayesian asymptotics. Ghosh and Ramamoorthi (2003) give many examples of prior distributions in nonparametric models, and discuss consistency. Results on non Bayesian methods can be found in Prakasa Rao (1999) and Kutoyants (1984).

Versions of the *parametric* Brownian semimartingale model, in which the process $\beta^{\theta,n}$ depends smoothly on a Euclidean parameter have been studied in detail. The *Gaussian white-noise model*, in which the drift coefficient is a deterministic function, is well understood, also from a Bayesian point of view. Here, the observational interval is supposed to be fixed and the diffusion-coefficients consist of a sequence of constants tending to zero as n tends to infinity. Results on parametric Bayesian estimation are summarized in Ibragimov and Has'minskii (1981), Theorem II.5.1, who prove asymptotic normality and efficiency for Bayes-estimators under various loss-functions under conditions that imply local asymptotic normality (LAN) of the statistical models. The rate of convergence in this case is equal to the size of the drift constants σ^n. The *perturbed dynamical system* is an extension of the white-noise model, which allows the drift-coefficient to depend on the solution X_t^n in addition to t. This model is treated in depth in the book Kutoyants (1994). Under natural regularity conditions these models are LAN, and Bayes estimators typically converge at rate σ^n and are asymptotically normal (Kutoyants (1994), Theorem 2.2.3). Results on nonstandard situations, such as model misspecification or nonregular parametrizations, can be found in this book too. In the *ergodic diffusion model* both the drift and diffusion coefficients may depend on the solution X_t^n, but they are assumed to have a form independent of n. The asymptotics here is on the endpoint T_n of the observational interval, which tends to infinity. Again these models are LAN under natural conditions, with scaling rate $\sqrt{T_n}$. Results on these models are derived in Kutoyants (2004).

Much less is known about the *nonparametric* Brownian semimartingale model, except for the very special case of the Gaussian white noise model. The Gaussian white noise model has been studied from many perspectives, and in the Bayesian set-up with many priors (see e.g. Zhao (2000), Shen and Wasserman (2001)). It was put in a more general framework of non-i.i.d. models in Ghosal and Van der Vaart (2004), Section 5. The general Brownian semimartingale model is much more complicated. The main focus of the present paper is on this general model.

A key difficulty of the general Brownian semimartingale model is that the *Hellinger semimetric* is, in general, a random process rather than a true semimetric. The square of the Hellinger semimetric h_n is given by

$$h_n^2(\theta,\theta_0)=\int_0^{T_n}\Big(\frac{\beta^{\theta,n}-\beta^{\theta_0,n}}{\sigma^n}\Big)^2(t,X^n)\,dt.$$

It is the natural semimetric to use, as the log-likelihood process (with respect to $P^{\theta_0,n}$) of the model can be written as $M-\frac{1}{2}[M]$ for a certain continuous local martingale M and the square Hellinger semimetric $h_n^2(\theta,\theta_0)=[M]_{T_n}$ is the quadratic variation of this martingale M.

The best possible rate of convergence is determined by the likelihood process of the model, and in a more technical way by the existence of appropriate tests of the true parameter versus balls of alternatives. The martingale representation of the log likelihood and Bernstein's inequality allow to construct such tests relative to the Hellinger semimetric. Unfortunately, the randomness of this semimetric causes complications that preclude straightforward extension of the Ghosal and Van der Vaart (2004)-result, and motivate the present chapter. In part we follow ideas from Van Zanten (2004), who considers convergence rates for the maximum likelihood estimator of the Brownian semimartingale model.

Our main theorem (Theorem 4.2) bounds the posterior rate of convergence in terms of the complexity of the model and the amount of prior mass given to balls centered around the true parameter. In the statement of the theorem, the distance of θ to the true parameter θ_0 is measured by the Hellinger semimetric, but it is often possible to translate this result in terms of a deterministic semimetric d_n. We illustrate the usefulness of our main result by three classes of examples of SDEs: the Gaussian white-noise model, the perturbed dynamical system, and the ergodic diffusion model. Explicit calculations using a variety of priors are included.

The organization of this chapter is as follows. In Section 4.2 we present our main result. We specialize this result to three classes of SDEs in Section 4.3. The proof of the main result and technical complements are deferred to Section 4.4.

Notation: The notation $\lesssim$ will be used to denote inequality up to a universal multiplicative constant, ore up to a constant that is fixed throughout. We use the notation Pf to abbreviate $\int f dP$. By $a_n\sim b_n$ ($n\to N\in[-\infty,\infty]$) we mean that $\lim_{n\to N}a_n/b_n=c$ for some positive constant c.

4.2 Main result

For $n\in\mathbb{N}$, given numbers $T_n>0$, and each θ in an arbitrary set Θ^n let $\beta^{\theta,n}$ and σ^n be measurable and nonanticipative [1] functions on $[0,T_n]\times C[0,T_n]$ such that the SDE

$$dX_t^n=\beta^{\theta,n}(t,X^n)\,dt+\sigma^n(t,X^n)\,dB_t^n,\quad t\in[0,T_n],\quad X_0^n=X_0 \tag{4.2}$$

[1] Let $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq0},P)$ be a filtered probability space. The random process $\xi=(\xi(t),t\geq0)$ is called *measurable* if, for all Borel sets B of the real line $\{(\omega,t):\xi_t(\omega)\in B\}\in\mathcal{F}\times\mathcal{B}([0,\infty))$, where $\mathcal{B}([0,\infty))$ is the σ-algebra of Borel sets on $[0,\infty)$. The measurable (with respect to a pair of variables (t,ω)) function $f=f(t,\omega)$, $t\geq0$, $\omega\in\Omega$ is called *nonanticipative* with respect to the family $F=(\mathcal{F}_t)_{t\geq0}$ if, for each t, it is $\mathcal{F}_t$-measurable. Such functions are often also called functions independent of the future. See Section 4.2 in Liptser and Shiryayev (1977).

possesses a unique weak solution $X^n = \big(X^n_t, t \in [0, T_n]\big)$. Here B^n is a standard Brownian motion. Denote the distribution of the process X^n on the Borel sets $\mathcal{C}^n$ of the space $C[0, T_n]$, endowed with the supremum norm, by $P^{\theta,n}$. The parameter value $\theta_0 \in \Theta^n$, which may also depend on n, will refer to the "true value" of the parameter: throughout we consider the distribution of X^n under the assumption that X^n satisfies the SDE with θ_0 instead of θ.

Under regularity conditions the measures $P^{\theta,n}$ are equivalent and possess densities

$$p^{\theta,n}(X^n) = \exp\bigg(\int_0^{T_n} \Big(\frac{\beta^{\theta,n}}{(\sigma^n)^2}\Big)(t, X^n)\, dX^n_t - \frac{1}{2}\int_0^{T_n} \Big(\frac{\beta^{\theta,n}}{\sigma^n}\Big)^2(t, X^n)\, dt\bigg) \tag{4.3}$$

relative to the common dominating measure which is the distribution of the process X^n, defined as the solution of (4.2) in case $\beta \equiv 0$. The following conditions are necessary and sufficient for this to be true, and are assumed throughout the paper:

- There exists a standard filtered probability space $(\Omega^n, \mathcal{F}^n, (\mathcal{F}^n_t, t \geq 0), \Pr^n)$ and a parameter value θ_0 on which the SDE (4.2) with θ_0 substituted for θ possesses a solution $X^n = \big(X^n_t, t \in [0, T_n]\big)$.
- This solution satisfies $\int_0^{T_n} (\beta^{\theta,n}/\sigma^n)^2(t, X^n)\, dt < \infty$ $\Pr^n$-almost surely for every $\theta \in \Theta^n$ and

$$E^n \exp\left(\int_0^{T_n} \Big(\frac{\beta^{\theta,n} - \beta^{\theta_0,n}}{\sigma^n}\Big)(t, X^n)\, dB^n_t - \frac{1}{2}\int_0^{T_n} \Big(\frac{\beta^{\theta,n} - \beta^{\theta_0,n}}{\sigma^n}\Big)^2 (t, X^n)\, dt\right) = 1.$$

The necessity of these conditions is clear (note that the exponential in the second condition is the quotient $p^{\theta,n}/p^{\theta_0,n}(X^n)$), and the sufficiency follows readily with the help of Girsanov's theorem (cf. Section 4.4.6). There are several approaches in the literature to verify the first condition under more concrete conditions on the drift and diffusion functions. The second condition is generally hardest to verify. Liptser and Shiryayev (1977) discuss this issue at length and provide elementary sufficient conditions. We defer a discussion of their results to the special examples in the next section.

We assume that the parameter set Θ^n is equipped with some σ-field $\mathcal{B}^n$ and that for all n, the map $(x, \theta) \mapsto p^{\theta,n}(x)$ is jointly measurable relative to $\mathcal{C}^n \times \mathcal{B}^n$. Then given a prior distribution Π^n, a probability distribution on $(\Theta^n, \mathcal{B}^n)$, the *posterior distribution* can be defined by

$$\Pi^n(B|X^n) = \frac{\int_B p^{\theta,n}(X^n)\, d\Pi^n(\theta)}{\int_{\Theta^n} p^{\theta,n}(X^n)\, d\Pi^n(\theta)}, \qquad B \in \mathcal{B}^n. \tag{4.4}$$

Because the measures $P^{\theta,n}$ are equivalent (by assumption), the expression on the right hand side is with probability one well defined, and apart from definition on a null set, gives a Markov kernel. In the Bayesian set-up it is the conditional distribution of the parameter given X^n, but in this paper we take the display as a definition of the kernel on the left, and study its behaviour under the measures $P^{\theta_0,n}$.

Under mild conditions $\Pi^n(B|X^n) \to 1$ in $P^{\theta_0,n}$-probability as $n \to \infty$ for any fixed "neighbourhood" B of θ_0. We are interested in the maximal rate at which we can shrink balls around θ_0, while still capturing almost all posterior mass. This can be formalized using a semimetric d_n on the parameter set Θ^n by saying that the sequence of posterior distributions converges to θ_0 (at least) at rate μ_n if for every sequence $M_n \to \infty$,

$$P^{\theta_0,n}\Pi^n(\theta \in \Theta^n : d_n(\theta, \theta_0) \geq M_n\mu_n | X^n) \to 0.$$

(It is a matter of taste, whether to write $E^{\theta_0,n}$ or $P^{\theta_0,n}$ for the above expectation.) The posterior rate of convergence reveals the size of Bayesian credibility regions (central regions of mass $1-\alpha$ in the posterior distribution). It also implies the same rate for a variety of derived point estimators, such as the posterior mode and (under some conditions) the posterior mean.

Our main result is formulated in terms of three semimetrics h_n, d_n and $\bar{d}_n$ on the parameter set Θ^n. The first is the *Hellinger semimetric* h_n given by

$$h_n^2(\theta, \psi) := \int_0^{T_n} \Big(\frac{\beta^{\theta,n} - \beta^{\psi,n}}{\sigma^n}\Big)^2 (t, X^n)\, dt, \quad \theta, \psi \in \Theta. \tag{4.5}$$

The Hellinger semimetric is random, unlike the other two semimetrics d_n and $\bar{d}_n$ we shall employ, which are ordinary semimetrics. They are related to the Hellinger semimetric through the following assumption. Let μ_n be the desired rate of convergence, a sequence of positive numbers.

Assumption 4.1 For every $\gamma > 0$ there exist positive constants $c = c_\gamma$, $C = C_\gamma$ and a non-negative constant $D = D_\gamma$ such that [2]

$$\begin{aligned}\liminf_{n\to\infty} P_*^{\theta_0,n}\Big(&c\, d_n(\theta, \theta_0) \leq h_n(\theta, \theta_0), \forall\, \theta \in \Theta^n \text{ with } h_n(\theta, \theta_0) \geq D\mu_n \text{ and} \\ &h_n(\theta, \psi) \leq C\, \bar{d}_n(\theta, \psi), \forall\, \theta, \psi \in \Theta^n \text{ with } h_n(\theta, \psi) \geq D\mu_n\Big) \geq 1 - \gamma. \end{aligned} \tag{4.6}$$

The *ε-covering number* of a set A for a semimetric ρ, denoted by $N(\varepsilon, A, \rho)$, is defined as the minimal number of ρ-balls of radius ε needed to cover the set A. The logarithm of the covering number is referred to as the entropy.

Our main theorem poses two conditions: the first one, (4.7), measures the complexity of the model by the so-called *local Kolmogorov entropy* or *Le Cam dimension*, the second condition, (4.8), requires that the prior puts sufficient mass close to the true parameter value θ_0. Denote by $B^n(\theta_0, \varepsilon)$ and $\bar{B}^n(\theta_0, \varepsilon)$ the balls of d_n- and $\bar{d}_n$-radius ε around θ_0.

Theorem 4.2 *Let μ_n be a sequence of positive numbers that is bounded away from zero. Suppose Assumption 4.1 is satisfied by the sequence μ_n and that for every $a > 0$ there exists a constant $g(a) < \infty$ such that*

$$\sup_{\mu > \mu_n} \log N\big(a\mu, B^n(\theta_0, \mu), \bar{d}_n\big) \leq \mu_n^2 g(a). \tag{4.7}$$

Furthermore, assume that for every $\xi > 0$ there exists an integer J such that for $j \geq J$

$$\frac{\Pi^n\big(B^n(\theta_0, j\mu_n)\big)}{\Pi^n\big(\bar{B}^n(\theta_0, \mu_n)\big)} \leq e^{\xi\mu_n^2 j^2}. \tag{4.8}$$

[2] We use inner-probability since the set in (4.6) may not be measurable. On a measure space $(S, \mathcal{S}, \mu)$ the *inner* μ-measure of a set $G \subseteq S$ is defined by $\mu_*(G) = \sup\{\mu(F) : F \in \mathcal{S}; F \subseteq G\}$. In the following, if measurability problems appear, the corresponding probabilities can be read as outer/inner-probabilities.

Then for every $M_n \to \infty$, we have that

$$P^{\theta_0,n}\Pi^n\big(\theta \in \Theta^n : h_n(\theta,\theta_0) \geq M_n\mu_n |\, X^n\big) \to 0. \tag{4.9}$$

If $\inf_{\gamma>0} c_\gamma/C_\gamma \geq a_0 > 0$, then the entropy condition (4.7) needs to hold for $a = a_0/8$ only. If $\inf_{\gamma>0} c_\gamma \geq c_0 > 0$, then the prior mass condition (4.8) needs to hold for $\xi = c_0^2/9216$ only.

The proof of the theorem is deferred to Section 4.4. The assertion of the theorem remains true if h_n in (4.9) is replaced by the lower semimetric d_n.

In our examples the semimetrics satisfy $d_n = c_n d$ and $\bar{d}_n = c_n\bar{d}$, for a sequence of positive numbers c_n and fixed semimetrics d and $\bar{d}$. Scaling properties of entropies and neighbourhoods then yield a rate of convergence $\mu_n = c_n\varepsilon_n$ (with respect to d_n) for ε_n satisfying the bounds

$$\sup_{\varepsilon>\varepsilon_n} \log N\big(a\varepsilon, B(\theta_0,\varepsilon), \bar{d}\big) \leq c_n^2\varepsilon_n^2 g(a). \tag{4.10}$$

$$\frac{\Pi^n\big(B(\theta_0, j\varepsilon_n)\big)}{\Pi^n\big(\bar{B}(\theta_0,\varepsilon_n)\big)} \leq e^{\xi c_n^2\varepsilon_n^2 j^2}. \tag{4.11}$$

Here $B(\theta_0,\varepsilon)$ and $\bar{B}(\theta_0,\varepsilon)$ are the balls of radius ε around θ_0 for the fixed semimetrics d and $\bar{d}$, respectively. These two equations replace (4.7) and (4.8) in the preceding theorem. It is then still assumed that Assumption 4.1 holds, with $\mu_n = c_n\varepsilon_n$, $d_n = c_n d$ and $\bar{d}_n = c_n\bar{d}$.

The prior mass conditions (4.8) and (4.11) concern the relative amount of prior mass close to θ_0 (denominator) and farther from θ_0 (numerator). Because the numerator is trivially bounded above by 1, (4.8) is implied by the condition

$$\Pi^n\big(\bar{B}^n(\theta_0,\mu_n)\big) \geq e^{-\mu_n^2}. \tag{4.12}$$

This is a lower bound on the prior mass close to θ_0.

The entropy condition (4.7) is sometimes restrictive, because it treats the parameter set in a uniform way, irrespective of the prior mass. The presence of a subset of parameters with large entropy, but small prior mass typically does not affect the rate of convergence. The following lemma allows to handle such situations. We first remark that the preceding theorem remains true if the prior measures Π^n are supported on larger parameter sets $\bar{\Theta}^n \supset \Theta^n$, where the balls $B^n(\theta_0,\varepsilon) = \{\theta \in \Theta^n : d_n(\theta,\theta_0) \leq \varepsilon\}$ are still defined to be subsets of the smaller set Θ^n and the assertion (4.9) remains unchanged. Thus the entropy (4.7) is measured only within Θ^n, but the assertion also only concerns the posterior within Θ^n. (The posterior distribution is now defined by (4.4) with Θ^n replaced by $\bar{\Theta}^n$, for measurable sets $B \subset \bar{\Theta}^n$.) The following lemma, whose proof is given in Section 4.4.4, allows to complement this with a result for parameter-values in $\bar{\Theta}^n \setminus \Theta^n$. It shows that sets $\bar{\Theta}^n \setminus \Theta^n$ with very small prior measure automatically have negligible posterior measure, and hence can be ignored.

Lemma 4.3 *If for every $\gamma > 0$,*

$$\frac{\Pi^n(\bar{\Theta}^n \setminus \Theta^n)}{\Pi^n\big(\bar{B}^n(\theta_0,\mu_n)\big)} = o\big(e^{-(C_\gamma \vee D_\gamma)^2\mu_n^2}\big), \tag{4.13}$$

then

$$P^{\theta_0,n}[\Pi^n(\bar{\Theta}^n \setminus \Theta^n |\, X^n)] \to 0, \quad n \to \infty.$$

The proof is given in Section 4.4.4.

4.3 Special cases

In this section we consider a number of special cases of the Brownian semimartingale model. We give examples of priors and derive the rate of convergence according to our main theorem.

4.3.1 Signal in white noise

In the signal in white noise model we observe the process X^n satisfying

$$dX_t^n = \theta_0(t)\,dt + \sigma_n\,dB_t, \quad t \leq T, \quad X_0^n = x_0.$$

We observe the process X^n up to a fixed endpoint T. The "noise level" σ_n is a deterministic sequence of positive numbers that tends to zero as $n \to \infty$. The parameter θ_0 is a deterministic function that belongs to a subset Θ of $L^2[0,T]$. Write $\|\cdot\|$ for the $L^2[0,T]$-norm.

In this case the Hellinger semimetric is nonrandom, and given by

$$h_n(\theta,\psi) = \frac{1}{\sigma_n}\|\theta - \psi\|.$$

It follows that Assumption 4.1 holds with $\gamma = 0$, $c = C = 1$, $D = 0$ and $d_n = \bar{d}_n = h_n$. Theorem 4.2 then yields the following theorem.

Theorem 4.4 *Let ε_n be a sequence of positive numbers such that ε_n/σ_n is bounded away from zero. Suppose that there exists a constant $K < \infty$ such that*

$$\sup_{\varepsilon>\varepsilon_n} \log N\big(\varepsilon/8, \{\theta \in \Theta : \|\theta - \theta_0\| < \varepsilon\}, \|\cdot\|\big) \leq K(\varepsilon_n/\sigma_n)^2.$$

and assume that there exists an integer J such that for $j \geq J$

$$\frac{\Pi^n\big(\theta \in \Theta : \|\theta - \theta_0\| \leq j\varepsilon_n\big)}{\Pi^n\big(\theta \in \Theta : \|\theta - \theta_0\| \leq \varepsilon_n\big)} \leq e^{j^2(\varepsilon_n/\sigma_n)^2/9216}. \tag{4.14}$$

Then for any sequence $M_n \to \infty$ we have

$$P^{\theta_0,n}\Pi^n(\theta \in \Theta : \|\theta - \theta_0\| \geq M_n\varepsilon_n |\, X^n) \to 0. \tag{4.15}$$

For $\sigma_n = n^{-1/2}$, we recover Theorem 6 in Ghosal and Van der Vaart (2004), who also give examples of priors. Note that the conditions are purely in terms of the L_2-distance.

4.3.2 Perturbed dynamical system

The "perturbed dynamical system" is described by the SDE

$$dX_t^n = \theta_0(X_t^n)\,dt + \sigma_n\,dB_t^n, \quad t \leq T, \quad X_0^n = x_0.$$

The "noise level" σ_n is a sequence of positive constants that tends to zero. We observe the process X^n up to a fixed time T. The parameter θ_0 belongs to a class of functions Θ on the real line, which we will specify below.

Under natural conditions, as $n \to \infty$ the processes X^n will tend to the solution $t \mapsto x_t$ of the unperturbed ordinary differential equation (ODE)

$$dx_t = \theta_0(x_t)\, dt.$$

For instance, if θ_0 is Lipschitz, then for some positive constant C

$$\begin{aligned} |X_t^n - x_t| &= \int_0^t |\theta_0(X_s^n) - \theta_0(x_s)| ds + \sigma_n |B_t^n| \\ &\leq C \int_0^t |X_s^n - x_s| ds + \sigma_n |B_t^n| \end{aligned}$$

and the Gronwall inequality (e.g. Karatzas and Shreve (1991), pp 287–288) implies that

$$\sup_{0 \leq t \leq T} |X_t^n - x_t| \leq \sigma_n \sup_{0 \leq t \leq T} \left(|B_t^n| + C \int_0^t \sigma_n |B_s^n| e^{C(t-s)} ds \right) = O_{P^{\theta_0,n}}(\sigma_n).$$

It follows that the process X^n will with probability tending to one take its values in a neighbourhood of the range of the deterministic function $t \mapsto x_t$, and hence in a compact set. The nature of the functions θ in the parameter set Θ therefore matters only through their restrictions to a compact set, and the semimetrics and entropies may be interpreted accordingly.

The convergence of the processes X^n is also the key to finding appropriate semimetrics d_n and $\bar{d}_n$. The Hellinger semimetric h_n is given by

$$h_n(\theta, \psi) = \frac{1}{\sigma_n} \sqrt{\int_0^T \left(\theta(X_t^n) - \psi(X_t^n)\right)^2 dt}.$$

The convergence of X^n to the solution $t \mapsto x_t$ of the corresponding ODE suggests that

$$\sigma_n^2 h_n^2(\theta, \theta_0) \to d^2(\theta, \theta_0),$$

for

$$d(\theta, \psi) = \sqrt{\int_0^T (\theta(x_t) - \theta_0(x_t))^2\, dt}. \tag{4.16}$$

We choose $(1/\sigma_n)$ times the semimetric d as both the lower semimetric d_n and upper semimetric $\bar{d}_n$ in the application of our main theorem. Typically, the solution of the ODE will be sufficiently regular to ensure that the semimetric d is equivalent to the L_2-semimetric on the range $\{x_t : t \in [0, T]\}$ of this solution. Of course, the semimetric d is always bounded above by the uniform norm on a neighborhood of the range $\{x_t : t \in [0, T]\}$ of the solution to the ODE, and hence we may use the uniform metric as well.

That the approximation d/σ_n of h_n satisfies Assumption 4.1 is made precise under a Lipschitz condition in the following theorem.

Theorem 4.5 *Let ε_n be a sequence of positive numbers such that ε_n/σ_n is bounded away from zero. Assume that*

$$\sup_{\theta \in \Theta} \sup_x |\theta(x)| < \infty, \qquad \sup_{\theta \in \Theta} \sup_{x \neq y} \frac{|\theta(x) - \theta(y)|}{|x - y|} < \infty, \tag{4.17}$$

Suppose there exists a constant $K < \infty$ such that

$$\sup_{\varepsilon > \varepsilon_n} \log N\bigl(\varepsilon/24, \{\theta \in \Theta^n : d(\theta,\theta_0) < \varepsilon\}, d\bigr) \le K(\varepsilon_n/\sigma_n)^2, \tag{4.18}$$

where d is given in (4.16). Furthermore, assume there exists an integer J such that for $j \ge J$

$$\frac{\Pi^n\bigl(\theta \in \Theta^n : d(\theta,\theta_0) < j\varepsilon_n\bigr)}{\Pi^n\bigl(\theta \in \Theta^n : d(\theta,\theta_0) < \varepsilon_n\bigr)} \le e^{\varepsilon_n^2 j^2/(20736\sigma_n^2)}. \tag{4.19}$$

Then for every sequence $M_n \to \infty$, we have

$$P^{\theta_0,n}\Pi^n\Bigl(\theta \in \Theta^n : \int_0^T \bigl(\theta(x_t) - \theta_0(x_t)\bigr)^2\,dt \ge M_n\varepsilon_n^2 \,|\, X^n\Bigr) \to 0. \tag{4.20}$$

Proof Under the Lipschitz condition (4.17) the Gronwall inequality mentioned previously shows that

$$\sup_{\theta,\psi\in\Theta} \bigl|\sigma_n h_n(\theta,\psi) - d(\theta,\psi)\bigr| = O_{P^{\theta_0,n}}(\sigma_n). \tag{4.21}$$

(Cf. the proof of Proposition 5.2 in Van Zanten (2004).) Using Lemma 4.10 from the appendix with $\varepsilon = 1/2$, we see that Assumption 4.1 is fulfilled for $c = 2/3$, $C = 2$ and $d_n = \bar{d}_n = (1/\sigma_n)d$. The theorem now follows from Theorem 4.2. □

Discrete priors

The current standard for *α-regular functions* on an interval $[-M,M] \subset \mathbb{R}$ is the Besov space $B^\alpha_{p,\infty}$ of functions $\theta : [-M,M] \to \mathbb{R}$ with

$$\|\theta\|^\alpha_{p,\infty} := \|\theta\|_p + \sup_{t>0} \frac{1}{t^\alpha} \sup_{0<h<t} \bigl\|\Delta_h^{\bar\alpha}\theta\bigr\|_p < \infty.$$

Here $\|\cdot\|_p$ is L_p-norm with respect to Lebesgue measure, $\bar\alpha$ is an integer strictly bigger than α, and $\Delta_h^{\bar\alpha}$ is the $\bar\alpha$-th difference operator, defined recursively by $\Delta_h^r = \Delta_h^{r-1}\Delta_h$ and $\Delta_h\theta(x) = \theta(x+h) - \theta(x)$ (cf. Devore and Lorentz (1993)). This Besov space contains in particular all functions that are $\bar\alpha$ times differentiable with bounded $\bar\alpha$-th derivative. See also Definition 9.2 (page 104) and Corollary 9.1 (page 123) in Härdle et al. (1998).

For $p > 1/\alpha$ the entropy of the unit ball of the Besov space $B^\alpha_{p,\infty}$ for the uniform norm is of the order $(1/\varepsilon)^{1/\alpha}$ (cf. Birgé and Massart (2000) and Kerkyacharian (2004)).

We choose a multiple of this unit ball as parameter set Θ, and define a prior Π^n by choosing for given numbers $\varepsilon_n > 0$ a minimal $\varepsilon_n/2$-net over Θ for the uniform norm and defining Π^n to be the discrete uniform measure on this finite set of functions. If N^n is the number of points in the support of this prior, then $\log N^n$ is of the order $(1/\varepsilon_n)^{1/\alpha}$. A uniform neighborhood of radius ε_n around some $\theta_0 \in \Theta$ contains at least one point of the support, and hence has prior mass at least $1/N^n$.

It follows that the entropy and prior mass conditions (4.18) and (4.19) are satisfied if

$$\begin{aligned} (1/\varepsilon_n)^{1/\alpha} &\le K(\varepsilon_n/\sigma_n)^2, \\ \exp\bigl(-(1/\varepsilon_n)^{1/\alpha}\bigr) &\ge e^{-\varepsilon_n^2/\sigma_n^2}. \end{aligned}$$

(Bound the numerator of (4.19) by one.) This is satisfied for $\varepsilon_n = \sigma_n^{2\alpha/(2\alpha+1)}$. If the parameters are also uniformly Lipschitz, then the rate of convergence relative to the semimetrics h_n/σ_n or d is $\sigma_n^{2\alpha/(2\alpha+1)}$.

Priors based on wavelet expansions

Consider as parameter space Θ the set of all functions $\theta : [-M, M] \to \mathbb{R}$ with a bounded αth derivative, for some given natural number α. This parameter set is contained in the Besov space $B^{\alpha}_{\infty,\infty}$ and therefore we can represent every parameter θ in a suitable orthonormal wavelet basis $(\psi_{j,k} : j \in \mathbb{N}, k = 1, \ldots, 2^j)$ in the form

$$\theta(x) = \sum_{j=1}^{\infty} \sum_{k=1}^{2^j} \theta_{j,k} \psi_{j,k}(x),$$

where the Fourier coefficients $\theta_{j,k}$ satisfy

$$\|\theta\|^{\alpha}_{\infty,\infty} := \sup_j 2^{j\alpha} 2^{j/2} \max_k |\theta_{j,k}| < \infty.$$

A prior on Θ can be defined structurally as

$$\theta \stackrel{d}{=} \sum_{j=1}^{J} \sum_{k=1}^{2^j} \delta_j Z_{j,k} \psi_{j,k},$$

where $J = J_n$ is chosen dependent on n at a rate to be determined later, δ_j are constants, and $(Z_{j,k} : j \in \mathbb{N}, k = 1, \ldots, 2^j)$ are i.i.d. standard normal random variables.

We shall show that if $2^{J_n} \sim \sigma_n^{-2/(2\alpha+1)}$ and $\delta_j = 2^{-j/2}$, then the Bayesian rate of convergence relative to the semimetrics h_n/σ_n or d is equal to $\sigma_n^{2\alpha/(2\alpha+1)}$ up to a logarithmic factor. The logarithmic factor is possibly a defect of our proof. The rate $\sigma_n^{2\alpha/(2\alpha+1)}$ is known to be the sharp estimation rate for non-Bayesian procedures, and hence can also not be improved in the present context.

We derive the rate from Theorem 4.5, setting Θ^n equal to the set of functions $\theta = \sum_{j=1}^{J} \sum_k \theta_{j,k} \psi_{j,k}$ with coefficients $\theta_{j,k}$ bounded absolutely by $M_{j,n} := \delta_j 2^{J/2} a_n$ for $k = 1, \ldots, 2^j$ and $\{a_n\}$ a sequence of positive numbers. Then

$$\begin{aligned} \Pi^n(\Theta \setminus \Theta^n) &= \Pr\big(\exists j, k : |\delta_j Z_{j,k}| > M_{j,n}\big) \le \sum_{j=1}^{J} 2^j 2\big(1 - \Phi(M_{j,n}/\delta_j)\big) \\ &\le \sum_{j=1}^{J} 2^{j+1} e^{-\frac{1}{2} M^2_{j,n}/\delta_j^2} \le 2^{J+2} e^{-2^J a_n^2/2}. \end{aligned}$$

We may then use Lemma 4.3 to show that (by an appropriate choice of the numbers $\{a_n\}$) the posterior mass within $\Theta \setminus \Theta^n$ is negligible, and concentrate on the posterior mass inside Θ^n.

The uniform norm of a function θ in the Besov space $B^{\alpha}_{\infty,\infty}$ is equivalent to the norm

$$\|\theta\|_{\infty} = \sum_{j=1}^{\infty} 2^{j/2} \max_k |\theta_{j,k}|,$$

on the Fourier coefficients of the function. If the true parameter θ_0 is contained in $B^{\alpha}_{\infty,\infty}$, then the uniform distance between θ_0 and its projection $\theta_0^J := \sum_{j=1}^{J} \sum_k \theta_{0;j,k} \psi_{j,k}$ on the space spanned by the wavelets of resolution up to J satisfies

$$\|\theta_0 - \theta_0^J\|_\infty = \sum_{j>J} 2^{j/2} \max_k |\theta_{0;j,k}| \le \sum_{j>J} \|\theta_0\|_{\infty,\infty}^\alpha 2^{-j\alpha} \le 2^{-J\alpha} \|\theta_0\|_{\infty,\infty}^\alpha.$$

See also Section 9.5 of Härdle et al. (1998), in particular formulas (9.34) - (9.35). By the triangle inequality it follows that for $2^{-J\alpha}\|\theta_0\|_{\infty,\infty}^\alpha < \varepsilon_n$,

$$\begin{aligned}
\Pi^n\bigl(\theta \in \Theta^n : \|\theta - \theta_0\|_\infty \le 2\varepsilon_n\bigr) \;&\ge\; \Pr\Bigl(\sum_{j=1}^J 2^{j/2} \max_k |\delta_j Z_{j,k} - \theta_{0;j,k}| \le \varepsilon_n\Bigr)\\
&\ge\; \prod_{j=1}^J \Pr\Bigl(2^{j/2} \max_k |\delta_j Z_{j,k} - \theta_{0;j,k}| \le \varepsilon_n / J\Bigr)\\
&\ge\; \prod_{j=1}^J \prod_k \Bigl[e^{-\theta_{0;j,k}^2/\delta_j^2} \frac{1}{\sqrt{2}} \Pr\Bigl(\frac{1}{\sqrt{2}}|Z_{j,k}| \le \frac{\varepsilon_n}{J 2^{j/2}\delta_j}\Bigr)
\end{aligned}$$

In the last step we use that the $N(\theta,1)$ density is bounded below by $e^{-\theta^2}/\sqrt{2}$ times the $N(0,1/2)$ density, so that $\Pr\bigl(|Z-\theta| \le \varepsilon\bigr) \ge (e^{-\theta^2}/\sqrt{2}) \Pr\bigl(|Z|/\sqrt{2} \le \varepsilon\bigr)$. For $\varepsilon_n/(J2^{j/2}\delta_j)$ bounded above, the right side is bounded below by, for some positive constant C,

$$\begin{aligned}
&C^{2^J} \exp\Bigl(-\sum_{j=1}^J \sum_k \frac{\theta_{0;,j,k}^2}{\delta_j^2}\Bigr) \prod_{j=1}^J \Bigl(\frac{\varepsilon_n}{J 2^{j/2}\delta_j}\Bigr)^{2^j}\\
&\quad \ge C^{2^J} \exp\Bigl(-\sum_{j=1}^J \frac{2^{-2j\alpha}}{\delta_j^2} \bigl(\|\theta_0\|_{\infty,\infty}^\alpha\bigr)^2\Bigr) \exp\Bigl(-\sum_{j=1}^J 2^j \log\Bigl(\frac{J 2^{j/2}\delta_j}{\varepsilon_n}\Bigr)\Bigr).
\end{aligned}$$

We shall use these estimates to verify the prior mass condition (4.19).

To compute the entropy of Θ^n we choose for each fixed j a minimal $2^{(j/2-J)} M_{j,n}\varepsilon/a_n$-net over the interval $[-M_{j,n}, M_{j,n}]^{2^j} \subset \mathbb{R}^{2^j}$ for the maximum norm on $\mathbb{R}^{2^j}$, and form a net over Θ^n by forming arrays $\theta = (\theta_{j,k})$ with the coefficients $(\theta_{j,1}, \ldots, \theta_{j,2^j})$ at each level $j \in \{1, \ldots, J\}$ chosen equal to an arbitrary element of the net over $[-M_{j,n}, M_{j,n}]^{2^j}$, and $\theta_{j,k} = 0$ for $j > J$. The logarithm of the total number of points θ is bounded by

$$\log \prod_{j=1}^J \Bigl(\frac{3 M_{j,n} a_n}{2^{(j/2-J)} M_{j,n}\varepsilon}\Bigr)^{2^j} \;\le\; \sum_{j=1}^J 2^j \Bigl(\log \frac{3a_n}{\varepsilon} + (J - j/2)\Bigr) \lesssim 2^J \Bigl(\log \frac{1}{\varepsilon} + \log a_n + J\Bigr).$$

The uniform distance of an arbitrary point $\theta \in \Theta^n$ to the net is bounded above by

$$\sum_{j=1}^J 2^{j/2} 2^{(j/2-J)} M_{j,n}\varepsilon/a_n = \varepsilon 2^{-J/2} \sum_{j=1}^J 2^j \delta_j.$$

If the right side is bounded by ε, then it follows that the ε-entropy of Θ^n for the uniform norm is bounded above by $2^J\bigl(\log \frac{1}{\varepsilon} + J\bigr)$.

Combining the preceding with Lemma 4.3 and Theorem 4.5, we see that the rate of convergence relative to the semimetrics h_n/σ_n or d is equal to ε_n if the following inequalities are satisfied:

$$\begin{aligned}
\sum_{j=1}^{J} 2^j \log\Big(\frac{J2^{j/2}\delta_j}{\varepsilon_n}\Big) + \sum_{j=1}^{J} \frac{2^{-2j\alpha}}{\delta_j^2} &\leq \frac{\varepsilon_n^2}{\sigma_n^2}, \\
\frac{\varepsilon_n}{J2^{j/2}\delta_j} &\lesssim 1, \\
2^{-J\alpha} &\lesssim \varepsilon_n \\
2^J\Big(\log\frac{1}{\varepsilon_n} + \log\alpha_n + J\Big) &\leq K\frac{\varepsilon_n^2}{\sigma_n^2}, \\
2^{-J/2}\sum_{j=1}^{J} 2^j\delta_j &\lesssim 1.
\end{aligned}$$

The first three conditions ensure that the prior-mass condition is satisfied, whereas the fourth and the fifth condition take care of the entropy condition. It can be verified that the above inequalities are satisfied for $2^J \sim \sigma_n^{-2/(2\alpha+1)}$ and $\varepsilon_n = \sigma_n^{2\alpha/(2\alpha+1)} \log(1/\sigma_n)$ if $a_n = \big(\log\sigma_n\big)^2$.

4.3.3 Ergodic diffusion

In this subsection we consider the SDE

$$dX_t = \theta_0(X_t)\,dt + \sigma(X_t)\,dB_t, \quad t \leq T_n,$$

for a given measurable function σ. Under regularity conditions (see e.g. Karatzas and Shreve (1991), Section 5.5), this equation generates a strong Markov process on an interval $I \subseteq \mathbb{R}$, with scale function s_{θ_0} given by

$$s_{\theta_0}(x) = \int_{x_0}^{x} \exp\left(-2\int_{x_0}^{y} \frac{\theta_0(z)}{\sigma^2(z)}\,dz\right)\,dy$$

(x_0 is an arbitrary, but fixed point in the state space) and speed measure

$$m_{\theta_0}(dx) = \frac{dx}{s'_{\theta_0}(x)\sigma^2(x)}.$$

We assume that m_{θ_0} has finite total mass, i.e. $m_{\theta_0}(I) < \infty$. Then the diffusion is ergodic, and the normalized speed measure $\mu_0 = m_{\theta_0}/m_{\theta_0}(I)$ is the unique invariant probability measure. For simplicity, the initial law of the diffusion is supposed to be degenerate in some point $x \in I$. The endpoint T_n of the observation interval is assumed to tend to infinity as $n \to \infty$. The parameter set Θ is a collection of real functions on the interval I which we will specify below.

In this model the square of the Hellinger semimetric h_n in (4.5) is given by

$$h_n^2(\theta,\psi) = \int_0^{T_n} \left(\frac{\theta(X_t) - \psi(X_t)}{\sigma(X_t)}\right)^2\,dt.$$

Using the occupation time formula $\int_0^t f(X_s)\,ds = \int_I f l_t\,dm_{\theta_0}$ we can rewrite this semimetric in terms of the diffusion local time $(l_t(x), t \geq 0, x \in I)$ of the process X relative to its speed measure m_{θ_0} (cf. e.g. Itô and McKean (1965)), as follows

$$h_n^2(\theta,\psi) = \int_I \left(\frac{\theta(x)-\psi(x)}{\sigma(x)}\right)^2 l_{T_n}(x)\,dm_{\theta_0}(x).$$

An immediate consequence is that for any interval $I^* \subset I$

$$m_{\theta_0}(I) \inf_{x\in I^*} l_{T_n}(x) \left\|\frac{\theta-\psi}{\sigma}\mathbf{1}_{I^*}\right\|^2_{L^2(\mu_0)} \leq h_n^2(\theta,\psi) \leq m_{\theta_0}(I)\sup_{x\in I} l_{T_n}(x)\left\|\frac{\theta-\psi}{\sigma}\right\|^2_{L^2(\mu_0)}. \quad (4.22)$$

In case I^* is compact, the infimum and supremum over the scaled local time $(1/T_n)l_{T_n}$ are appropriately bounded away from zero and infinity (see the proof below) and we can choose

$$d_n(\theta,\psi) = \sqrt{T_n}\left\|\frac{\theta-\psi}{\sigma}\mathbf{1}_{I^*}\right\|_{L^2(\mu_0)} \quad \text{and} \quad \bar{d}_n(\theta,\psi) = \sqrt{T_n}\left\|\frac{\theta-\psi}{\sigma}\right\|_{L^2(\mu_0)}.$$

Let $\bar{d}(\theta,\psi) = \bar{d}_n(\theta,\psi)/\sqrt{T_n} = \|(\theta-\psi)\sigma^{-1}\|_{L^2(\mu_0)}$.

Theorem 4.6 *Let ε_n be a sequence of positive numbers such that $T_n\varepsilon_n^2$ is bounded away from zero. Let I^* be a compact subinterval of I. Suppose that for every $a > 0$ there exists a constant $K < \infty$ such that*

$$\sup_{\varepsilon>\varepsilon_n} \log N\big(a\varepsilon, \{\theta \in \Theta : \|(\theta-\theta_0)\mathbf{1}_{I^*}/\sigma\|_{L_2(\mu_0)} < \varepsilon\}, \bar{d}\big) \leq KT_n\varepsilon_n^2. \quad (4.23)$$

Furthermore, assume that for every $\xi > 0$ there is a constant J such that for $j \geq J$

$$\frac{\Pi^n\big(\theta \in \Theta : \|(\theta-\theta_0)\mathbf{1}_{I^*}/\sigma\|_{L_2(\mu_0)} < j\varepsilon_n\big)}{\Pi^n\big(\theta \in \Theta : \|(\theta-\theta_0)/\sigma\|_{L_2(\mu_0)} < \varepsilon_n\big)} \leq e^{\xi T_n \varepsilon_n^2 j^2}. \quad (4.24)$$

Then for every $M_n \to \infty$, we have that

$$P^{\theta_0,n}\Pi^n\big(\theta \in \Theta^n : \|(\theta-\theta_0)\mathbf{1}_{I^*}/\sigma\|_{L_2(\mu_0)} \geq M_n\varepsilon_n | X^n\big) \to 0. \quad (4.25)$$

Proof The assertion follows from Theorem 4.2 once it has been established that Assumption 4.1 is satisfied for $d_n := \sqrt{T_n}d$ and $\bar{d}_n := \sqrt{T_n}\bar{d}$, where d and $\bar{d}$ are the L_2-metrics appearing on the left and right side of (4.22).

Now, according to Theorems 3.1 and 3.2 of Van Zanten (2003) respectively, it holds that, with $M = m_{\theta_0}(I)$,

$$\sup_{x\in I} l_{T_n}(x) = O_P(T_n)$$
$$\sup_{x\in I^*}\left|\frac{1}{T_n}l_{T_n}(x) - \frac{1}{M}\right| \xrightarrow{\text{P}} 0.$$

Hence for $\gamma > 0$ there exists a constant $C = C_\gamma > 0$ such that

$$P\Big(\frac{1}{T_n}\sup_{x\in I} l_{T_n}(x) \leq C\Big) \geq 1-\gamma,$$

and we have that

$$P\Big(\inf_{x\in I^*}\frac{1}{T_n}l_{T_n}(x)\geq\frac{1}{2M}\Big)\geq P\Big(\sup_{x\in I^*}\Big|\frac{1}{T_n}l_{T_n}(x)-\frac{1}{M}\Big|\leq\frac{1}{2M}\Big)\to 1.$$

Therefore, the events

$$U^n=\{1/(2M)\leq(1/T_n)l_{T_n}(x)\;\forall x\in I^*\}\cap\{(1/T_n)l_{T_n}(x)\leq C\;\forall x\in I\}$$

have probability satisfying $\liminf_{n\to\infty}P(U^n)\geq 1-\gamma$, and on U^n we have $\frac{1}{2}d_n^2\leq h_n^2\leq MC\bar{d}_n^2$ for all $\theta,\psi\in\Theta^n$. Thus Assumption 4.1 is satisfied with $\mu_n=1$. □

From a modelling perspective the most interesting case is that the state space I of the diffusion is a bounded open interval. Then we shall never observe the full state space in finite time, as the sample paths $t\mapsto X_t$ are continuous functions with range strictly within the state space. A model will specify the parameters $\theta: I\to\mathbb{R}$ on an interval containing the range of the observed sample path. (Note that correspondingly the preceding theorem gives consistency of the estimator on compact subintervals of the state space only.) These parameters should also be specified so that the resulting diffusion equation possesses an ergodic solution that remains within the interval. The most interesting (and simplest) case is that the diffusion function σ is strictly positive on the state space I and tends to zero at its boundaries, so that the diffusion part of the differentials dX_t become negligible as the sample path $t\mapsto X_t$ approaches the boundary. The drift parameters θ should then be positive near the left boundary of I and negative near the right boundary, so that the differentials dX_t become positive and negative at these two boundaries, thus deflecting the sample path near the boundaries of the state space.

Following Liptser and Shiryayev (1977) we give conditions that make the preceding precise and ensure that the conditions at the beginning of Section 4.2 are satisfied. After that, we discuss examples of prior distributions. For simplicity of notation we take the state space equal to the open unit interval $I=(0,1)$. We assume that the drift function $\sigma:(0,1)\mapsto$ is strictly positive and Lipschitz, with, for some numbers $p,q\geq 0$,

$$\sigma(x)\sim x^{1+p},\quad\text{as } x\downarrow 0,\qquad\sigma(x)\sim(1-x)^{1+q},\quad\text{as } x\uparrow 1.$$

Then the diffusion equation

$$dX_t=\theta(X_t)\,dt+\sigma(X_t)\,dB_t,\quad t\leq T_n,\qquad X_0=x_0$$

possesses a unique strong solution X for any initial value $x_0\in(0,1)$ for any Lipschitz function $\theta:(0,1)\to\mathbb{R}$ that is positive and bounded away from zero in a neighborhood of 0 and negative and bounded away from zero in a neighborhood of 1. The corresponding scale function s_θ can be seen to satisfy $s_\theta(x)\to-\infty$ as $x\downarrow 0$ and $s_\theta(x)\to\infty$ as $x\uparrow 1$ and hence maps I onto $\mathbb{R}$ (Proposition 5.22(a) in Karatzas and Shreve (1991)). It follows that the diffusion X is recurrent on the state space I with speed measure m_θ that has a continuous density, which is bounded by

$$\frac{C_1}{x^{2+2p}}e^{-C_2x^{-1-2p}}\qquad\text{and}\qquad\frac{C_1}{(1-x)^{2+2q}}e^{-C_3(1-x)^{-1-2q}}$$

near 0 and 1, respectively. Here C_1, C_2 and C_3 are positive constants. In particular, the speed measure m_θ is finite, so that the diffusion is positive recurrent and ergodic. We also have that $\int_0^1 \sigma^{-2}(x)\, dm_\theta(x) < \infty$, so that

$$\int_0^{T_n} \Big(\frac{\vartheta}{\sigma}\Big)^2 (X_t)\, dt \le \sup_x l^{\vartheta}_{T_n}(x) \int_0^t \Big(\frac{\vartheta}{\sigma}\Big)^2 (x)\, dm_\theta(x) < \infty,$$

for any bounded function $\vartheta : (0,1) \to \mathbb{R}$. According to theorems 7.19 and 7.20 of Liptser and Shiryayev (1977) the induced distributions $P^{\theta,n}$ on the Borel sets of $C[0,T_n]$ of the solutions are equivalent, and their likelihood process is given by (4.3).

Thus for a diffusion function σ as given we obtain a valid statistical model for the parameter set Θ equal to the set of Lipschitz functions $\theta : [0,1] \to \mathbb{R}$ that are positive and bounded away from zero near 0, and negative and bounded away from zero near 1. In the following sections we discuss examples of priors on this parameter set.

Monotone drift functions

Let the parameter set Θ be the set of all monotone, Lipschitz functions $\theta : [0,1] \to \mathbb{R}$ with $\theta(0) > 0$ and $\theta(1) < 0$. Given a finite measure α with a continuous positive density on $(0,1)$ and a positive integer L we define a prior on this parameter set through the following steps:

- $\big(D(1/L), D(2/L) - D(1/L), \ldots, D(1) - D(1-1/L)\big)$ is Dirichlet distributed on the unit simplex in $\mathbb{R}^L$ with parameter vector $\big(\alpha(0,1/L], \alpha(1/L,2/L], \ldots, \alpha(1-1/L,1)\big)$.
- D is extended to a function $D : (0,1) \to [0,1]$ by setting $D(0) = 0$, $D(1) = 1$, and linear interpolation on the intervals $\big((j-1)/L, j/L\big]$.
- U and V are independent random variables, both independent of D. U is uniformly distributed on $(0,1)$ and V has a distribution on $[0,\infty)$ with bounded, strictly positive density such that $P(V \ge v) \le e^{-e^v}$ for large values of v.
- $\theta \stackrel{d}{=} VU - VD$.

Thus D is a random distribution function on $(0,1)$ that is reflected $(-D)$ shifted up to cross the horizontal axis at a random location $(U - D)$ and finally scaled by multiplication with V.

We shall now show that for any $\theta_0 \in \Theta$ the rate of convergence relative to the L_2-metric on a compact subinterval $I^* \subset I$ is at least $T_n^{-1/3} \log T_n$.

We apply Theorem 4.6 with Θ^n equal to $\{\theta \in \Theta : \|\theta\|_\infty \le K_n\}$ for $K_n \sim (\log T_n)^2$. Because a function $VU - VD$ decreases from VU at 0 to $V(U-1)$ at 1, its absolute value can take values larger than K only if $V \ge K$. Consequently, for n sufficiently large

$$\Pi^n(\Theta \setminus \Theta^n) \le \Pr(V \ge K_n) \le e^{-e^{K_n}}.$$

With the help of Lemma 4.3 we shall be able to discard this part of the prior.

The set Θ^n consists of monotone functions $\theta : [0,1] \to [-K_n, K_n]$. The measure Q_0 defined by $dQ_0(x) = \sigma^{-2}(x)\, d\mu_0(x)$ is finite. Therefore the ε-entropy of Θ^n relative to the $L_2(Q)$-semimetric is bounded above by a multiple of K_n/ε. (see Theorem 2.7.5 in Van der Vaart and Wellner (1996).)

To lower bound the prior mass of a neighborhood of θ_0, we first note that, by the triangle inequality, with $D_0 = (\theta_0(0) - \theta_0)/\big(\theta_0(0) - \theta_0(1)\big)$,

$$\|VU - VD - \theta_0\|_\infty \le \big|VU - \theta_0(0)\big| + \Big\|D_0 - D\Big\|_\infty \big(\theta_0(0) - \theta_0(1)\big) + \big|\theta_0(0) - \theta_0(1) - V\big|.$$

Here $\theta_0(0)$ and $\theta_0(0)-\theta_0(1)$ are positive numbers by assumption, and hence the probability of the intersection of the events that $\big|VU - \theta_0(0)\big| < \varepsilon$ and $\big|\theta_0(0) - \theta_0(1) - V\big| < \varepsilon$ is of the order ε^2 as $\varepsilon \downarrow 0$. By Lemma 3 in Ghosal and Van der Vaart (2003) we also have that, for $J\varepsilon \le 1$ and positive constants c and C,

$$\Pr\Big(\sum_{j=1}^{L}\Big|D\Big(\frac{j-1}{L}, \frac{j}{L}\Big] - p_j\Big| < \varepsilon\Big) \ge Ce^{-cL\log(1/\varepsilon)},$$

uniformly in $(p_1, \dots, p_L)$ in the unit simplex. The function D_0 is the cumulative distribution of a probability distribution on $(0, 1)$ and is Lipschitz. It can be seen that

$$\|D_0 - D\|_\infty \le \|D_0\|_{Lip}\frac{1}{L} + \sum_{j=1}^{L}\Big|D\Big(\frac{j-1}{L}, \frac{j}{L}\Big] - D_0\Big(\frac{j-1}{L}, \frac{j}{L}\Big]\Big|.$$

Here, the Lipschitz-norm of a function f is defined by $\|f\|_{Lip} = \sup_{x\neq y}|f(x)-f(y)|/|x-y|$. Combining these facts it follows that

$$\begin{aligned}
&\Pi^n\big(\theta \in \Theta \ : \ \|\theta - \theta_0\|_\infty \le 3\varepsilon\big) \\
&\quad \ge \Pr\big(\big|VU - \theta_0(0)\big| < \varepsilon, \big|\theta_0(0) - \theta_0(1) - V\big| < \varepsilon\big)\Pr\big(\|D_0 - D\|_\infty < \varepsilon\big) \\
&\quad \gtrsim \varepsilon^2 e^{-cJ\log(1/\varepsilon)}.
\end{aligned}$$

If we choose $J \sim T_n^{1/3}\log T_n$, $K_n = (\log T_n)^2$, and $\varepsilon_n \sim T_n^{-1/3}\log T_n$, then the entropy and prior mass conditions are satisfied.

Parametric models

Consider the ergodic diffusion model with the drift function taking a parametric form. We shall denote the parameter again as θ and write the drift function in the form β^θ. Thus the process X satisfies the SDE

$$dX_t = \beta^\theta(X_t)\,dt + \sigma(X_t)\,dB_t,$$

for a given measurable function σ.

Let the parameter θ range over a subset of k-dimensional Euclidean space $(\mathbb{R}^k, \|\cdot\|)$, and assume that there exist functions $\underline{\beta}$ and $\bar\beta$ satisfying

$$0 < \int_{I^*}\big(\underline{\beta}/\sigma\big)^2\,d\mu_0, \qquad \int_I \big(\bar\beta/\sigma\big)^2\,d\mu_0 < \infty,$$

and such that, for all $x \in I$ and all $\theta, \psi \in \Theta$,

$$\underline{\beta}(x)\,\|\theta - \psi\| \le |\beta^\theta(x) - \beta^\psi(x)| \le \bar\beta(x)\,\|\theta - \psi\|.$$

For our purpose it suffices that the first inequality be satisfied for $x \in I^* \subseteq I$.

Under this assumption the entropy and prior mass conditions of Theorem 4.6 can be expressed in corresponding ones with respect to Euclidean distance, and we obtain the following corollary.

Corollary 4.7 *Let the prior Π^n be independent of n and possess a Lebesgue density that is bounded and bounded away from zero on a neighborhood of θ_0. Let functions $\underline{\beta}$ and $\bar{\beta}$ as in the preceding exist. Then for every $M_n \to \infty$, we have, as $n \to \infty$,*

$$P^{\theta_0,n}\Pi^n\Big(\theta \in \Theta^n : \|\theta - \theta_0\| \geq M_n/\sqrt{T_n} | \, X^n\Big) \to 0. \tag{4.26}$$

Proof The assumptions imply the existence of positive constants L, U such that

$$L\|\theta - \psi\| \leq \Big\|\frac{\beta^\theta - \beta^\psi}{\sigma}\mathbf{1}_{I^*}\Big\|_{L_2(\mu_0)} \leq \Big\|\frac{\beta^\theta - \beta^\psi}{\sigma}\Big\|_{L_2(\mu_0)} \leq U\|\theta - \psi\|.$$

These inequalities allow to perform the calculations for Theorem 4.6 using Euclidean balls and distances.

First the bounds imply that the left side of (4.23) is bounded above by

$$\sup_{\varepsilon > \varepsilon_n} \log N\big(a\varepsilon/U, \{\theta : \|\theta - \theta_0\| \leq \varepsilon/L\}, \|\cdot\|\big) \leq k \log\Big(\frac{5\varepsilon/L}{a\varepsilon/U}\Big),$$

(cf. Pollard (1990), Lemma 4.1) which is bounded above by a constant, independently of ε.

Secondly, the comparison of norms shows that the quotient in the left side of (4.24) is bounded above by

$$\frac{\Pi^n\big(\theta \in \Theta^n : \|\theta - \theta_0\| \leq j\varepsilon/L\big)}{\Pi^n\big(\theta \in \Theta^n : \|\theta - \theta_0\| \leq \varepsilon/U\big)} \leq \frac{M}{m}\Big(\frac{j\varepsilon/L}{\varepsilon/U}\Big)^k = \frac{M}{m}\Big(\frac{jU}{L}\Big)^k,$$

where m and M are lower and upper bounds on the density of the prior. □

4.4 Proofs

4.4.1 Proof of Theorem 4.2

For given μ_n and $M_n \to \infty$ denote by U^n the random set

$$U^n = \big\{\theta \in \Theta^n : h_n(\theta, \theta_0) \geq M_n\mu_n\big\}.$$

For given positive constants c, C, D define events

$$\begin{aligned}\bar{A}_{n,C,D} &:= \big\{\omega : h_n(\theta,\psi)(\omega) \leq C\bar{d}_n(\theta,\psi),\ \forall\, \theta,\psi \in \Theta^n \text{ with } h_n(\theta,\psi)(\omega) \geq D\mu_n\big\},\\ A_{n,c,D} &:= \big\{\omega : h_n(\theta,\theta_0)(\omega) \geq c\, d_n(\theta,\theta_0)),\ \forall\, \theta \in \Theta^n \text{ with } h_n(\theta,\theta_0)(\omega) \geq D\mu_n\big\}.\end{aligned}$$

According to Assumption (4.1) there exist positive constants c, C, D such that the events $A_{n,c,D} \cap \bar{A}_{n,C,D}$ have probability arbitrarily close to one as $n \to \infty$. It therefore suffices to show that the sequence $P^{\theta_0,n}\Pi^n\big(U^n | \, X^n\big)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}$ tends to zero for fixed positive constants c, C, D. Furthermore, if the constants c_γ, C_γ in Assumption (4.1) satisfy $\inf_{\gamma>0} c_\gamma/C_\gamma \geq a_0 > 0$ and/or $\inf_{\gamma>0} c_\gamma \geq c_0 > 0$, then it suffices to consider c, C, D satisfying these restrictions only.

In Lemma 4.8 we construct test functions $\varphi^n : \Omega \to [0,1]$ that are consistent for the null hypothesis $H_0 : \theta = \theta_0$, i.e. $P^{\theta_0,n}\varphi^n \to 0$ as $n \to \infty$. Since $1 = \varphi^n + (1-\varphi^n)$, we can bound

$$P^{\theta_0,n}\Pi^n(U^n|X^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}} \le P^{\theta_0,n}\varphi^n + P^{\theta_0,n}\Pi^n(U^n|X^n)(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}. \quad (4.27)$$

Here the first term on the right tends to zero by consistency, and hence it suffices to concentrate on the second term. We rewrite the posterior distribution (4.4) as

$$\Pi^n(B|X^n) = \frac{\int_B p^{\theta,n}/p^{\theta_0,n}(X^n)\, d\Pi^n(\theta)}{\int_{\Theta^n} p^{\theta,n}/p^{\theta_0,n}(X^n)\, d\Pi^n(\theta)}, \quad B \in \mathcal{B}^n. \quad (4.28)$$

The set of interest is the union $U^n = \cup_{i\ge M_n}\Theta_i^n$ of the random annuli defined by

$$\Theta_i^n = \{\theta \in \Theta^n : i\mu_n \le h_n(\theta,\theta_0) < (i+1)\mu_n\}, \quad i \in \mathbb{N}.$$

Therefore, we can bound the second term on the right in (4.27) by

$$\sum_{i\ge M_n} P^{\theta_0,n}\left[\frac{\int_{\Theta_i^n} p^{\theta,n}/p^{\theta_0,n}\, d\Pi^n(\theta)}{\int p^{\theta,n}/p^{\theta_0,n}\, d\Pi^n(\theta)}(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}\right]. \quad (4.29)$$

The main part of the proof is to construct the test functions in such a way that the terms in this sum are small. Here we bound the denominator from below by a constant, and use Fubini's theorem to bound

$$P^{\theta_0,n}\int_{\Theta_i^n}\frac{p^{\theta,n}}{p^{\theta_0,n}}\, d\Pi^n(\theta)(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}} \le \int P^{\theta,n}\mathbf{1}_{\{\theta\in\Theta_i^n\}}(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}\, d\Pi^n(\theta). \quad (4.30)$$

The following two lemmas assert the existence of appropriate test functions φ^n, and give the lower bound on the denominator.

Lemma 4.8 *If condition (4.7) holds, then for all positive constants μ_n, c, C, D and sufficiently large integer I there exists a test φ^n (depending on μ_n and c, C, D, I) such that*

$$P^{\theta_0,n}\varphi^n \le \exp\Big(\mu_n^2 g\Big(\frac{c}{8C}\Big)\Big)\sum_{i\ge I} e^{-i^2\mu_n^2/512}, \quad (4.31)$$

and for all $i \ge I$

$$P^{\theta,n}(1-\varphi^n)\mathbf{1}_{\{\theta\in\Theta_i^n\}}\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}} \le e^{-i^2\mu_n^2/1152}. \quad (4.32)$$

Lemma 4.9 *For every $\varepsilon > 0$ and $K > 0$,*

$$P^{\theta_0,n}\left(\int \frac{p^{\theta,n}}{p^{\theta_0,n}}\, d\Pi^n(\theta) \le e^{-(\frac{1}{2}(C^2\varepsilon^2\vee D^2\mu_n^2)+K\varepsilon^2)}\Pi^n\big(\bar{B}^n(\theta_0,\varepsilon)\big), \bar{A}_{n,C,D}\right) \le \exp\left(-\frac{K^2\varepsilon^4}{2(C^2\varepsilon^2\vee D^2\mu_n^2)}\right).$$

The proofs of these lemmas are deferred to the next sections. We first proceed with the proof of the main theorem. Choose $I = M_n \to \infty$ and let φ^n be tests as in Lemma 4.8.

Since $g(c/8C) < \infty$, assertion (4.31) of Lemma 4.8 implies that $P^{\theta_0,n}\varphi^n \to 0$ if μ_n is bounded away from zero and $I = I_n \to \infty$.

By Lemma 4.9, applied with $\varepsilon = \mu_n$, the expression (4.29) can be bounded by

$$\sum_{i \geq M_n} P^{\theta_0,n} \frac{\int_{\Theta_i^n} p^{\theta,n}/p^{\theta_0,n}\, d\Pi^n(\theta)}{e^{-(\frac{1}{2}(C\vee D)^2+K)\mu_n^2}\Pi^n\big(\bar{B}^n(\theta_0,\mu_n)\big)}(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap \bar{A}_{n,C,D}} + e^{-K^2\mu_n^2/(2(C\vee D)^2)}.$$

The second term can be made arbitrarily small by choice of K. The first term can be handled using Fubini's theorem as in (4.30), and inequality (4.32). Here, since $\Theta_i^n(\omega) \subset B^n(\theta_0, 2i\mu_n/c)$ if $\omega \in A_{n,c,D}$ and $i \geq D\vee 2$, we may restrict the integral to the (nonrandom) set $B^n(\theta_0, 2i\mu_n/c)$. Thus, for n sufficiently large, we obtain the bound

$$\sum_{i \geq M_n} \frac{\Pi^n(B^n(\theta_0, 2\mu_n i/c))}{\Pi^n(\bar{B}^n(\theta_0,\mu_n))} e^{(\frac{1}{2}(C\vee D)^2+K)\mu_n^2 - i^2\mu_n^2/1152}.$$

Taking $\xi = c^2/(8\cdot 1152)$ in condition (4.8), we see that the latter is for sufficiently large n bounded by

$$\sum_{i \geq M_n} \exp\Big\{(\tfrac{1}{2}(C\vee D)^2 + K)\mu_n^2 - \frac{1}{2} i^2\mu_n^2/1152\Big\},$$

which tends to zero, as $M_n \to \infty$, for any fixed C, D, K. This concludes the proof of the main theorem.

4.4.2 Proof of Lemma 4.8

The proof is based on the following version of *Bernstein's inequality*: if M is a continuous local martingale vanishing at 0 with quadratic variation process $[M]$, then, for any stopping time T and all $x, L > 0$,

$$P\Big(\sup_{0\leq t\leq T} |M_t| \geq x, [M]_T \leq L\Big) \leq e^{-x^2/(2L)}$$

(see for instance Revuz and Yor (1999), pp. 153–154). We shall apply this inequality to two local martingales derived from the log likelihood.

First (cf. (4.3) or Section 4.4.6) the log likelihood ratio process can be written as

$$\ell(\theta) := \log \frac{p^{\theta,n}}{p^{\theta_0,n}}(X^n) = M_{T_n}^{\theta,n} - \tfrac{1}{2}[M^{\theta,n}]_{T_n},$$

where $M^{\theta,n}$ is the $P^{\theta_0,n}$-local martingale

$$M_t^{\theta,n} = \int_0^t \left(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\right) dB_s^n, \quad t \geq 0, \quad \theta \in \Theta, \tag{4.33}$$

for B^n a Brownian motion under $P^{\theta_0,n}$. The quadratic variation of $M^{\theta,n}$ at T_n is precisely the square Hellinger semidistance $h_n^2(\theta,\theta_0) = [M^{\theta,n}]_{T_n}$.

Under $P^{\theta,n}$ the process $M^{\theta,n}$ is not a local martingale. However, by Girsanov's theorem the process

$$B_t^{\theta,n} = B_t^n - \int_0^t \Big(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\Big)\, ds$$

is a $P^{\theta,n}$-Brownian motion, and we can write

$$\ell(\theta_1) = Z_{T_n}^{\theta_1,\theta,n} + \tfrac{1}{2}[Z^{\theta_1,\theta,n}]_{T_n} + \int_0^{T_n} \Big(\frac{\beta_t^{\theta_1,n} - \beta_t^{\theta_0,n}}{\sigma_t^n}\Big)\Big(\frac{\beta_t^{\theta,n} - \beta_t^{\theta_1,n}}{\sigma_t^n}\Big)\, dt, \tag{4.34}$$

for the $P^{\theta,n}$-local martingale $Z^{\theta_1,\theta,n}$ defined by

$$Z_t^{\theta_1,\theta,n} = \int_0^t \Big(\frac{\beta_s^{\theta_1,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\Big)\, dB_s^{\theta,n} \quad \theta \in \Theta^n.$$

The quadratic variation of the process $Z^{\theta_1,\theta,n}$ at T_n is again equal to the squared Hellinger semidistance $h_n^2(\theta_1,\theta_0) = [Z^{\theta_1,\theta,n}]_{T_n}$. (The process $Z^{\theta_1,\theta_0,n}$ is equal to the process $M^{\theta_1,n}$ introduced earlier.)

For fixed natural numbers i and n let $\theta_1,\ldots,\theta_N \in \Theta^n$ be a minimal $\mu_n i/(4C)$-net for $\bar{d}_n$ over the set $B^n(\theta_0, 2i\mu_n/c)$. For sufficiently large i we have $2i\mu_n/c \geq \mu_n$ and hence by condition (4.7) the number of points in the net is bounded by

$$N \leq N\Big(\frac{\mu_n i}{4C}, B^n\Big(\theta_0, \frac{2i\mu_n}{c}\Big), \bar{d}_n\Big) \leq \exp\Big(\mu_n^2 g\big(\frac{c}{8C}\big)\Big). \tag{4.35}$$

Define for each $i \in \mathbb{N}$ a deterministic map $\sigma_{ni} : \Theta^n \to \{\theta_1,\ldots,\theta_N\}$ by mapping each $\theta \in B^n(\theta_0, 2i\mu_n/c)$ into a closest point of the net and mapping each other $\theta \in \Theta^n$ in an arbitrary point of the net. For each $\theta \in \Theta^n$ and $i \in \mathbb{N}$ define a test by

$$\varphi_i^{\theta,n} := \mathbf{1}\{\ell(\theta) > 0, i\mu_n/2 < h_n(\theta,\theta_0) < 2i\mu_n\},$$

and set

$$\varphi^n := \sup_{i\geq I} \sup_{\theta\in\sigma_{ni}(\Theta^n)} \varphi_i^{\theta,n},$$

We shall show that the latter tests satisfy (4.31) and (4.32) if I is sufficiently large.

The error of the first kind (4.31) of these tests satisfies

$$P^{\theta_0,n}\varphi^n \leq \sum_{i\geq I} \sum_{\theta\in\sigma_{ni}(\Theta^n)} P^{\theta_0,n}\varphi_i^{\theta,n} \leq \Big(\sup_{i\geq I} \#\sigma_{ni}(\Theta^n)\Big) \sum_{i\geq I} \max_{\theta\in\sigma_{ni}(\Theta^n)} P^{\theta_0,n}\varphi_i^{\theta,n}.$$

Here the cardinality of the sets $\sigma_{ni}(\Theta^n)$ is bounded above in (4.35). The probabilities in the right side of the last display can be bounded with the help of Bernstein's inequality

$$\begin{aligned} P^{\theta_0,n}\varphi_i^{\theta,n} &= P^{\theta_0,n}\Big(M_{T_n}^{\theta,n} - \tfrac{1}{2}h_n^2(\theta,\theta_0) > 0, i\mu_n/2 < h_n(\theta,\theta_0) < 2i\mu_n\Big) \\ &\leq P^{\theta_0,n}\Big(M_{T_n}^{\theta,n} > \tfrac{1}{2}(i\mu_n/2)^2, [M^{\theta,n}]_{T_n} < (2i\mu_n)^2\Big) \leq e^{-i^2\mu_n^2/512}, \end{aligned}$$

uniformly in $\theta \in \Theta^n$. Inserting this bound and the bound (4.35) in the preceding display, we obtain (4.31).

The expectation in (4.32) is restricted to the intersection of the events $\bar{A}_{n,C,D} \cap A_{n,c,D}$ and $\theta \in \Theta_i^n$. By construction of the net $\theta_1, \ldots, \theta_N$,

$$\bar{d}_n\big(\theta, \sigma_{ni}(\theta)\big) \leq \frac{\mu_n i}{4C}, \quad \text{if } \theta \in B^n\Big(\theta_0, \frac{2i\mu_n}{c}\Big).$$

We have $\Theta_i^n(\omega) \subset B^n(\theta_0, 2i\mu_n/c)$ if $\omega \in A_{n,c,D}$ and $i \geq D \vee 2$. Furthermore, if $\omega \in \bar{A}_{n,C,D}$, then either $h_n\big(\theta, \sigma_{ni}(\theta)\big) \leq D\mu_n$ or the Hellinger semimetric is bounded above by $C\bar{d}_n$. It follows that for $i \geq I \geq 4D$, if $\omega \in \bar{A}_{n,C,D} \cap A_{n,c,D}$ and $\theta \in \Theta_i^n(\omega)$, then

$$h_n\big(\theta, \sigma_{ni}(\theta)\big) \leq \frac{\mu_n i}{4}. \tag{4.36}$$

By the triangle inequality it then follows that

$$\frac{3i\mu_n}{4} \leq h_n\big(\theta_0, \sigma_{ni}(\theta)\big) \leq (i + 1 + \tfrac{1}{4}i)\mu_n < 2\mu_n i, \qquad (i \geq 2). \tag{4.37}$$

Therefore, if $\omega \in A_{n,c,D} \cap \bar{A}_{n,C,D}$ and $\theta \in \Theta_{n,i}(\omega)$,

$$1 - \varphi^n \leq 1 - \varphi_i^{\sigma_{ni}(\theta),n} \leq \mathbf{1}\{\ell\big(\sigma_{ni}(\theta)\big) \leq 0\}.$$

We write the log likelihood ratio $\ell\big(\sigma_{ni}(\theta)\big)$ in terms of the process $Z^{\sigma_{ni}(\theta),\theta,n}$ as in (4.34), where by the Cauchy-Schwarz inequality the inner product in (4.34) can be bounded as

$$\begin{aligned}
\Big|\int_0^{T_n} \Big(\frac{\beta_t^{\sigma_{ni}(\theta),n} - \beta_t^{\theta_0,n}}{\sigma_t^n}\Big)\Big(\frac{\beta_t^{\theta,n} - \beta_t^{\sigma_{ni}(\theta),n}}{\sigma_t^n}\Big)\, dt\Big| \;&\leq\; h_n\big(\sigma_{ni}(\theta), \theta_0\big)\, h_n\big(\theta, \sigma_{ni}(\theta)\big) \\
&\leq\; \frac{1}{3} h_n^2\big(\sigma_{ni}(\theta), \theta_0\big),
\end{aligned}$$

for $\omega \in A_{n,c,D} \cap \bar{A}_{n,C,D}$ and $\theta \in \Theta_{n,i}(\omega)$, since $h_n\big(\theta, \sigma_{ni}(\theta)\big) \leq \mu_n i/4 \leq h_n\big(\theta_0, \sigma_{ni}(\theta)\big)/3$ on this event, by (4.36) and (4.37). It follows that the variable $\ell\big(\sigma_{ni}(\theta)\big)$ is bounded below by $Z_{T_n}^{\sigma_{ni}(\theta),\theta,n} + [Z^{\sigma_{ni}(\theta),\theta,n}]_{T_n}/6$, and therefore

$$\begin{aligned}
&P^{\theta,n}(1 - \varphi^n)\mathbf{1}_{\{\theta \in \Theta_i^n\}}\mathbf{1}_{A_{n,c,D} \cap \bar{A}_{n,C,D}} \\
&\quad \leq P^{\theta,n}\big(Z_{T_n}^{\sigma_{ni}(\theta),\theta,n} + [Z^{\sigma_{ni}(\theta),\theta,n}]_{T_n}/6 \leq 0, \{\theta \in \Theta_i^n\}\big) \\
&\quad \leq P^{\theta,n}\Big(\big|Z_{T_n}^{\sigma_{ni}(\theta),\theta,n}\big| \geq \frac{1}{12}\mu_n^2 i^2, [Z^{\sigma_{ni}(\theta),\theta,n}]_{T_n} \leq 4\mu_n^2 i^2\Big) \\
&\quad \leq e^{-i^2\mu_n^2/1152},
\end{aligned}$$

by Bernstein's inequality.

4.4.3 Proof of Lemma 4.9

Let $\tilde{\Pi}^n$ be equal to the measure Π^n restricted and renormalized to be a probability measure on $\bar{B}^n(\theta_0, \varepsilon)$. By Jensen's inequality, with $M^{\theta,n}$ the local martingale in (4.33),

$$\begin{aligned}
\log \int \frac{p^{\theta,n}}{p^{\theta_0,n}} \frac{d\Pi^n(\theta)}{\Pi^n\big(\bar{B}^n(\theta_0, \varepsilon)\big)} \;&\geq\; \int \log \frac{p^{\theta,n}}{p^{\theta_0,n}} d\tilde{\Pi}^n(\theta) \\
&=\; \int \big(M_T^{\theta,n} - \tfrac{1}{2}h_n^2(\theta, \theta_0)\big)\, d\tilde{\Pi}^n(\theta) \\
&\geq\; Z_T^n - \tfrac{1}{2}(C^2\varepsilon^2 \vee D^2\mu_n^2),
\end{aligned} \tag{4.38}$$

on $\bar{A}_{n,C,D}$, where the process Z^n is defined by

$$Z_t^n := \int M_t^{\theta,n}\, d\tilde{\Pi}^n(\theta) = \int_0^t \int \Big(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\Big)\, d\tilde{\Pi}^n(\theta)\, dB_s^{\theta_0}.$$

The last equality follows from the stochastic Fubini theorem (see e.g. Protter (2004), Theorem 64 of Chapter IV). The process Z^n is a continuous local martingale with respect to $P^{\theta_0,n}$ with quadratic variation process

$$[Z^n]_t = \int_0^t \Big(\int \Big(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\Big)\, d\tilde{\Pi}^n(\theta)\Big)^2\, ds,$$

By Jensen's inequality and Fubini's theorem,

$$[Z^n]_T \leq \int_0^T \int \left(\frac{\beta_t^{\theta,n} - \beta_t^{\theta_0,n}}{\sigma_t^n}\right)^2 d\tilde{\Pi}^n(\theta)\, dt = \int h_n^2(\theta, \theta_0)\, d\tilde{\Pi}^n(\theta),$$

Thus $[Z^n]_T \leq C^2\varepsilon^2 \vee D^2\mu_n^2$ on the event $\bar{A}_{n,C,D}$. In view of (4.38) the probability in the lemma is bounded by

$$P^{\theta_0,n}(Z_T^n \leq -K\varepsilon^2, [Z^n]_T \leq C^2\varepsilon^2 \vee D^2\mu_n^2) \leq e^{-K^2\varepsilon^4/(2(C^2\varepsilon^2 \vee D^2\mu_n^2)},$$

by Bernstein's inequality for continuous local martingales.

4.4.4 Proof of Lemma 4.3

By Fubini's theorem and the fact that $P^{\theta_0,n}(p^{\theta,n}/p^{\theta_0,n}) \leq 1$,

$$P^{\theta_0,n}\Big[\int_{\bar{\Theta}^n \setminus \Theta^n} \frac{p^{\theta,n}}{p^{\theta_0,n}}\, d\Pi^n(\theta)\Big] \leq \Pi^n(\bar{\Theta}^n \setminus \Theta^n).$$

By Lemma 4.9 with $\varepsilon = \mu_n$ and arbitrary $K > 0$ on the event $\bar{A}_{n,C,D}$ the denominator of the posterior distribution is bounded below by $e^{-(\frac{1}{2}(C\vee D)^2 + K)\mu_n^2}\Pi^n\big(\bar{B}^n\big(\theta_0, \mu_n\big)\big)$ with probability at least $1 - e^{-K^2\mu_n^2/(2(C\vee D)^2)}$. Choosing $C = C_\gamma$, $D = D_\gamma$, and combining this with the previous display we obtain

$$\begin{aligned} P^{\theta_0,n}\left[\Pi^n(\bar{\Theta}^n \setminus \Theta^n |\, X^n)\mathbf{1}_{\bar{A}_{n,C,D}}\right] &\leq \frac{\Pi^n(\bar{\Theta}^n \setminus \Theta^n)e^{(\frac{1}{2}(C\vee D)^2 + K)\mu_n^2}}{\Pi^n\big(\bar{B}^n(\theta_0, \mu_n)\big)} + e^{-K^2\mu_n^2/(2(C\vee D)^2)} \\ &\leq o(1)e^{(-\frac{1}{2}(C\vee D)^2 + K)\mu_n^2} + e^{-K^2\mu_n^2/(2(C\vee D)^2)}, \end{aligned}$$

by Assumption (4.13).

If $\mu_n \to \infty$, then we choose $K < (C \vee D)^2/2$, and both terms on the right tend to zero. If μ_n remains bounded, then so is the factor $\exp\big(-(C \vee D)^2/2 + K)\mu_n^2\big)$ and hence the first term on the right tends to zero for any fixed K. Furthermore, the second term on the right can be made arbitrarily small by choosing large K in this case.

Thus we have proved the assertion of the lemma on the event $\bar{A}_{n,C,D}$ for each $\gamma > 0$. This suffices, since the probability of this event can be made arbitrarily large by choice of γ.

4.4.5 A technical result

The following lemma is helpful to verify Assumption 4.1. It gives a sufficient condition for Assumption 4.1 with $\mu_n = 1$ (and hence also $\mu_n \to \infty$).

Lemma 4.10 *If h_n and d_n are random semimetrics on a set Θ^n with*

$$\sup_{\theta,\psi\in\Theta^n} \big|h_n(\theta,\psi) - d_n(\theta,\psi)\big| = O_{P^{\theta_0,n}}(1), \tag{4.39}$$

then for all $\gamma > 0$ there exists a positive constant L such that for all $\varepsilon \in (0,1)$

$$\liminf_{n\to\infty} P^{\theta_0,n}\Big(\frac{1}{2-\varepsilon}\, d_n(\theta,\psi) \le h_n(\theta,\psi) \le \frac{1}{\varepsilon} d_n(\theta,\psi) \\ \text{for all } \theta,\, \psi \in \Theta^n \text{ with } h_n(\theta,\psi) \ge L/(1-\varepsilon)\Big) \ge 1-\gamma.$$

Proof For any $\gamma > 0$ there exists a constant $L_\gamma < \infty$ such that on an event with probability at least $1-\gamma$

$$h_n(\theta,\psi) - L_\gamma \le d_n(\theta,\psi) \le h_n(\theta,\psi) + L_\gamma, \quad \forall\, \theta,\psi \in \Theta^n.$$

If $h_n(\theta,\psi) \ge L_\gamma/(1-\varepsilon)$, then on the same event

$$\varepsilon\, h_n(\theta,\psi) \le d_n(\theta,\psi) \le (2-\varepsilon) h_n(\theta,\psi), \quad \forall\, \theta,\psi \in \Theta^n.$$

This is the same event as in the assertion of the lemma. □

4.4.6 Likelihoods

In this section we derive an expression of the likelihood function using Girsanov's theorem. Let B be a Brownian motion on the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \ge 0), \mathrm{Pr})$ and for every $\theta \in \Theta$ let β^θ and σ be adapted processes with

$$\int_0^T |\beta_t^\theta|\, dt < \infty, \qquad \int_0^T |\sigma_t^2|\, dt < \infty.$$

Time is restricted throughout to a compact interval $[0,T]$. Assume that there exists a solution X to the SDE

$$dX_t = \beta_t^{\theta_0}\, dt + \sigma_t\, dB_t, \qquad X_0 = x_0.$$

Here a "solution" is an adapted process X defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathrm{Pr})$ with continuous sample paths that satisfies

$$X_t = x_0 + \int_0^t \beta_s^{\theta_0}\, ds + \int_0^t \sigma_s\, dB_s.$$

Suppose that $\int_0^T (\beta_t^\theta/\sigma_t)^2\, dt$ is finite Pr-almost surely, and assume

$$E \exp\Big(\int_0^T \Big(\frac{\beta_t^\theta - \beta_t^{\theta_0}}{\sigma_t}\Big)\, dB_t - \tfrac{1}{2} \int_0^T \Big(\frac{\beta_t^\theta - \beta_t^{\theta_0}}{\sigma_t}\Big)^2\, dt \Big) = 1. \tag{4.40}$$

Under (4.40) we can define a probability measure $\Pr^\theta$ on $(\Omega, \mathcal{F})$ by

$$\frac{d\Pr^\theta}{d\Pr} = \exp\Big(\int_0^T \Big(\frac{\beta_t^\theta - \beta_t^{\theta_0}}{\sigma_t}\Big)\, dB_t - \tfrac{1}{2} \int_0^T \Big(\frac{\beta_t^\theta - \beta_t^{\theta_0}}{\sigma_t}\Big)^2 dt\Big). \tag{4.41}$$

By Girsanov's theorem the process B^θ given by

$$B_t^\theta = B_t - \int_0^t \Big(\frac{\beta_s^\theta - \beta_s^{\theta_0}}{\sigma_s}\Big) ds$$

is a Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \Pr^\theta)$. By algebra we have

$$dX_t = \beta_t^\theta\, dt + \sigma_t\, dB_t^\theta, \qquad X_0 = x_0.$$

Therefore the process X on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \Pr^\theta)$ is a weak solution of the SDE for θ. Let P^θ be its induced law on $C[0, T]$.

Since by definition $\Pr^\theta \ll \Pr^{\theta_0} = \Pr$ it follows that $P^\theta \ll P^{\theta_0}$. By general considerations of densities

$$\frac{dP^\theta}{dP^{\theta_0}}(X) = E\Big(\frac{d\Pr^\theta}{d\Pr^{\theta_0}} \,|\, X^{-1}(\mathcal{B})\Big),$$

Pr-almost surely on Ω, for $\mathcal{B}$ the Borel sets in $C[0, T]$. If $d\Pr^\theta/d\Pr$ is a measurable function of X, then the conditional expectation on the right is unnecessary and the likelihood of P^θ relative P^{θ_0} evaluated at X is given by (4.41). It can also be written in the form

$$\frac{d\Pr^\theta}{d\Pr} = \exp\Big(\int_0^T \Big(\frac{\beta_t^\theta - \beta_t^{\theta_0}}{\sigma_t^2}\Big)\, dX_t - \frac{1}{2} \int_0^T \Big(\frac{\beta_t^\theta}{\sigma_t}\Big)^2 - \Big(\frac{\beta_t^{\theta_0}}{\sigma_t}\Big)^2 dt\Big). \tag{4.42}$$

The measurability depends on the nature of the processes β^θ and σ and the original filtration.

Condition (4.40) is crucial, but not easy to verify. It is implied by Novikov's condition

$$E \exp\Big(\tfrac{1}{2} \int_0^T \Big(\frac{\beta_t^\theta - \beta_t^{\theta_0}}{\sigma_t}\Big)^2 dt\Big) < \infty.$$

References

Applebaum, D. (2004) *Lévy processes and stochastic calculus*, Cambridge University Press.

Barndorff-Nielsen, O.E. and Shephard, N. (2001) *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, Journal of the Royal Statistical Society, Series B, Vol. **63**, pp. 167-241.

Barndorff-Nielsen, O.E. and Shephard, N. (2001a) *Modelling by Lévy processes for Financial Econometrics*, in "Lévy Processes — Theory and Applications" (eds. Barndorff-Nielsen, O.E., Mikosch T. and Resnick, S), Boston: Birkhäuser.

Barndorff-Nielsen, O.E. and Shephard, N. (2002) *Normal modified stable processes*, Teor. Ĭmovīr. Mat. Stat. **65** pp. 1-19, 2001. Translation in Theory Probab. Math. Statist. **65**, pp 1-20.

Basawa, I.V. and Brockwell, P.J. (1982) *Nonparametric estimation for nondecreasing Lévy processes*, Journal of the Royal Statistical Society, Series B, Vol. **44**, pp. 262-269.

Bertoin, J. (1996) *Lévy processes*, Cambridge University Press.

Billingsley, P. (1968) *Convergence of probability measures*, Wiley, New York.

Birgé, L. and Massart, P., *An Adaptive Compression Algorithm in Besov Spaces*, Constr. Approx. **16**, 1-36.

Carr, P., Geman, H., Madan, D.B. and Yor, M. *The Fine Structure of Asset Returns: An Empirical Investigation*, Journal of Business, vol. **75.2**, pp. 305-332.

Chung, K.L. (2001) *A course in probability theory* (3rd Edition), San Diego, CA: Academic Press.

Csörgö, S. and Totik, V. (1983) *On how long interval is the empirical characteristic function uniformly consistent?*, Acta Sci. Math. (Szeged) **45**, pp. 141-149.

Davydov, Yu.A. (1973) *Mixing condtions for Markov chains*, Theory Probab. Appl. **18**, pp. 312-328.

Devore, R.A. and Lorentz, G.G. *Constructive Approximation.* Berlin: Springer-Verlag.

Eberlein, E. and Keller, U. *Hyperbolic distributions in finance.* Bernoulli **1**, 281-290.

Eraker, B. (2001) *MCMC Analysis of Diffusion Models With Application to Finance*, Journal of Business & Economic Statistics **19(2)** 177-191.

Feuerverger, A. and McDunnough, P. (1981) *On the efficiency of empirical characteristic function procedures*, Journal of the Royal Statistical Society, Series B, Vol. **43**, pp. 20-27.

Feuerverger, A. and McDunnough, P. (1981a) *On some Fourier Methods for Inference*, Journal of the Amer. Stat. Ass. **76**, pp. 379-387.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications, vol. 2*, 2nd Edition, John Wiley & Sons, New York.

Figueroa-López, E. and Houdré, C. (2004) *Nonparametric estimation for Lévy processes with a view towards mathematical finance*, submitted to Ann. Stat.

Getoor, R.K. (1975) *Markov processes: Ray processes and right processes*, Lecture notes in mathematics 440, Berlin: Springer-Verlag.

Ghosh J.K. and Ramamoorthi, R.V. (2003) *Bayesian Nonparametrics*, Springer.

Ghosal, S. and Vaart, A.W. van der (2003) *Posterior convergence rates of Dirichlet* *mixtures of normal distributions for smooth densities.*

Ghosal, S. and Vaart, A.W. van der (2004) *Convergence rates of posterior distributions for noniid observations*, submitted Ann. Stat.

Giné, E. and Zinn, J. (1986) *Lectures on the central limit theorem for empirical processes* Lecture notes in mathematics **1221**, p. 50-113.

Groeneboom, P., Jongbloed, G. and Wellner, J.A. (2003) *The support reduction algorithm for computing nonparametric function estimates in mixture models*, submitted.

Härdle W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998) *Wavelet, Approximation and Statistical Applications.* Lectures notes in statistics **129**. Springer, New York.

Harn, K. van and Steutel, F.W. (2004) *Infinite divisibility of probability distributions on the real lihne*, New York: Marcel Dekker.

Ibragimov, I.A. and Has'minskii R.Z. (1981) *Statistical Estimation: Asymptotic Theory*, Springer-Verlag New York.

Itô, K. and McKean, H.P., Jr. (1965) *Diffusion processes and their sample paths.* Springer-Verlag, Berlin.

Jacod, J. and Shiryaev, A. N. (1987). *Limit theorems for stochastic processes.* Springer-Verlag, Berlin.

Jameson, G.J.O. (1974) *Topology and Normed Spaces*, London: Chapman and Hall.

Karatzas, I. and Shreve, S.E. (1991) *Brownian motion and stochastic calculus* 2nd ed., Springer-Verlag, New York.

Kaspi, H. and Mandelbaum, A. (1994) *On Harris recurrence in continuous time*, Mathematics of operations research, Vol **19** No. 1.

Kerkyacharian, G., Picard, D, (2004) *Entropy, Uniersal Coding, Approximation and Bases Properties.* Constr. Approx. **20** 1–37.

Knight, J.L. and Satchell, S.E. (1997) *The cumulant generating function method: implementation and asymptotic efficiency*, Econometric Theory **13**, pp 170-184.

Koponen, I. (1995) *Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic processl.* Physics Review E **52**, 1197-1199.

Korsholm, L. (1999) *Some Semiparametric Models*, Ph.D. Thesis, University of Aarhus.

Krengel, U. (1985) *Ergodic theorems*, Berlin: De Gruyter.

Kutoyants, Yu,A. *Parameter estimation for stochastic processes*, Berlin: Heldermann Verlag.

Kutoyants, Yu,A. (1994) *Identification of dynamical systems with small noise*, Kluwer Academic Publishers.

Kutoyants, Yu,A. (2004) *Statistical inference for ergodic diffusion processes*, Springer.

Lang, S. (1985) *Complex Analysis* 2nd Edition, Springer-Verlag.

Liptser, R.S. and Shiryayev, A.N. (1977). *Statistics of random processes I.* Springer-Verlag, New York.

Luong, A. and Thompson, M.E. (1987) *Minimum-distance methods based on quadratic distances for transforms*, Canad. J. Stat. **15** pp. 239-251.

Madan, D.P. and Seneta, E. (1990) *The VG model for Share Market Returns.* Journal of Business **63**, 511-524

Masuda, H. (2004) *On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process*, Bernoulli **10**, p. 97-120.

Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability*, London:

Springer-Verlag.

Meyn, S.P. and Tweedie, R.L. (1993a) *Stability of Markovian processes II: Continuous-time processes and sampled chains*, Adv. Appl. Prob. **25**, pp. 487-517.

Nielsen, B. and Shephard, N. (2003) *Likelihood analysis of a first order autoregressive model with exponential innovations*, Journal of Time Series Analysis **24**, p. 337-344.

Nummelin, E. (1984) *General irreducible Markov chains and non-negative operators*, Cambridge.

Olerian, O., Chib, S. and Shephard, N. (2001) *Likelihood inference for discretely observed nonlinear diffusions*, Econometrica **69**, 959-993

Pollard, D. (1984) *Convergence of Stochastic Processes*, Springer-Verlag, New York.

Pollard, D. (1990) *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics **2**, Institute of Mathematical Statistics and American Statistical Association

Protter, P. (2004) *Stochastic integration and differential equations*, Second edition. Berlin: Springer-Verlag.

Prakasa Rao, B.L.S. (1999), *Statistical inference for diffusion type processes*, London: Edward Arnold.

Remmert, R. (1991) *Theory of Complex Functions*, Springer.

Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer.

Revuz, D. and Yor, M. (1999) *Continuous martingales and Browmian Motion* (3rd edition), Berlin: Springer-Verlag.

Rio, E. (1998) *Processus empiriques absolument réguliers et entropie universelle*, Prob. Theory Relat. Fields **111**, p. 585-608.

Rio, E. (2000) *Théorie asymptotique des processus aléatoires faiblement dependants*, In: Collection Mathématiques & Applications, vol. **31**, Berlin: Springer-Verlag.

Roberts, G., Papaspiliopoulos, O. and Dellaportas, P. (2004) *Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes*, Journal of Royal Statistical Society, Series B, Vol. **66**, pp. 369-393.

Rogers, L.C.G. and Williams, D. (2000) *Diffusions, Markov processes and martingales, vol.1*, 2nd Edition, Cambridge university press.

Rogers, L.C.G. and Williams, D. (2000) *Diffusions, Markov processes and martingales, vol.2*, 2nd Edition, Cambridge university press.

Rosiński, J (2001) *Series representations of Lévy processes from the perspective of point processes*, In: *Lévy processes*, Birkhäuser Boston, pp. 401–415.

Rubin, H. and Tucker, H.G. (1959) *Estimating the parameters of a differential process*, Ann. Math. Statist. **30**, pp. 641-658.

Rudin, W. (1987) *Real and comples analysis* (3rd edition), New York: McGraw-Hill.

Sato, K.I. (1999) *Lévy processes and infinitely divisible distributions*. Cambridge University Press.

Schorr, B. (1975) *Numerical inversion of a class of characteristic functions*, BIT **15**, pp. 94-102.

Shen, X. and Wasserman, L. (2001), *Rates of convergence of posterior distributions*, Ann. Statist. **29** 687-714.

Shiga, T. (1990) *A recurrence criterion for Markov processes of Ornstein-Uhlenbeck type*, Prob. Theory Related Fields **85**, pp. 425-447.

Shiryaev, A.N. (1999) *Essentials of stochastic finance*, Advanced Series on Statistical

Science & Applied Probability, **3**, (Translated from the Russian manuscript by N. Kruzhilin),World Scientific Publishing Co.

TUCKER, H.G. (1967) *A Graduate Course in Probability*. New York: Academic Press.

USHAKOV, N.G. (1999) *Selected Topics in Characteristic Functions*, Utrecht: VSP BV.

VAART, A.W. VAN DER (1998) *Asymptotic Statistics*, Cambridge University Press.

VAN DER VAART, A.W. AND WELLNER, J.A. (1996). *Weak convergence and empirical processes with applications to statistics*. Springer-Verlag, New York.

VAN ZANTEN, J.H. (2003), *On uniform laws of large numbers for ergodic diffusions and consistency of estimators*, Stat. Inference Stoch. Process, **6**(2), 199–213.

ZANTEN, J.H. VAN (2004) *On the rate of convergence of the MLE in Brownian semimartingale models*, accepted by Bernoulli.

ZHAO, L.H. (2000), *Bayesian aspects of some nonparametric problems.* Ann. Statist. **28** 532-552.

Samenvatting (Summary in Dutch)

Statistische schatting voor Lévy aangestuurde OU-processen en Brownse semimartingalen

In dit proefschrift behandelen we statistische methoden voor stochastische differentiaalvergelijkingen. In de financiële wiskunde worden zulke vergelijkingen gebruikt om processen waarbij onzekerheid optreedt te modelleren. Hierbij valt bijvoorbeeld te denken aan aandelenkoersen of rentestanden. De nadruk in dit proefschrfit ligt op het analyseren van asymptotische eigenschappen van statistische schattings-methoden.

Simpel gezegd zijn stochastische differentiaalvergelijkingen normale differentiaalvergelijkingen, waarbij een stochastische component is toegevoegd. Tezamen met de drift- en diffusie-coëfficient legt deze component de vergelijking vast. De resulterende oplossing is dan een stochastisch proces, met eigenschappen die nauw gerelateerd zijn aan de onderliggende stochastische component. Vaak wordt voor het onderliggende stochastische proces een Brownse beweging gekozen, maar in toenemende mate zien we modellen waarin de Brownse beweging vervangen wordt door een meer algemene semimartingaal en, als een speciaal geval hiervan, door een Lévy proces. Laatstgenoemde is een proces met onafhankelijke en stationaire incrementen. Standaard voorbeelden hiervan zijn de Brownse beweging en het compound Poisson proces. De introductie van steeds complexere modellen noodzaakt de ontwikkeling van nieuwe statistische methoden. In dit proefschrift ontwikkelen we zulke methoden voor een tweetal modellen. Hoofdstukken 1 tot en met 3 concentreren zich op Ornstein-Uhlenbeck processen aangestuurd door een stijgend Lévy process; hoofdstuk 4 behandelt Bayesiaanse schatters voor Brownse semimartingalen.

In hoofdstuk 1 introduceren we de meest belangrijke concepten die nodig zijn voor de hoofdstukken 2 en 3. We starten met een introductie tot Lévy processen en oneindig deelbare verdelingen. Vervolgens bespreken we stationaire Ornstein-Uhlenbeck processen (OU-processen), waarbij de stochastische component een stijgend Lévy proces is. Ruwweg is zo'n OU-proces een proces dat positieve sprongen maakt volgens het onderliggende Lévy proces, en exponentieel daalt tussen de opeenvolgende sprongtijdstippen. De stationaire verdeling van zo'n proces behoort tot de klasse van positief zelf-ontbindbare verdelingen, een subklasse van de oneindig deelbare verdelingen. Als zodanig wordt een positief zelf-ontbindbare verdeling vastgelegd door haar Lévy maat, waarvan de dichtheid gekarakteriseerd wordt door een dalende functie op de positieve halflijn, de zogenaamde kanonieke functie. Het schatten van deze kanonieke functie is het onderwerp van hoofdstuk 2. Naast de introductie van de diverse processen, geven we in hoofdstuk 1 verschillende voorbeelden van Lévy en OU-processen, behandelen we enkele simulatie-methoden en leiden we enige probabilistische eigenschappen van OU-processen af.

Uitgangspunt voor hoofdstuk 2 is de observatie van een stationair OU-proces op discrete tijdstippen. Op basis van deze data definiëren we een niet-parametrische schatter voor de kanonieke functie, welke gebaseerd is op cumulant functies. Een cumulant functie is de logaritme van een karakteristieke functie. Startpunt voor onze schatter is een initiële schatter voor de cumulant van de stationaire verdeling. Dit kan bijvoorbeeld de cumulant behorende bij de empirische verdelingsfunctie van de waarnemingen zijn. Door een kwadratische afstand tussen de initiële cumulant en de theoretische cumulant te minimaliseren over de verzameling van kanonieke functies wordt een schatter verkregen. We noemen deze schatter een cumulant-M-schatter. We tonen aan hoe deze schatter numeriek benaderd kan worden met een support-reductie algoritme en analyseren het (asymptotische) gedrag van de verkregen schatter. De resultaten worden geïllustreerd aan de hand van voorbeelden.

In hoofdstuk 3 analyseren we het gedrag van de parametrische variant van de cumulant-M-schatter. Deze schatter wordt verkregen via minimalisering van dezelfde criterium-functie als hiervoor, zij het dat we nu veronderstellen dat de Lévy dichtheid geparametriseerd wordt door een eindig-dimensionale parameter. We laten zien dat de schatter consistent en asymptotisch normaal verdeeld is. Voor een tweetal voorbeelden geven we aan hoe de verkregen schatter numeriek benaderd kan worden.

Het laatste hoofdstuk richt zich op Brownse semimartingalen, een klasse stochastische processen waartoe bijvoorbeeld diffusie-processen behoren. Uitgangspunt is de veronderstelling dat we zo'n proces continu in de tijd kunnen monitoren. De te schatten parameter is dan de drift-coëfficient, of een parameter in de drift-coëfficient. De focus ligt vervolgens op het analyseren van het gedrag van niet-parametrische Bayesiaanse schatters. Ons hoofdresultaat is een stelling die voorwaarden oplegt van waaruit de convergentie-snelheid van Bayesiaanse schatters verkregen kan worden. Deze voorwaarden zijn tweeledig: enerzijds is er een voorwaarde op de complexiteit van het model, anderzijds is er een voorwaarde op de a-priori verdeling. Voor een drietal speciale modellen werken we deze voorwaarden uit. Vervolgens geven we enige voorbeelden voor specifieke a-priori verdelingen.

Dankwoord

Op deze plaats wil ik een aantal mensen bedanken dat geholpen heeft bij de totstandkoming van dit proefschrift.

- Op de eerste plaats wil ik Geurt Jongbloed bedanken voor veel vertrouwen, steun en geduld. Geurt, naast de altijd leuke gesprekken met je, heb ik mijn vaardigheden als onderzoeker voor een heel groot deel aan jou te danken. Ik denk met veel plezier terug aan de vele middagen op jouw kamer en de conferenties waar we samen geweest zijn.
- Aad van der Vaart wil ik bedanken voor een prettige samenwerking en het gegeven vertrouwen om bij de Vrije Universiteit te kunnen starten als promovendus. De gegeven vrijheid en de mogelijkheid om altijd aan te kloppen voor advies heb ik als zeer prettig ervaren.
- Het onderzoek in het vierde hoofdstuk is gezamenlijk gedaan met Harry van Zanten. Ik wil Harry bedanken voor zijn input en voor zijn deelname in de leescommissie.
- Het idee om na mijn afstuderen als promovendus aan de universiteit te gaan werken is afkomstig van Piet Groeneboom. Ik realiseer me dat ik de geboden mogelijkheid bij de VU mede aan Piet te danken heb. Ik wil hem hiervoor bedanken, alsmede voor zijn deelname in de leescommissie.
- Dank aan Peter Spreij voor deelname in de leescommissie en het gedetailleerd lezen van een eerdere versie van dit proefschrift. De vele op- en aanmerkingen zijn het voor u liggende werk zeker ten goede gekomen.
- I would like to thank Michael Sørensen for participating in the reading committee.

Verder dank ik een aantal kamergenoten die ik gedurende de tijd als promovendus gehad heb, te weten Martine Reurings, Bas Kleijn en Sanne ter Horst bij de Vrije Universiteit en Henk de Koning bij de Universiteit van Amsterdam. Voorts dank ik Ronald Does voor de geboden tijd om bij IBIS-UvA mijn proefschrift af te ronden.

Op persoonlijk vlak kan ik vele mensen bedanken voor steun en de nodige afwisseling buiten werktijd. Ik houd het hier kort en noem alleen de belangrijksten voor mij. Ik bedank Christiaan Plaat en Patrick van Puffelen voor hun optreden als paranimf en voor een mooie vriendschap. Verder wil ik mijn ouders bedanken voor het feit dat ze altijd voor me klaar staan met raad en daad.

Curriculum Vitae

De schrijver van dit proefschrift werd geboren op 28 maart 1977 te Papendrecht. Van 1989 tot 1995 bezocht hij het Willem de Zwijger-college in Papendrecht, alwaar hij het VWO diploma behaalde. Vervolgens studeerde hij van 1995 tot 2001 Technische Wiskunde aan de Technische Universiteit in Delft. In februari 2001 startte hij als promovendus in de statistiek aan de Vrije Universiteit in Amsterdam. In april 2005 kreeg hij de mogelijkheid bij het Instituut voor Bebrijfs en-Industriële Statistiek aan de Universiteit van Amsterdam (IBIS-UvA) te gaan werken. Hier werd de laatste hand aan het voor u liggende proefschrift gelegd. Thans is hij werkzaam bij IBIS-UvA.